Collège Doctoral du Languedoc-Roussillon

THÈSE

Pour obtenir le grade de Docteur

Délivré par **Montpellier SupAgro**

Préparée au sein de l'école doctorale SIBAGHE
Et de l'unité de recherche AGAP

Spécialité **: Biologie intégrative**

Présentée par **Gabriel Sérgio COSTA ALVES**

CHARACTERIZATION OF A CANDIDATE GENE FOR DROUGHT TOLERANCE IN *COFFEA* : THE *CcDREB1D* GENE, IN CONTRASTING GENOTYPES OF *COFFEA CANEPHORA* AND RELATED SPECIES

Soutenue le 27 février 2015 devant le jury composé de

Dr. Jacqueline Grima-Pettenati, DR, CNRS	Rapporteur
Pf. Luiz Vieira, PR, U. do Oeste Paulista	Rapporteur
Dr. Pierre Marraccini, PR associé, U. Lavras	Co-directeur
Dr. Dominique This, MCF, Montpellier SupAgro	Co-directrice
Dr. Hervé Etienne, chercheur, CIRAD	Invité
Dr. Adilson Kobayashi, chercheur, EMBRAPA	Invité

Thèse préparée en cotutelle avec
l'Université Fédérale de Lavras, Brésil

UFLA UNIVERSIDADE FEDERAL DE LAVRAS

Montpellier SupAgro

Sumary

1. Introduction ... 4
1.1. Coffee production: global context and challenges ... 4
1.2. Abiotic stress in coffee ... 6
1.2.1. Temperature ... 6
1.2.2. Water availability ... 7
1.2.3. Coffee adaptation to water scarcity ... 9
1.3. Molecular bases of plants responses to drought ... 11
1.3.1. Signal transduction in the perception of dehydration ... 14
1.3.2. The rapid and emerging response via transcription factors ... 19
1.3.3. *Cis*-regulatory elements and the adaptive response to drought ... 20
1.4. AP2/ERF family of transcription factors ... 25
1.4.1. DREB transcription factors ... 26
1.4.2. Expression pattern and regulation of DREB1/CBF subgroup in abiotic stress response ... 30
1.4.3. Homologous genes and function prediction ... 33
1.5. Natural diversity and genetic variability in coffee ... 34
1.5.1. Structuration of *Coffea* diversity ... 35
1.5.2. *Coffea arabica* ... 35
1.5.3. *Coffea canephora* ... 36
1.5.4. Single Nucleotide Polymorphisms (SNP) as new molecular markers ... 38
1.5.5. Intra- e inter-genomic diversity ... 40
1.6. Associations between genetic diversity and adaptations ... 43
1.6.1. Molecular signatures of selection ... 43
1.6.2. Evolutionary significance of *cis*-regulatory mutations ... 45
1. Reference ... 47
CHAPTER 1 ... 58
Promoter of *Coffea DREB1D* as a major source of diversity for this gene and potential implication in local adaptation to drought tolerance. ... 59
1. Introduction ... 60
2. Materials and methods ... 64

3. Results 67

4. Discussion 70

5. Conclusions 76

6. Accession numbers 76

7. Acknowledgments 76

8. References 78

9. Main Figures List 82

10. Tabels 91

CHAPTER 2 96

CcDREB1D as part of the coffee DREB gene family: difficulties in extrapolating functional information from Arabidopsis to coffee. 97

1. Introduction 97

2. Materials and Methods 98

3. Results 100

4. Discussion 112

5. References 113

CHAPTER 3 116

Characterization of CcDREB1D promoter region from Coffea canephora and functional analysis to assess the responsiveness of the CcDREB1D alleles promoter Hp_15, Hp_16 and Hp_17 under drought stress. 117

1. Background 117

2. Introduction 120

3. Materials and Methods 122

4. Results 138

5. Discussion 166

6. References 175

7. Conclusions and Perspectives 183

8. Reference 187

Résume 189

Abstract 190

1. Introduction

1.1. Coffee production: global context and challenges

Coffee (*Coffea* L.) is the world's second most valuable traded commodity after crude oil. In 2013/14 coffee exportations were estimated in 107.6 million bags accounting for a US$ 17 billion exchange on trade market (International Coffee Organization (ICO), 2014). The coffee sector employs approximately 26 million people in 52 producing countries having a social and economic importance to developing countries (International Coffee Organization (ICO), 2014). Arabica coffee (*Coffea arabica* L.) and Robusta coffees (*C. canephora* Pierre) are the two main species used in the production of coffee, providing approximately 70% and 30% of commercial production, respectively. Three countries alone have in recent years produced around 55% of the world's coffee. The largest producer and exporter is Brazil (32%–34%), followed by Viet Nam (12%–13%) and Colombia (8%–9%) (Table 1). Around the globe, the annual consumption of coffee has expanded to 145 million bags (Figure 1).

Table 1 Total production in exporting countries

Crop year commencing	2011	2012	2013	2014*	% change 2013-14
Total	**136 246**	**147 477**	**146 772**	**141 420**	**-3.60%**
Arabicas	81 952	88 345	87 124	83 914	-3.70%
Colombian Milds	8 720	11 523	13 506	13 935	3.20%
Other Milds	31 913	28 914	26 935	26 653	-1.00%
Brazilian Naturals	41 319	47 908	46 683	43 326	-7.20%
Robustas	54 295	59 132	59 648	57 505	-3.60%
Africa	15 738	16 521	16 189	16 905	4.40%
Asia & Oceania	41 913	45 328	46 550	44 709	-4.00%
Mexico & Central America	20 194	18 504	16 800	17 313	3.10%
South America	58 401	67 124	67 233	62 492	-7.10%

Production (*estimated) given in thousand bags.
Full production data are available on the ICO website at www.ico.org/prices/po.htm
(Source: Coffee market report 2014 – ICO)

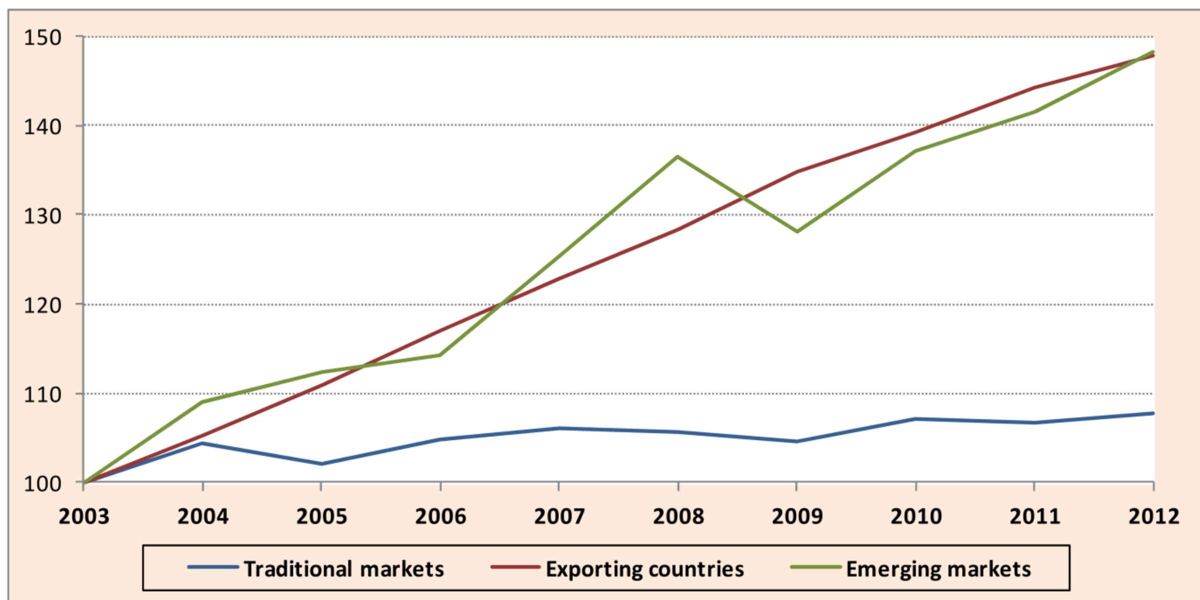


Figure 1 Index of consumption in traditional markets, exporting countries and emerging markets. The major consumers on traditional markers are USA, Germany, Japan, France, Italy, Canada, Spain and United Kingdom. The major consumers on emerging markets are Russia, Algeria, Poland, Republic of Korea and Australia. As for exporting countries, Brazil is the major consumer followed by Ethiopia, Indonesia, Mexico, Philippines, India, Venezuela, and Vietnam.
(Source: Coffee market report 2013 – ICO)

The productivity (green bean yield) of Arabica is tightly linked to climatic variability, and is thus strongly influenced by natural climatic oscillations (DaMatta and Ramalho, 2006; Camargo, 2010). Climatic variability has always been the main factor responsible for the fluctuation of coffee yields worldwide, and the climate change, as a result of global warming, is expected to presents a major challenge to the coffee industry (van Hilten, 2011). Despite of the uncertainty surrounds the effect of climate changes on individual producing regions and overall coffee production, some significant changes are expect to occur in some regions . Furthermore, the potential impact will not only vary between countries but also within producing areas in individual countries (van Hilten, 2011).

Scientific evidence suggests that climate change is accelerating at a much faster pace than previously thought and that important tipping points, leading to irreversible changes in major Earth systems and ecosystems, may already have been reached or even overtaken (IPCC, 2013). The Intergovernmental Panel on Climate Change predicts in the best scenario that global surface temperature change is likely to exceed 1.5°C by the end of the twenty- first century. Global temperatures have increased by an average of 0.85°C (0.65°C to 1.06°C) in the last 132 years (1880–2012), and the last three decades has

been successively warmer than any preceding decade since 1850 (IPCC, 2013). On this basis it has been forecast that the sustainability of the coffee industry faces serious challenges in the coming decades (Camargo, 2010; van Hilten, 2011; Davis et al., 2012). The future distribution of indigenous Arabica coffee has a profoundly negative trend under the influence of accelerated global climate change (Davis et al., 2012). In the most favorable outcome, the reduction in the number of pre-existing bio climatically suitable localities in Africa will be in the order of 65%, and at the worst, an almost 100% reduction, by 2080, in any case the outcome will be negative for the coffee industry (Davis et al., 2012).

Optimum cultivation requirements are likely to become increasingly difficult to achieve in many pre-existing coffee growing areas, leading to a reduction in productivity, increased and intensified management (e.g. the use of irrigation), and crop failure (some areas becoming unsuitable for coffee cultivation) (Davis et al., 2012).

One of the possible strategies to better prepare coffee producers to this future challenge is the development of new varieties that could cope with higher temperatures and water limitations, and remain highly productive contributing to the long-term sustainability of coffee cultivation in lands potentially affected by climate change.

1.2. Abiotic stress in coffee

1.2.1. Temperature

The favorable mean temperature for coffee plant growth lie between 18°C to 26°C, with the optimum annual mean temperature from 18°C to 21°C for Arabica coffee (Alègre, 1959) and 22°C to 30°C for Robusta coffee (Matiello, 1998; Willson, 1999). Although, some selected cultivars of Arabica under intensive management conditions could be successfully cultivated in marginal regions with average temperatures as high as 24-25ºC (DaMatta, 2004a). The optimum day/night mean temperature is 18/22°C, with tolerated extremes for Arabica coffee extending to a 15°C minimum during the night, and between 25°C - 30°C during the day (Descroix and Snoeck, 2004). Less adaptable to lower

temperatures than Arabica coffee, Robusta do not withstand temperatures at 15ºC for long periods (Willson, 1999).

High or low temperatures can significantly damage the productivity of coffee plants in several ways. Temperatures greater than 23°C can accelerated development and ripening of fruits, often leading to the loss of beverage quality (Camargo, 1985), and above 25°C can reduce the photosynthetic rate (Descroix and Snoeck, 2004). Continuous exposure to temperatures as high as 30°C leads to depressed growth and abnormalities, such as the yellowing of leaves, growth of tumors on the stem and during blossoming may cause abortion of flowers, reducing coffee yields (DaMatta, 2004a; Descroix and Snoeck, 2004). In regions with a mean annual temperature below 17–18°C growth is also depressed (DaMatta and Ramalho, 2006). Occurrence of frosts, even if sporadic, may strongly limit the economic success of the crop. In that case, low temperatures cause the discoloration of leaves, exposure to temperatures below 1°C can result in serious lesions on both the leaves and coffee cherries and below -2°C for durations of 6 hours or more can cause plant death (Descroix and Snoeck, 2004). Arabica coffee trees are severely damaged by frost, and are therefore not suited to regions that experience sub-zero temperatures, even for short periods of time (Gay et al., 2006). Large changes in diurnal temperatures can affect yield and quality, and the maximum tolerance is a range of 19°C (Descroix and Snoeck, 2004).

In addition to the direct control on coffee growth, temperature has an indirect influence. For example, high and low temperature extremes increase the threat to coffee from pests and disease. High temperatures favor the development of Coffee Leaf Rust (*Hemileia vastatrix*) and fruit blight, while Coffee Berry Disease is more prominent in cool regions (Descroix and Snoeck, 2004).

1.2.2. Water availability

Water availability is determined by total levels of precipitation and atmospheric humidity. Rainfall is the most restrictive factor in coffee growing regions, and both the total annual rainfall, and monthly distributions of precipitation are important (Descroix and Snoeck, 2004). Rainfall requirements depend on several factors like absorption capacity of the soil, atmospheric humidity, sky cover and

also cultivation practices (DaMatta and Ramalho, 2006). A soil water deficit decreases the biological and economic productivity of coffee, by lowering the quantity and quality of the yield (DaMatta et al., 1998).

If the dry season is not prolonged, and soils have a high water retention capacity, coffee can be grown without irrigation in areas where total precipitation exceeds1100mm a year (Descroix and Snoeck, 2004). Although coffee can be viable in areas that receive between 800mm and 1000mm of precipitation a year, even if this is ideally distributed in time, overall productivity will be low (Descroix and Snoeck, 2004). For higher yields, more precipitation is required. It is generally considered that the optimal annual rainfall for Arabica coffee lies within the range of 1200-1800mm (Alègre, 1959; Wrigley, 1995; DaMatta, 2004a; Descroix and Snoeck, 2004), though a range between 762 and 2540 mm remains acceptable. In coffee producing areas if total annual rainfall lies between 2500mm and 3000mm and providing soils have good drainage properties, Arabica coffee is productive (Descroix and Snoeck, 2004). If total precipitation exceeds 3000mm, leaf diseases from fungal infections can develop more easily and the region is usually less successful in producing economically viable coffee.

It is important that the rains are well distributed over the season or are continuous for about 7-8 months. The nature of the rainy season in terms of length and intensity of the rains is a key ecological factor determining the interval between flowering and seed maturation (Carr, 2001). From both field observations and controlled irrigation experiments, it has been shown that coffee plants require a period of water shortage to trigger floral initiation (Alvim, 1960; Crisosto et al., 1992; Gay et al., 2006). A dry season of 12 – 14 weeks is critical for the growth and development of Arabica coffee, as it enables internal water stress to induce flowering after dry season. Months in which rainfall is less than twice the monthly average temperature can be considered as ‘dry months’ (Descroix and Snoeck, 2004). When the rains occur, plants break out of dormancy, triggering blossoming and rapid shoot growth. If the soil conditions are favorable, Arabica coffee can withstand a dry season of up to six months . In regions lacking a dry season, harvest periods are scattered throughout the year, and overall annual yields are low (DaMatta, 2004a). High relative humidity can reduce the water loss from the coffee plant and soil, and optimal atmospheric humidity is 60% for Arabica coffee. Relative humidity

above 85% can negatively affect the quality of coffee (Descroix and Snoeck, 2004). Cloud cover and mist increases the relative humidity, which is advantageous during the dry season. Morning dew can also provide an additional water source, especially in mountainous regions.

1.2.3.Coffee adaptation to water scarcity

Dehydration triggers a series of physiological and metabolic changes in coffee plants (Montagnon and Leroy, 1993). A complementary approach to improve plant performance for drought stress conditions involves the identification and selection of traits that enhances drought tolerance or water- use efficiency (WUE) (DaMatta, 2004b). Coffee retains high leaf relative water content (RWC) under dehydrating conditions, being considered a water-saving rather than a dehydrating tolerant species. A small water loss should cause a shift in turgor and as consequence leaves tend to maintain a high RWC by retaining a high symplast volume. Therefore, the rates of water use or efficiency of extraction of soil water appears to be a major component of differential adaptation to drought among Arabica and Robusta coffee genotypes under water scarcity. This may be attributed to an efficient stomatal control on transpiration and/or low cell-wall elasticity (Pinheiro et al., 2005). To prevent further loss of water, the plant induces stomatal closure resulting in reduction in carbon dioxide (CO_2) uptake and consequent change in photosynthetic rate.

Plants of *C. canephora* have less stomatal conductance sensitivity to variations in level of soil water than *C. arabica* . Such differences may be explained by the origin of species. The Ethiopia native region of *C. arabica*, is situated between 1600-2800 m high, has low atmospheric humidity and a dry season 4-5 months a year. *C. canephora* is coming from the Congo River basin region of low altitude, humidity close to saturation, tropical equatorial climate and high rainfall with homogeneous distribution of rainfall over 9-10 months. Thus, the requirement for an efficient stomatal control on transpiration would not therefore be as imperative for Robusta as for Arabica coffee. However, *C. canephora* has a considerably greater genetic diversity than *C. arabica*.

In *C. canephora*, the better crop yield of a drought-tolerant conilon clone, compared with a drought-sensitive one, was associated with maintenance of leaf area and higher tissue water potentials, as a

consequence of smaller stomatal conductance (g_s) (DaMatta, 2003). Longer root depth and more efficient stomatal control of water use were directly correlated to the differences in relative tolerance to drought stress of conilon clones 14 and 120 compared with clones 46 and 109A (Pinheiro et al., 2005) (Figure 2). Conilon coffee clones drought-tolerant (14 and 120) and –sensitive (46 and 109A) showed an intra-specific variation in stomatal behavior and in leaf water relations adjusted to changes in soil water supply and evaporative demand indicating a feed forward response to avoid large internal water deficit (Pinheiro et al., 2005) (Figure 3). Drought tolerance in coffee results from the combination of several physiological mechanisms.

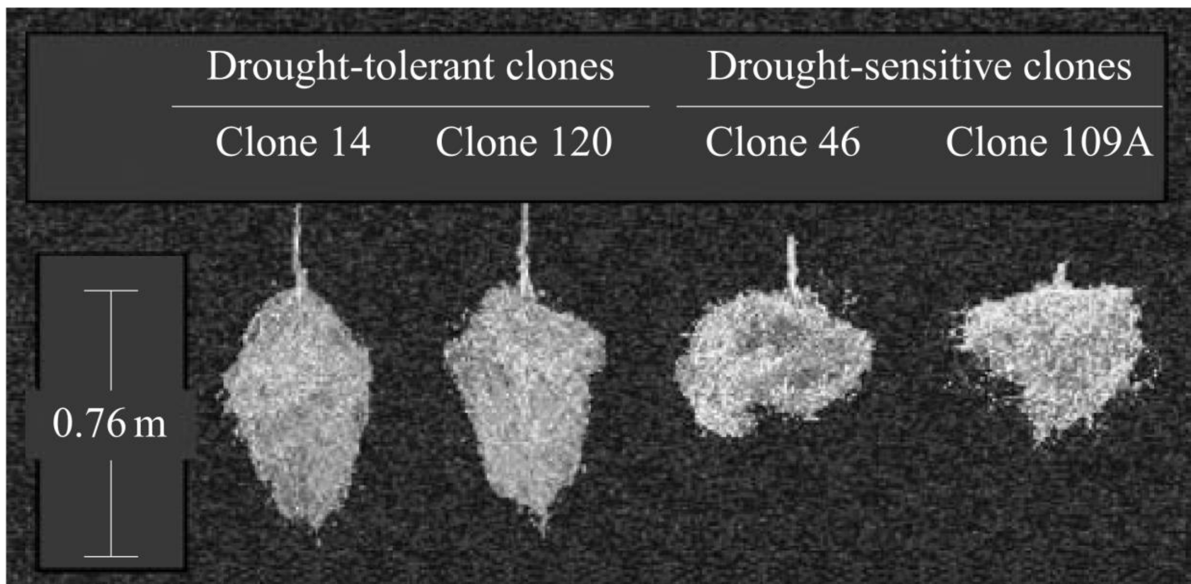


Figure 2 Typical root systems of four clones of Robusta coffee grown under full irrigation. (Source: Pinheiro et al 2005).

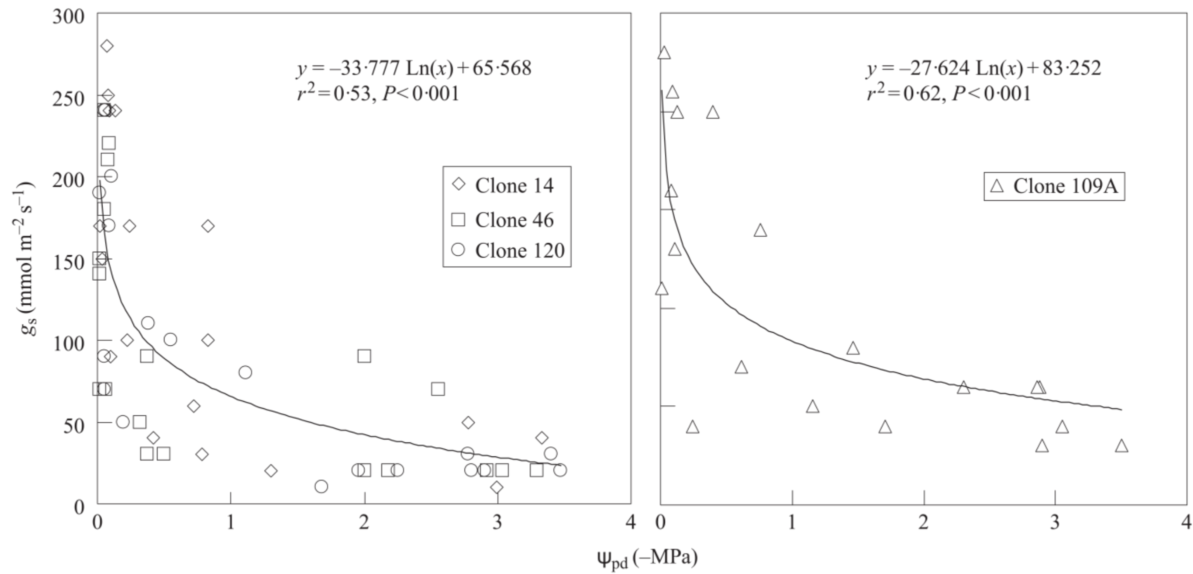


Figure 3 Stomatal conductance (g_s) in relation to pre-dawn leaf xylem pressure potential (Ψ_{pd}) in four clones of Robusta coffee. The g_s (expressed in mmol m^{-2} s^{-1}) was measured between 0700–0900 h and represents the entire data set from plants during dehydration after selecting a narrow range of leaf-to-air vapor pressure deficits (1.5 kPa at most). Note that stomatal conductance decreased sharply with decreasing Ψ_x, with no apparent value of Ψ_{pd} (expressed in MPa) at which stomatal closure was systematically observed. This stomatal sensitivity to evaporative demand observed in clones 14, 46 and 120, might indicate a feedforward response that would avoid large internal water deficits. (Source: Pinheiro et al 2005).

1.3. Molecular bases of plants responses to drought

Water deficit conditions cause a disruption of osmotic and ionic homeostasis in plant cells often resulted by the production of reactive oxygen species (ROS) and accumulation of free radicals ensuing lipid peroxidation, membrane damage, and inactivation of enzymes thus influencing cell viability (Bartels and Sunkar, 2005). Molecular responses to abiotic stress consist on the perception of these imbalances and its subsequently transduction in signal, gene expression and ultimately metabolic changes in the plant thus re-establishing cellular homeostasis (Wang et al., 2003) (Figure 4). Gene expression may be triggered directly by the stress conditions or result from secondary stresses and/or injury responses. Nonetheless, it is well established that drought tolerance is a complex phenomenon involving the concerted action of many genes. The genes activated at the transcriptional level in response to abiotic stresses can provide stress tolerance by the production of vital metabolic proteins and also by regulating downstream genes. Genes that normally are silent and activated under osmotic

stress conditions are often referred as osmotic stress-responsive (OR) genes and its products can largely be classified into two groups (Bohnert et al., 2001; Fowler and Thomashow, 2002; Seki et al., 2002). First group comprises of genes that encode for proteins that defend the cells from the effects of water-deficit, mainly including those that regulate the accumulation of compatible solutes (enzymes for osmolyte biosynthesis like proline, betaine, sugars, etc.); passive and active transport systems across membranes (water channel proteins and membrane transporters); and protection and stabilization of cell structures from damage by reactive oxygen species (ROS) (the detoxification enzymes such as catalase, superoxide dismutase, ascorbate peroxidase, etc.); fatty acid metabolism enzymes, proteinase inhibitors, ferritin and lipid-transfer proteins; and other proteins for the protection of macromolecules (Late embryogenesis abundant [LEA] protein, osmotin, chaperons, etc.) (Figure 4). The other group of genes includes regulatory proteins that further regulate the stress signal transduction and alter gene expression and hence possibly function in stress response. They comprise several transcription factors (TFs); protein kinases (microtubule-associated proteins [MAP] and CCAAT displacement protein [CDP] kinases, receptor protein kinase, etc.); protein phosphatases and proteinases implicated in the regulation of stress signaling and gene expression (Figure 4).

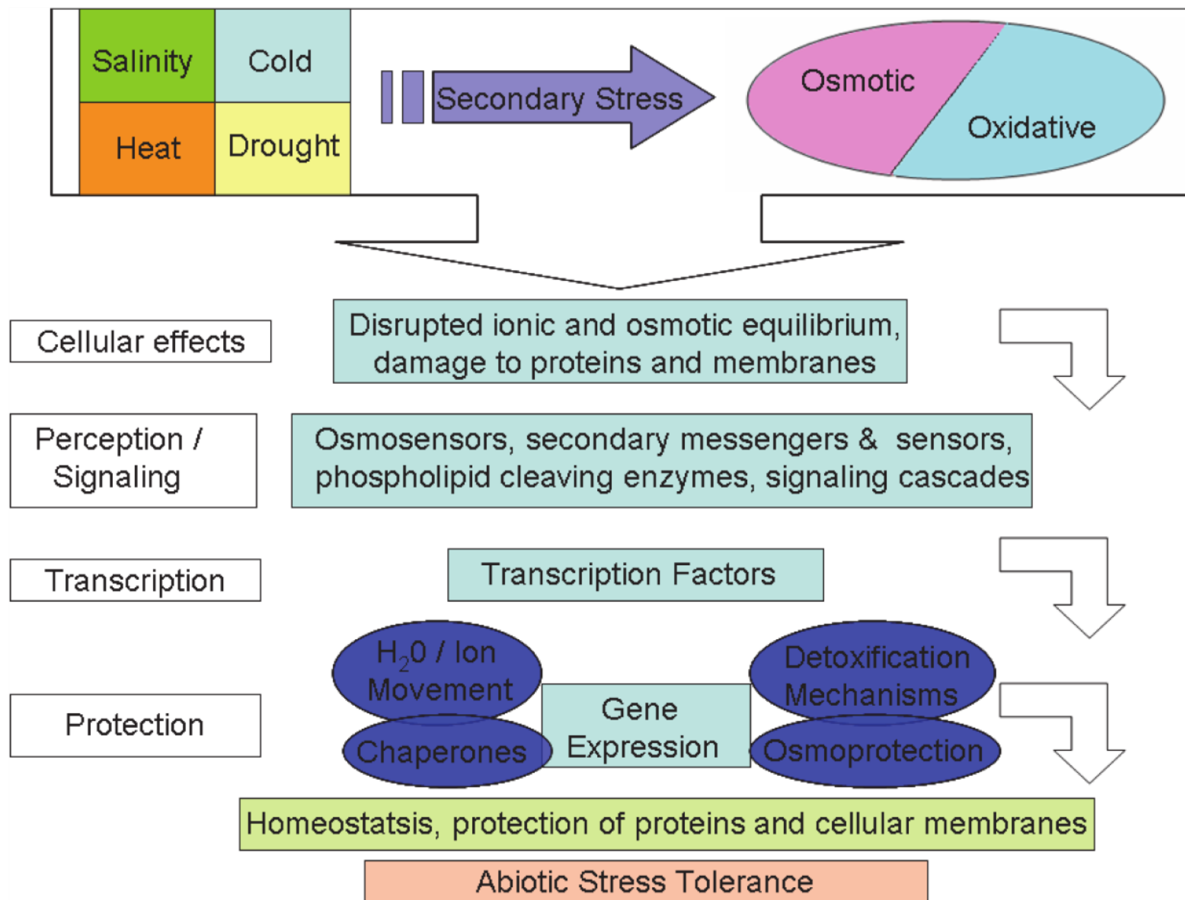


Figure 4 Depiction of the overlapping and complex cellular responses resulting from abiotic stress. Abiotic stress such as cold, drought, heat, and cold, are primary stresses causing cellular damages and secondary stresses, such as altered osmotic or ionic potential. The cell senses initial stress signals, such as osmotic and ionic effects, temperature, membrane fluidity changes, and triggers downstream signaling cascades, and that amplify of the signal. The final layer is the transcription factors, which initiate stress-responsive mechanisms in order to protect the cell, and restore homeostasis. Adapted from Wang et al. (2003).

1.3.1.Signal transduction in the perception of dehydration

Signal transduction of dehydration cascades from the sensing of water stress signals to the expression of various genes. As a result, differences in stress tolerance between genotypes or different developmental stages of a single genotype may arise from differences in signal perception and transduction mechanisms (Roychoudhury et al., 2013). Stomata closure is well characterized as a model system in the responses of plant cells to water stress. Prevention of osmotic stress caused by drought depends upon minimizing stomatal and cuticular water loss and maximizing water uptake. Increase in temperature or a rapid drop in humidity often results in acute water deficit condition in plants. Moreover, dry air mass, which moves into the environment, can also add to rapid and acute water losses from plants. Such atmospheric changes result in a dramatic increase in the vapor pressure gradient between leaf and the ambient air. This results in increased rate of transpiration. Moreover, increase in vapor pressure gradient enhances water loss from the soil. The first response of virtually all the plants to acute water deficit is the closure of their stomata to prevent the transpiratory water loss (MansWeld and Atkinson, 1991). Stomata closure may result from direct evaporation of water from the guard cells with no metabolic involvement, a process also called hydropassive closure. Stomatal closure may also be metabolically dependent and involve processes that result in reversal of the ion Xuxes that cause stomatal opening. This process of stomatal closure, which requires ions and metabolites, is known as hydroactive closure (Mahajan and Tuteja, 2005). The guard cell signaling is of critical importance because it plays a major role in plant water-use efficiency (WUE) (Roychoudhury et al., 2013). During stomata closure, the level of cytoplasmic Ca^{2+} increases, which suggests that Ca^{2+} functions as a second messenger in the osmotic stress response (Schwartz et al., 1988). This perturbation in cytosolic Ca^{2+} level is sensed by calcium binding proteins, also known as Ca^{2+} sensors that apparently lack any enzymatic activity and change their conformation in a calcium dependent manner. These sensory proteins then interact with their respective interacting partners often initiating a phosphorylation cascade and target the major stress responsive genes or the TFs controlling these genes (Huang et al., 2012). Several protein kinases reported in plants function in

phosphorylation processes of many signal transduction pathways. Phosphoinositide (IP) signaling has been implicated in the elevation of cytoplasmic Ca^{2+} in guard cells. The elevation of IP results in Ca^{2+} mobilization and the IP content has been demonstrated to increase following hyperosmotic stress. Signaling molecules like inositol 1,4,5-triphosphate (IP3) are generated by membrane phospholipids dynamic system under the perception of stress stimuli. In that case, phospholipase C (PLC) catalyzes the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP2) into diacylglycerol (DAG) and IP3, which acts as second messengers. IP3 releases Ca^{2+} from internal stores (Mahajan and Tuteja, 2005) (Figure 5). Calcium is perhaps the main signal transducer in the signaling cascades activated in plant response to any stimulus or stress, and the ubiquitous characteristic of this molecule in stress signaling justifies the role of the Ca^{2+} cation as an important node for all pathways of abiotic stress response (Shanker and Venkateswarlu, 2011).

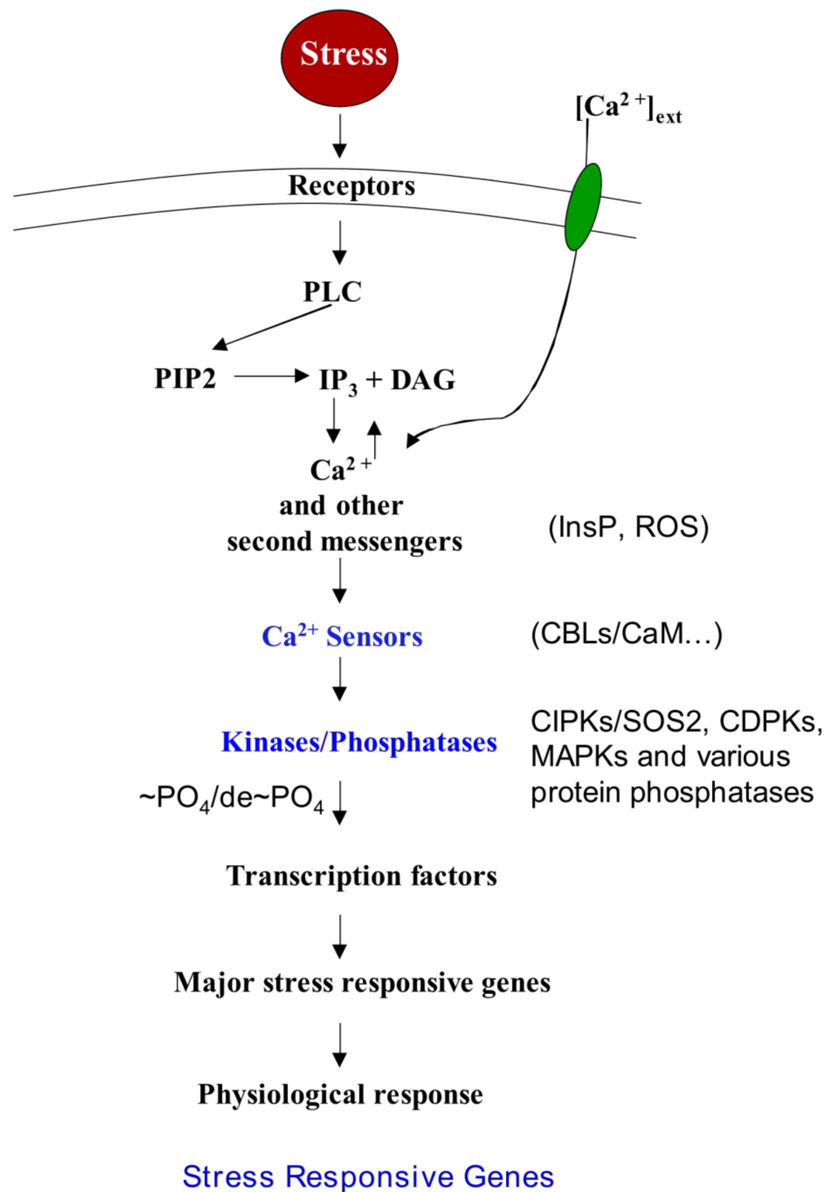


Figure 5 Overview of signaling pathway under stress condition highlighting the participation of Ca^{2+}. Stress signal is first perceived by the membrane receptor, which activates PLC and hydrolyses PIP2 to generate IP3 as well as DAG. Following stress, cytoplasmic calcium levels are up-regulated via movements of Ca^{2+} ions from apoplast or from its release from intracellular sources mediated by IP3. This change in cytoplasmic Ca^{2+} level is sensed by calcium sensors which interact with their downstream signaling components which may be kinases and/or phosphatases. These proteins affect the expression of major stress responsive genes leading to physiological responses.
(Source: Mahajan and Tuteja, 2005).

Stress induced changes in gene expression resulting of the first signaling cascade in turn may participate in the generation of hormones like abscisic acid (ABA), salicylic acid (SA) and ethylene. These molecules may amplify the initial signal and initiate a second round of signaling that may follow the same pathway or use altogether different components of signaling pathway (Mahajan and Tuteja, 2005; Shao et al., 2007). Cytoplasmic pH is another possible second messenger of ABA signaling in guard cells, and it functions in the Ca^{2+} -independent pathway. The endogenous ABA level is triggered in a plant system in response to various stress signals, mainly due to the induction of genes encoding enzymes responsible for ABA biosynthesis from *β*-carotene (Roychoudhury and Basu, 2012). Various stress signals and ABA share common elements in the signaling pathway and these common elements cross-talk with each other, to maintain cellular homeostasis. ABA controls many stress adaptation responses, activation of OR genes involved in osmotic adjustment (OA), helps in limiting transpiration rate and controls wilting, thus reducing water loss in plants (Agarwal and Jha, 2010a; Roychoudhury et al., 2013). Once present, ABA evokes an alkalization of the cytoplasm of guard cells, and this has a relationship with the activation of outward-rectifying K^+ channels by ABA. The interaction between the Ca^{2+} pathways and ABA has a crucial role on the induction of stomatal closure. ABA-induced stomatal closure is reduced to only 30% of the normal stomatal closure response under inhibition of Ca^{2+} cytoplasmic accumulation in guard cells (Siegel et al., 2009). Further, ABA enhances the ability of cytosolic Ca^{2+} to activate S-type anion channels and down-regulate inward-rectifying K^+ channels, increasing the [Ca^{2+}] sensitivity of stomatal closure mechanisms. While measured in a low-extracellular-Ca^{2+} bath, guard cell S-type anion channels show little response to an increase in cytoplasmic [Ca^{2+}] to 2 μM (Siegel et al., 2009; Kim et al., 2010). However, if ABA is added to the same solution during preincubation of guard cells, the ability of 2μM cytoplasmic [Ca^{2+}] to activate anion channel currents increase strongly (Siegel et al., 2009). These findings provide evidence that ABA enhances/primes the ability of guard cells to respond to increased cytoplasmic [Ca^{2+}] levels and to activate anion channels (Siegel et al., 2009) (Figure 6).

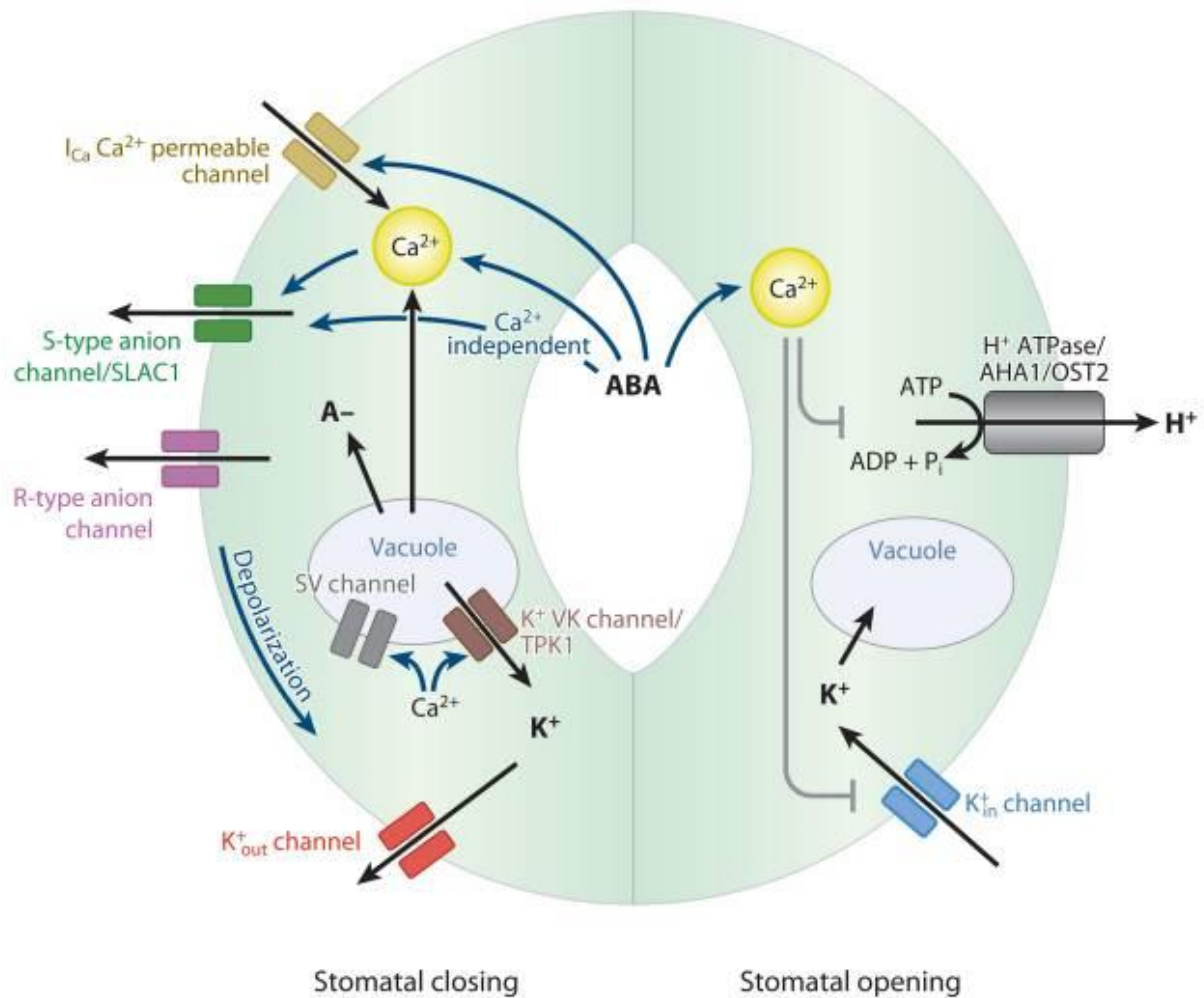


Figure 6 Summary of guard cell signaling and ion channel regulation. This model focuses on guard cell ion channel functions and ABA-induced signal transduction across the plasma membrane and vacuolar membrane of guard cells. Signaling events during stomatal closing are shown in the left guard cell, and major regulation steps for ABA-inhibition of stomatal opening mechanisms are shown in the right guard cell. Abbreviations: ABA, abscisic acid; I_{Ca}, inward Ca^{2+} current; S-type, slow-type; SLAC1 (Slow anion channel associated 1); R-type, rapid-type; SV, slow vacuolar; VK, vacuolar K^+ selective; TPK1 (Two pore K^+ channel 1); AHA1 (*Arabidopsis* H^+ ATPase 1); OST2 (Open stomata 2).
(Source: Kim et al., 2010).

Parallel signaling pathway could provide a molecular basis for Ca^{2+}- and ABA- induced Ca^{2+} sensitivity priming. Experimentally imposed variations on $[Ca^{2+}]$cyt (concentration of cytoplasmic calcium), regardless of the variation pattern, have been shown to trigger a Ca^{2+}-reactive stomatal closure response (Allen et al., 2001; Yang et al., 2003; Li et al., 2004). The chloroplastic protein CAS (Calcium sensing receptor) was reported to regulate $[Ca^{2+}]$cyt elevations in response to elevated extracellular $[Ca^{2+}]$ext (Webb, 2008; Weinl et al., 2008). Ca^{2+} sensitivity priming of specific Ca^{2+} sensors may explain specificity in Ca^{2+} signaling and responses and how opposing signaling pathways like ABA-induced stomatal closure and blue-light- and low-$[CO_2]$- induced stomatal opening can both employ $[Ca^{2+}]$cyt elevations as a secondary messenger and nonetheless retain specificity (Harada et al., 2003; Young et al., 2006).

1.3.2. The rapid and emerging response via transcription factors

Transcription factors are proteins that act together with other transcriptional regulators, including chromatin remodeling/modifying proteins, to employ or obstruct RNA polymerases to the DNA template (Lata et al., 2007). Plant genomes assign approximately 7% of their coding sequence to TFs, which proves the complexity of transcriptional regulation (Agarwal et al., 2006b). Transcriptome data in *Arabidopsis* and in numerous other plants suggest that there are several pathways that independently respond to environmental stresses (in both ABA dependent- and independent- manner), suggesting that stress tolerance or susceptibility is controlled at the transcriptional level by an extremely intricate gene regulatory network (Agarwal and Jha, 2010b). The phytohormone ABA is the central regulator of abiotic stress particularly drought resistance in plants, and coordinates a complex gene regulatory network enabling plants to cope with decreased water availability (Lindemose et al., 2013). ABA-dependent signaling systems have been illustrated as pathways that mediate stress adaptation by induction of at least two separate regulons (a group of genes controlled by a certain TFs): the ABA-responsive element-binding protein/ABA-binding factor (AREB/ABF) regulon and (2) the myelocytomatosis oncogeneb/myeloblastosis oncogene (MYC/MYB) regulon (Agarwal and Jha,

2010a). While ABA-independent regulons are the DRE-binding protein/C-repeat binding factor (DREB/CBF) regulon, and the NAC and zinc- finger homeodomain (ZF-HD) regulon.

In contrast, there are genes or TFs containing both DRE/CRT and ABRE, which can integrate input stimuli from salinity, drought, cold and ABA signaling pathways, thereby enabling cross-tolerance to multiple stresses. Transcription factors fall in the category of early genes and are induced within minutes of stress (Huang et al., 2012). The TFs interact with *cis* elements in the promoter regions of various abiotic stress-related genes and thus upregulate the expression of many secondary responsive downstream genes, imparting stress tolerance (Figure 7).

1.3.3. *Cis*-regulatory elements and the adaptive response to drought

Cis-acting regulatory elements are important molecular switches involved in the transcriptional regulation of a dynamic network of gene activities. These regulatory elements participate in the control of various biological processes, including abiotic stress responses, hormone responses and developmental processes (Yamaguchi-Shinozaki and Shinozaki, 2005). *Cis*-acting elements are specific binding sites for proteins involved in the initiation and regulation of transcription (Hernandez-Garcia and Finer, 2014). Prior to initiation of transcription, a protein complex is formed with RNA polymerase II and general TFs such as TFIID and TFIIB in the "core promoter" (a region located ~40 bp upstream of the transcriptional initiation site of protein-encoding genes). This protein complex/DNA interaction is modulated by general TFs and *cis*-regulatory elements present on the core promoter, like TFIID TBP (TATA-box-Binding Protein) which binds to TATA box (Molina and Grotewold, 2005) (Figure 7A). However, the presence of *cis*-regulatory motifs is not restricted to the core promoter. Upstream of the core promoter are the proximal and distal regions of promoters. These regions contain different *cis*-regulatory sequences that contribute to the fine regulation of gene expression, such as enhancers, silencers/repressors and insulators (Hernandez-Garcia and Finer, 2014) (Figure 7B). The physical demarcation of the upstream regions that contribute to the "full promoter" is more fitful than that of the core promoter. The size of the active, fully functional promoter depends directly on the positional and combinatorial understanding of the *cis*-acting elements present in both

the proximal and distal regions (Hernandez-Garcia and Finer, 2014). During transcription, co-activators and TFs bind to these *cis*-elements and simultaneously interact with the transcriptional machinery attached to the core promoter (Lee and Young, 2000) (Figure 7B). These multiples interactions function as molecular switches that lead to the activation, enhancement, or suppression of transcription. Thus, regulation of transcription depends on the (i) availability and activity of TFs, and (b) presence or absence, copy number, location and combinatorial relationship of *cis*-regulatory elements present in and around the promoter (Zou et al., 2011; Hernandez-Garcia and Finer, 2014; Mishra et al., 2014).

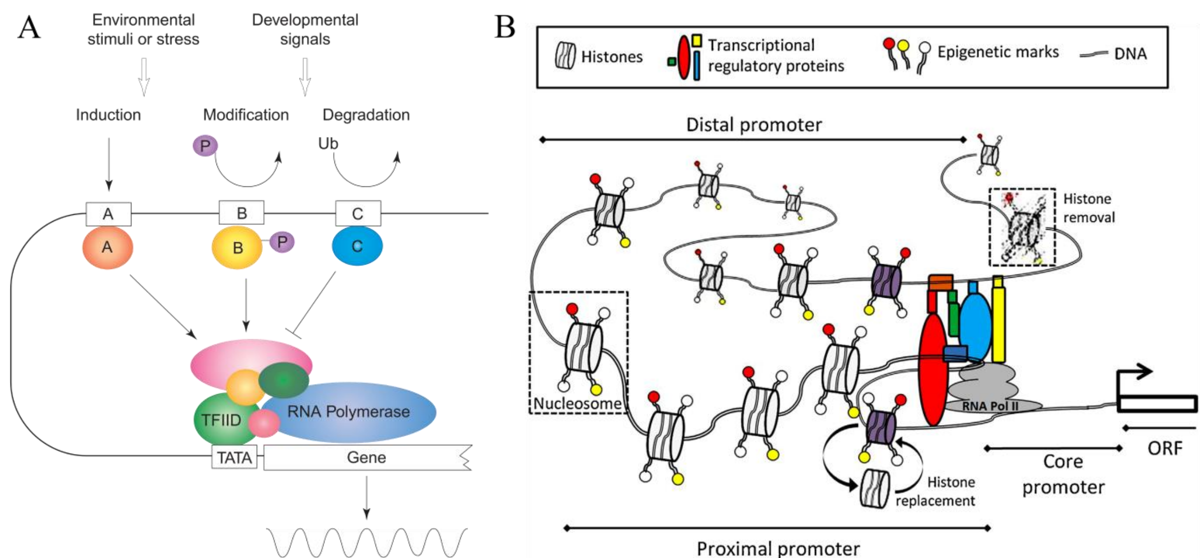


Figure 7 Simplified model of transcriptional regulation of protein-encoding genes. (A) Scheme of transcriptional regulatory networks. The transcriptional initiation complex is regulated by transcription factors that are activated or repressed by environmental stimuli and/or developmental signals. The rectangular boxes labeled A, B and C represents the *cis*-acting factors and the ellipses labeled A, B and C represents the transcription factors. (B) Model of promoter showing core promoter, proximal promoter, and distal promoter regions. This model shows dynamic regulation of gene transcription, integrating the 3D structure of chromatin, epigenetic marks, and folding of the distal promoter sequences to become proximal to the transcriptional complex. Abbreviation: Ub, ubiquitin. Adapted from Yamaguchi-Shinozaki and Shinozaki (2005), and Hernandez-Garcia and Finer (2014).

Abiotic stress signals activate TFs by transcriptional induction of their genes, activation of proteins (such as phosphorylation), and degradation of proteins through the proteasome system (Yamaguchi-Shinozaki and Shinozaki, 2005). Once activated, the TFs bind to *cis*-regulatory elements present in promoters of OR genes. There are two major *cis*-regulatory elements, ABRE and DRE, that function, respectively, in ABA-dependent and ABA-independent gene expression under osmotic stress responses (Yamaguchi-Shinozaki and Shinozaki, 2005) (Figure 8).

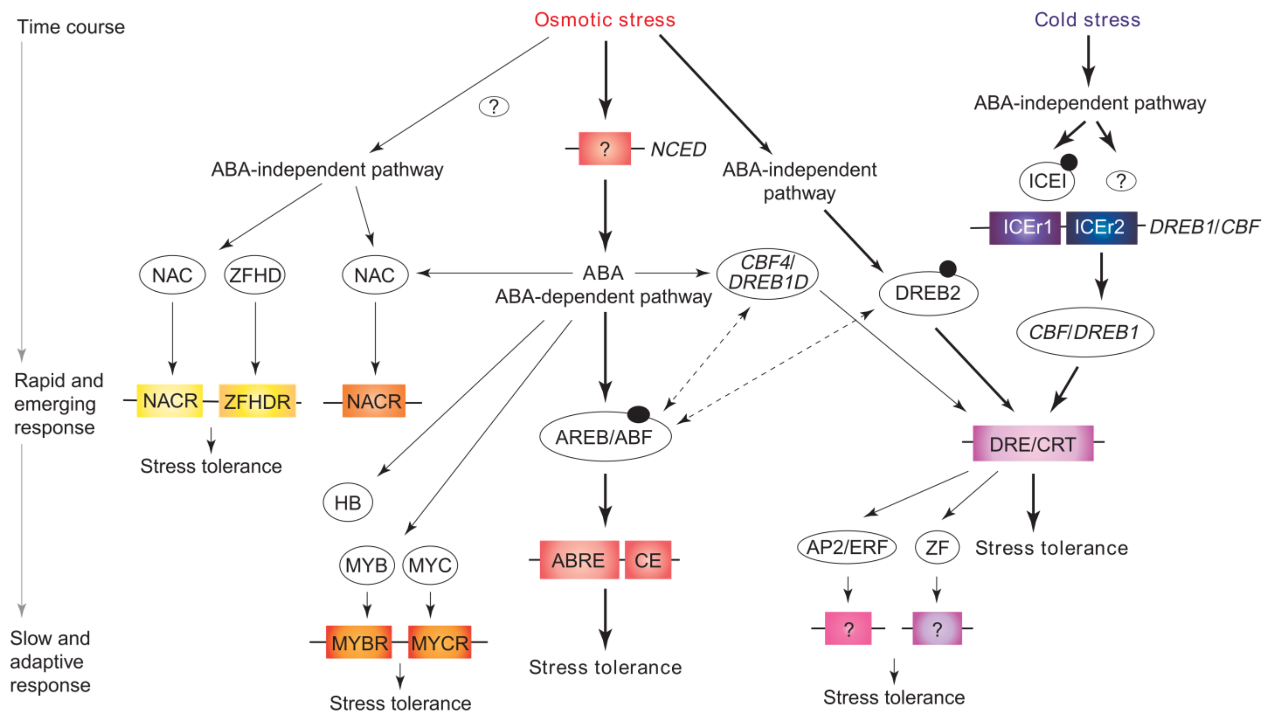


Figure 8 Regulatory networks of *cis*-acting elements and transcription factors involved in osmotic- and cold-stress responsive gene expression. Transcription factors controlling stress-inducible gene expression are depicted as ellipses. *Cis*-acting elements involved in stress responsive transcription are depicted as colored boxes. Small, black, filled circles reveal modification of transcription factors in response to stress signals for their activation, such as phosphorylation. Regulatory cascade of stress-responsive gene expression is shown from top to bottom. Early and emergency responses of gene expression are shown in the upper part, and late and adaptive responses in the lower part. Thick black arrows indicate the major signaling pathways; these pathways regulate many downstream genes. Broken arrows indicate protein–protein interactions. Abbreviations: ABA, abscisic acid; AREB, ABRE-binding proteins; ABRE, ABA-responsive element; CBF, C-repeat-binding factor; DRE/CRT, dehydration- responsive element/C-repeat; DREB, DRE-binding protein; ICE, Inducer of CBF Expression; MYBR, MYB recognition site; MYCR, MYC recognition site; NACR, NAC recognition site; ZFHDR, zinc-finger homeodomain recognition site.
(Source: Yamaguchi-Shinozaki and Shinozaki, 2005).

Analysis in promoter regions of ABA-inducible genes identified the conserved *cis*-regulatory sequence PyACGTGGC designated as abscisic acid-responsive element (ABRE). ABREs contain a core ACGT (also termed G-boxes) usually present in single/multiple copies in the upstream regions of the genes belonging to ABREB regulon (Yamaguchi-Shinozaki and Shinozaki, 2005; Mishra et al., 2014). The presence of single copy of ABRE is not sufficient for ABA-responsive transcription, either multiple copies of ABREs or a conjunction with another GC-rich sequence, such as coupling element (CE), is necessary (Roychoudhury et al., 2013; Mishra et al., 2014). The ABREs motifs are mostly recognized by the basic leucine zipper-transcription (bZIP) factors AREB or ABFs that after binding to the *cis*-regulatory sequences activates ABA-dependent gene expression (Figure 8).

The conserved sequence called the dehydration-responsive element (DRE) (TACCGACAT) regulates the response to cold and osmotic stress through the ABA- independent pathway (Figure 8). DRE is recognized by the DRE-binding protein (DREB) (Roychoudhury et al., 2013). A single copy of DRE is sufficient for ABA-independent stress-responsive gene expression, indicating that DRE does not require other elements for its function in stress-inducible gene expression, unlike ABRE (Yamaguchi-Shinozaki and Shinozaki, 1994). A striking similarity occurs between the DRE and others *cis*-regulatory sequences reported to regulate cold-inducible promoters, like C-repeat (CRT) and low-temperature-responsive element (LTRE) which share the DRE core sequence CCGAC. The presence of a common core sequence suggest a cross-talk and integration of cold and drought stimuli at the promoters of several downstream target genes (Yamaguchi-Shinozaki and Shinozaki, 1994; Haake et al., 2002). The secondary structure of the DNA sequences around *cis*-acting elements might be important for the binding of transcription factors (Yamaguchi-Shinozaki and Shinozaki, 2005). An assay evaluating the affinity of DREB proteins to degenerated DRE sequences showed that the binding characteristics for analogues of the optimum binding sequence differ between the DRE targeting proteins (Sakuma et al., 2002). The DREB1A and DREB2A proteins recognize and bind to the 6 bases optimum DRE sequence: A/GCCGAC. However, DREB1A can also weakly bind to CCCGAC, TCCGAC, and ATCGAC alternative DNA boxes. In contrast, DREB2A cannot bind to these sequences at all. Such slight differences in the recognition sequences between DREB1A and DREB2A TFs seem to influence the expression profile of each DREB target gene by cold, drought, and high salt

stress. This indicates that the surrounding sequences of the conserved motif are also important for specific binding of the transcription factors.

Multiple *cis*-regulatory elements are necessary for expressing a gene under stress conditions. Concomitant analysis of microarray data, localization and positioning of ABRE and DRE *cis*-motifs from 155 TFs in *Arabidopsis* revealed that the combinatorial relationship of these *cis*-regulatory elements in promoter regions of TFs may affect the gene expression in response to osmotic stress and ABA treatment (Mishra et al., 2014). Out of 155 TFs, 44 transcription factors highly expressed in response to osmotic stress and ABA treatment (upregulated by more than twofold) were checked for gap distance between ABRE and DRE motifs and orientation thereof. Approximately 60% of them presented one (ABRE or DRE) in direct and the other one in reverse orientation with a gap distance of less than 400 bp between them (Mishra et al., 2014) (Figure 9).

Besides these major path ways, many other transcriptional regulatory systems are involved in stress-responsive gene expression in plants. MYC/MYB transcription factors participate in the ABA-dependent pathway for the upregulation of the abiotic stress responsive genes (Agarwal and Jha, 2010a). Different MYC/MYB proteins bind to different *cis*-elements in their target gene's promoter. The plant MYB proteins can bind to several MYB recognition sequences (MYBRS) like MBSI (T/CAACG/TGA/C/TA/C/T), MBSII (TAACTAAC), MYB1AT (A/TAACCA) and others (Agarwal and Jha, 2010a). The core DNA sequence recognized by the MYC proteins is a consensus hexanucleotide sequence (CANNTG) known as the MYC recognition sequence (MYCRS). There are different types of MYCRS, and one of the most common ones is CACGTG. Certain variants of MYCRS can offer higher bind affinity to MYC proteins, as CACATG and CAAATG are the preferable target sequence of the inducer of CBF expression 1 (ICE1) protein.

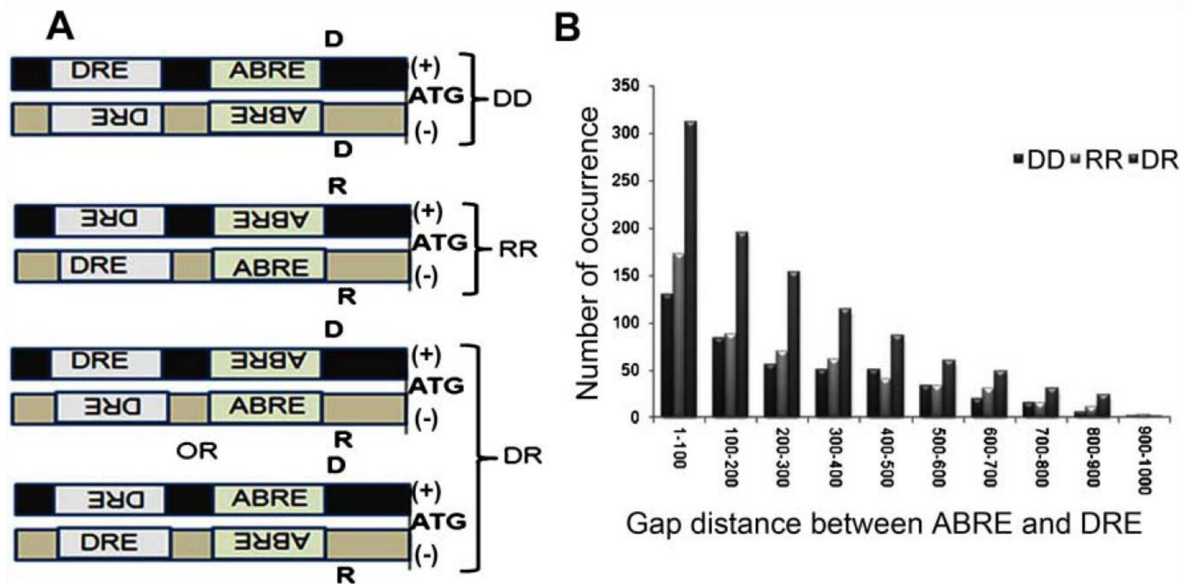


Figure 9 Gap length, distribution, and orientation pattern of DRE and ABRE cis-motifs. (A) The predicted DRE and ABRE were differentiated into D (direct), that is (+) strand, R(reverse), that is (-) strand, and DR or RD, that is, +/- or -/+ strand. The total number of all three possible orientations, that is, DD, RR, and DR, were counted and plotted with respect to total occurrence. (B) In the promoter regions of *Arabidopsis* genes the most preferable form was found to be DR form, with 72% of predictions with a gap length of ≤ 400 between DRE and ABRE motifs. Approximately 60% of genes with twofold change in expression under drought stress (osmotic and ABA treatments) were of DR form as a prevalent form of orientation with a gap distance of less than 400 bp between DRE and ABRE motifs.
(Source: Mishra et al., 2014)

1.4. AP2/ERF family of transcription factors

The APETALA2/Ethylene Responsive Factor (AP2/ERF) family is a large group of plant-specific transcription factors containing AP2/ERF-type DNA binding domains of approximately 60 amino acids. The AP2/ERF domain was firstly described in the *Arabidopsis* homeotic gene APETALA 2 (Jofuku, 1994), and a similar domain was found in tobacco (*Nicotiana tabacum*) ethylene-responsive element binding proteins (EREBPs) (Ohme-Takagi and Shinshi, 1995). Structural determination by nuclear magnetic resonance (NMR) of the *Arabidopsis* ERF1 protein in complex with a DNA molecule revealed that the AP2/ERF domain contains an N-terminal three-strand β-sheet that recognizes the target DNA sequence and a C-terminal α-helix (Allen et al., 1998). Based on the overall structure of AP2/ERF domains and the additional B3 DNA-binding domain, the AP2/ERF family can be divided into four family groups: AP2, Related-to-ABI3/VP1 (RAV), ERF and DREB (Sakuma et

al., 2002). Members of the AP2 subfamily contain double AP2/ERF domains, RAV subfamily harbor an AP2/ERF domain and additionally a B3 domain, while ERF and DREB present only a single AP2/ERF domain (Sakuma et al., 2002). In *Arabidopsis*, DREB and ERF subfamily were both divided into six subgroups, subgroup A1 to A6 and B1 to B6, respectively (Sakuma et al., 2002). An reviewed classification of AP2/ERF family considering gene structure and proteins motif structural analysis propose the union of DREB and ERF subfamilies in only one family, ERF, which can be further divided in twelve groups, I to X, VI-L and Xb-L (Nakano et al., 2006). Nakano et al (2006) investigated the conserved motifs outside of the AP2/ERF Domain in all members of each group in ERF family. Several motifs were described and based on these analyses each group was further divided into several distinct subgroups. Sequences both N-terminal and C-terminal of the AP2 domain are important short signature for the binding affinity to the *cis*-regulatory element (Canella et al., 2010).

1.4.1.DREB transcription factors

The DREB proteins are important transcription factors that induce a set of abiotic stress-related genes and impart stress tolerance to plants. DREB TFs have two conserved functional amino acids, valine (14) and glutamic acid (19), in the AP2/ERF domain that play crucial role in the binding specificity of DREBs to DRE core (Liu et al., 1998; Sakuma et al., 2002). Members of the DREB and ERF subfamilies recognize similar, but slightly different sequences. For example, DREB/CBF TFs has highest affinity for the DRE sequence (A/GCCGAC), in turn, ERF TFs has greatest affinity for the GCC-box sequence (AGCCGCC). Amino acid alignment of different DREB proteins showed a high sequence similarity in the nuclear localization signal (NLS) at the alkaline N-terminal region and in the Ser/Thr-rich region adjacent to the ERF/AP2 domain responsible for phosphorylation of DREB proteins (Liu et al., 1998). The C-terminal acidic domain functions in trans-activation activity (Stockinger et al., 1997). The DREB subfamily has 56/57 members divided into groups A1 to A6 (Sakuma et al., 2002)/I to IV (Nakano et al., 2006) (Figure 10A). Among them, DREB1 and DREB2

are the two most important subgroups and integrate two different signal transduction pathways under cold and dehydration respectively.

DREB1/CBF proteins consists of six members in *Arabidopsis*: DREB1A/CBF3; DREB1B/CBF1; DREB1C/CBF2; DREB1D/CBF4; DREB1E/DDF2; DREB1F/DDF1 (Sakuma et al., 2002). As part of its intrinsic characteristics, the DREB1s/CBFs subgroup has two highly conserved regions on both side of the AP2/ERF domain (CMIII-3) and two consensus motif in the C-terminal region (CMIII-2 and CMIII-4) (Sakuma et al., 2002; Nakano et al., 2006; Canella et al., 2010). Compose the CMIII-3 motif a NLS-type consensus PKK/RPAGRxKFxETRHP and a DSAWR motif at the N- and C-terminal side of the AP2/ERF domain, respectively. Homology comparison of the protein sequence revealed the conservation of these motif in several plants species (Jaglo et al., 2001; Agarwal et al., 2006b) (Figure 10B).

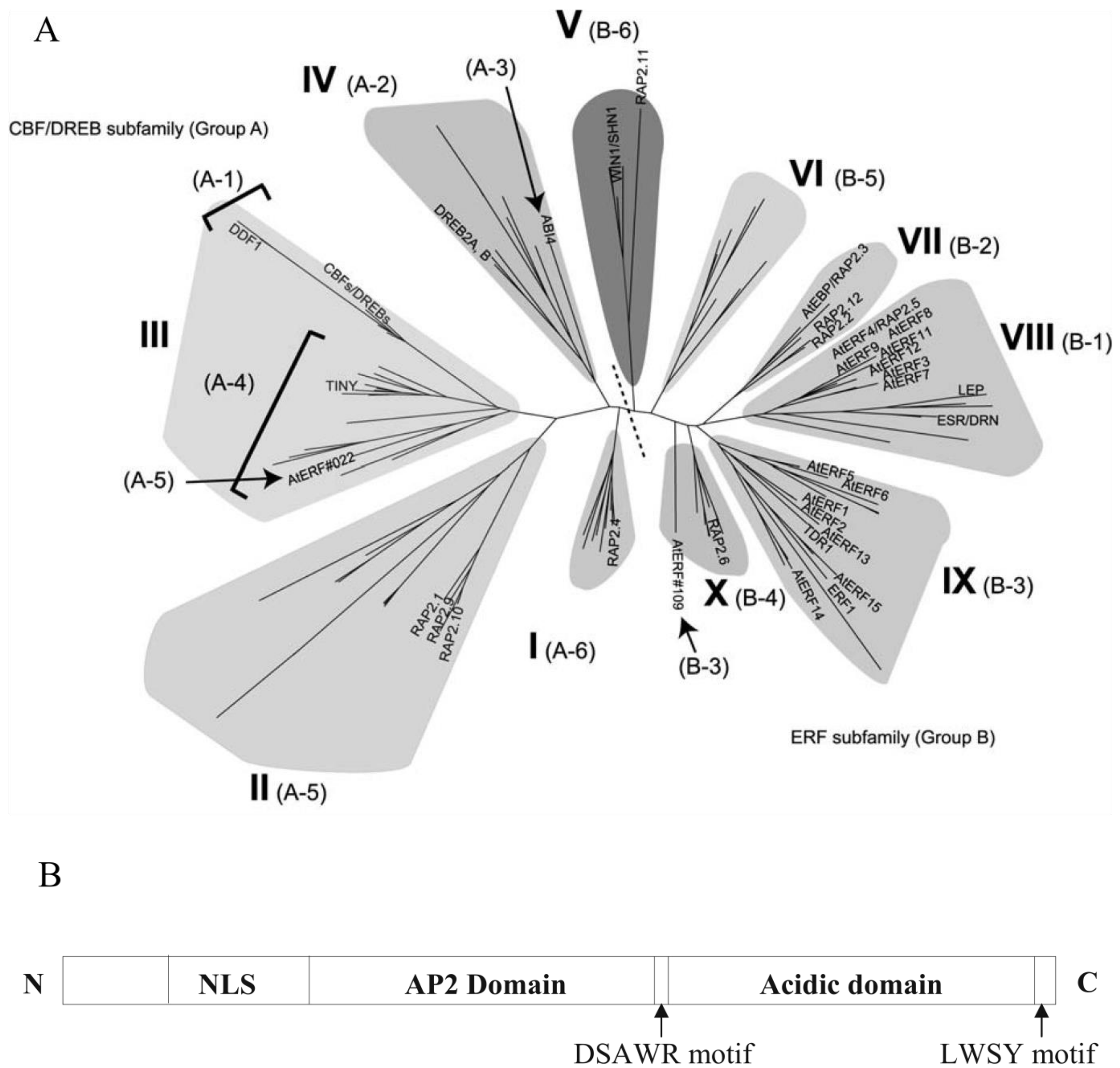


Figure 10 Phylogenetic tree of *Arabidopsis* ERF proteins, and structure and domain of DREB1/CBF transcription factors. (A) An unrooted phylogenetic tree of *Arabidopsis* ERF proteins according to Nakano et al (2006) classification. The names of some already reported ERF genes are indicated. The broken line divides the CBF/DREB and ERF subfamilies. Classification by Sakuma et al. (2002) is indicated in parentheses. (B) Schematic structure and domains of DREB1/CBF transcription factors. The entry of DREB proteins in the nucleus is mediated by one or two nuclear localization signal (NLS), in DREB1/CBF proteins NLS have as consensus sequence PKRPAGRTKFRETRHP. Other motifs intrinsic of DREB1/CBF members and are distinguished from the other DREB subgroups, DSAW motif at the end of the AP2/ERF domain and LWSY motif at the end of the C-terminal region The carboxyl-terminal acidic region is supposed to be a transcriptional activator motif. Adapted from Nakano et al (2005) and Akhtar et al (2012).

The DREB2 subfamily is induced by drought and high-salinity stress indicating their important role in stress-responsive gene expression. Eight proteins compose the DREB2 subgroup in *Arabidopsis*, DREB2A to H. Like DREB1s, a high homology can be found throughout the N-terminal region outside the AP2/ERF domain in the protein sequences of DREB2 members. This conserved region includes two motifs, CMIV-1 and CMIV-2, being the latest a putative NLS (Nakano et al., 2006).

Other DREB subfamily members have been reported to be stress-responsive in transgenic plants. The ABI4, a A-3 subgroup member, can regulates the expression of more than a hundred ABA-dependent genes and present a synergistic effect with several ABA-responsive bZIP TFs (Reeves et al., 2011). Although A-3 members are related to DREB2s, they have specialized functions and recognize a target sequence different to DRE, suggesting that they use a pathway different from the DREB2 in stress response (Niu et al., 2002; Nakano et al., 2006; Koussevitzky et al., 2007; Matsukura et al., 2010). The A-4 proteins are related to and share conserved motifs to DREB1s/CBFs (Nakano et al., 2006). Among them, *TINY* is capable of binding to both DRE and ERE with similar affinity and can activate the expression of reporter genes driven by either pathways.

The RELATED TO AP2.1 (RAP2.1) gene of A-5 subgroup act as a transcriptional repressor to keep cold and drought stress responses under tight control, among its downstream targets are the Cold-regulated15A (COR15A) and RD29A (Dong and Liu, 2010). Thus the A-5 subgroup proteins may function in negative feedback regulation of the DREB1/CBF and DREB2 pathways. The ABA-responsive ERF subfamily member RAP2.6 and RAP2.6L play a dual role in abiotic and biotic stress responses. The overexpression of RAP2.6L in *Arabidopsis* enhance salt and drought tolerance in transgenic plants, as well vigor and rooting in seedlings compared to WT without affecting the phenotype (Krishnaswamy et al., 2011). Representing the A-6 subgroup, RAP2.4 and RAP2.4B respond to cold and heat shocks, respectively, additionally they are both expressed in response to dehydration, high salinity and heat (Lin et al., 2008; Rae et al., 2011). The RAP2.4A integrates the redox-response pathway and binds to a C3-like regulatory sequence (Shaikhali et al., 2008). Thus, it is possible that A-6 proteins have different target genes from those of DREB1s/CBFs and DREB2s.

1.4.2.Expression pattern and regulation of DREB1/CBF subgroup in abiotic stress response

Differences in abiotic stress tolerance between ecotypes or cultivars are correlated with the expression levels of DREB1/CBF genes. Therefore, regulation of their expression level is a major regulatory mechanism for DREB1/CBF activity (Mizoi et al., 2012). The DREB1/CBF genes are rapidly and transiently induced most often by cold stress. The overexpression of DREB1s/CBFs genes in transgenic plants revealed an up regulation of more than 40 cold-regulated (COR) genes (Fowler and Thomashow, 2002; Seki et al., 2002). The COR genes encode cold-inducible proteins that function in survival at low temperatures, including LEA proteins (i.e., hydrophilic proteins abundantly expressed during seed maturation and in response to cold or dehydration) and enzymes for sugar metabolism and fatty acid desaturation (Heidarvand and Maali Amiri, 2010). Among these, DREB1A/CBF3, DREB1B/CBF1 and DREB1C/CBF2 TFs are rapidly induced in cold acclimation (Stockinger et al., 1997; Shinwari et al., 1998). DREB1E/DDF2 and DREB1F/DDF1 are up-regulated by salinity stress, and DREB1F/DDF1 directly upregulated the expression of the gibberellin-deactivating gene. Distinctively to others DREB1/CBF genes, DREB1D/CBF4 is responsive to drought stress and ABA, but not to cold stress (Haake et al., 2002; Mizoi et al., 2012). Overexpression of CBF4 in *Arabidopsis* resulted in constitutive expression of CRT/DRE containing stress-responsive genes and enhanced tolerance to drought and freezing stresses (Haake et al., 2002).

A considerable progress has been made in the past decade in elucidating the DREB1/CBF regulatory network and several of its components has been identified in *Arabidopsis*. TFs have been reported to interact directly with the promoters of the *DREB1/CBF* genes and activate their expression upon stress. ICE1, a MYC-type basic helix– loop–helix (bHLH) transcription factor, has been described to bind to MYC recognition elements in the *CBF3* promoter and activate the expression of *CBF3* during cold acclimation (Chinnusamy et al., 2003) (Figure 11). In addition, constitutive overexpression of *ICE1* also enhanced the expression of CBF2 and COR genes during cold acclimation, and increased freezing tolerance of the transgenic *Arabidopsis*. Interestingly, *ICE1* is constitutively expressed and localized in the nucleus, but it induces expression of *CBFs* only under cold stress suggesting that cold

stress-induced post-translational modifications of ICE1 is necessary to activate downstream genes (Chinnusamy et al., 2003). Indeed, cold stress induces phosphorylation and sumoylation of ICE1. Mutation of a potential phosphorylation site, Serine 403, increased transactivational activity and inhibited cold-induced degradation of ICE1 (Miura et al., 2011). ICE1 phosphorylation is mediated by kinases which are activated following the increase of $[Ca^{2+}]$cyt (Figure 11). The SUMO E3 ligase, SIZ1, promotes the sumoylation of ICE1 during cold stress enhancing its stability and activity, and at the same time reduces the polyubiquitination of ICE1. In contrast, RING-type E3 High Expression of Osmotically Responsive Gene 1 (HOS1) is a mediator of ICE1 ubiquitination and subsequent degradation, thus regulating negatively ICE1 (Dong et al., 2006). ICE2, an ICE1 homologue, is involved in regulating the cold induction of CBF1 (Fursova et al., 2009). An R2R3-MYB transcription factor, MYB15, expressed even in the absence of cold stress, was shown to negatively regulate the cold induction of the *CBFs* (*CBF1*, *CBF2* and *CBF3*) genes in *Arabidopsis* by bind to MYB recognition elements located in their promoters (Agarwal et al., 2006a). Another transcription factor that also appears to function as a negative regulator of CBFs is the zinc finger, ZAT12. The overexpression of ZAT12 in transgenic *Arabidopsis* decreases the expression of CBFs under cold stress (Vogel et al., 2005). Different members of calmodulin-binding transcription activators (CAMTAs) have been uncovered that bind to the *CBF2* promoter inducing the expression of *CBF2* (Doherty et al., 2009). Finally, the bHLH factor PIF7 has been found to bind to the G-box element present in the *CBF2* promoter and to function as a repressor in transient assays (Kidokoro et al., 2009).

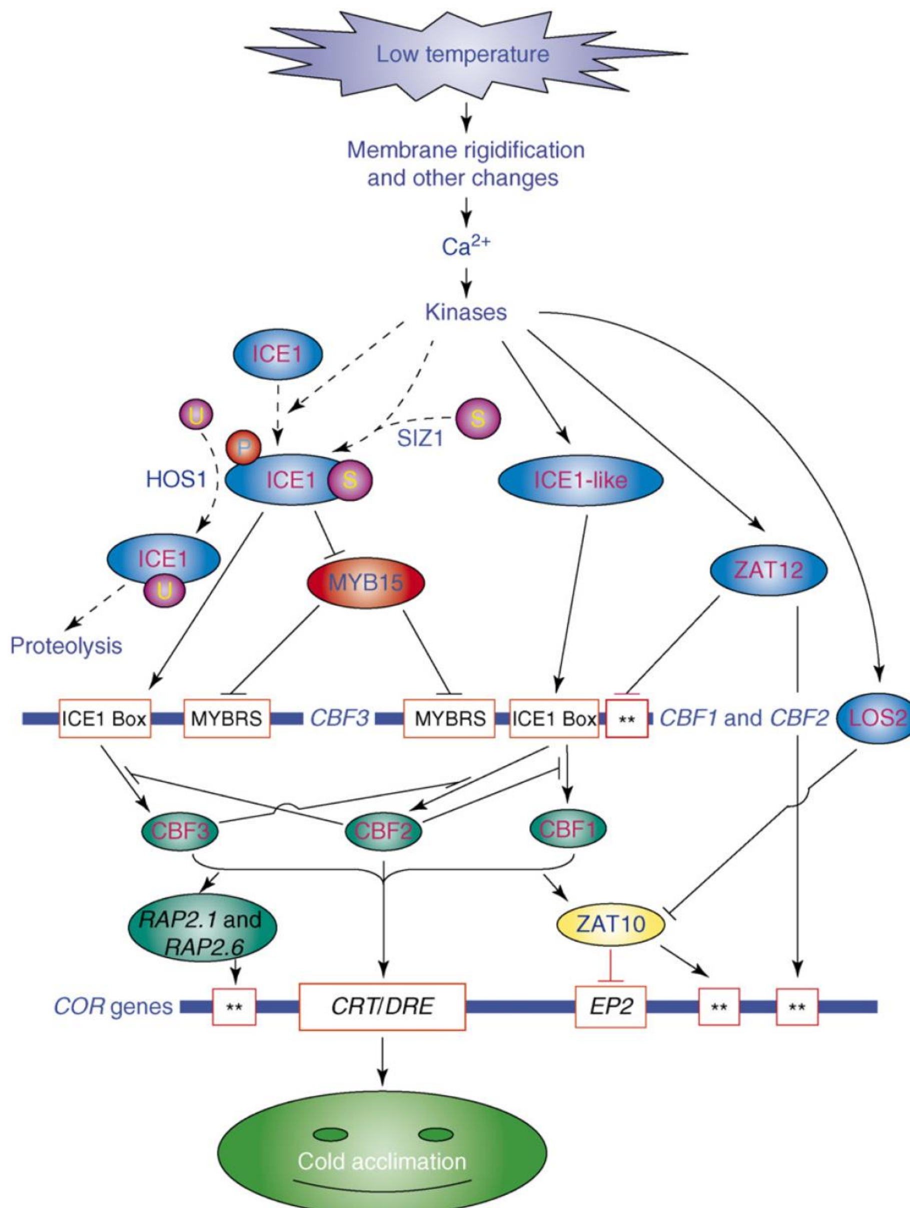


Figure 11 Diagram of cold-responsive transcriptional network in *Arabidopsis*. Plants probably sense low temperatures through membrane rigidification and/or other cellular changes, which might induce a calcium signature and activate protein kinases necessary for cold acclimation. Constitutively expressed ICE1 is activated by cold stress through phosphorylation and sumoylation, the latter is critical for ICE1-activation of transcription of CBFs and repression of MYB15. The expression of CBFs is negatively regulated by MYB15 and ZAT12. HOS1 mediates the ubiquitination and proteosomal degradation of ICE1 and, thus, negatively regulates CBF regulons. CBFs induce the expression of ZAT10 (=STZ), which might downregulate the expression of COR genes. Cold-upregulated LOS2 represses the transcription of ZAT10. ZAT10 and ZAT12 are two C2H2 zinc finger transcription factors. Broken arrows indicate post-translational regulation; solid arrows indicate activation, whereas lines ending with a bar show negative regulation; the two stars (**) indicate unknown *cis*-elements. Abbreviations: CBF, C-repeat binding factor (an AP2-type transcription factor); CRT, C-repeat elements; DRE, dehydration-responsive elements; HOS1, high expression of osmotically responsive genes1 (a RING finger ubiquitin E3 ligase); ICE1, inducer of CBF expression 1 (a MYC-type bHLH transcription factor); LOS2, low expression of osmotically responsive genes 2 (a bifunctional enolase with transcriptional repression activity); MYB, myeloblastosis; MYBRS, MYB transcription factor recognition sequence; SIZ1, SAP and MiZ1 (a SUMO E3 ligase); P, phosphorylation; S, SUMO (small ubiquitin-related modifier); U, ubiquitin.
Source: Chinnusamy et al., 2007

1.4.3.Homologous genes and function prediction

As previously detailed, *DREB1/CBFs* genes from different plant species have been showing a conservative pattern of stress response. It is well known that expressions of the A-1 group of genes is induced by low temperature, but not by drought or high-salt stress, while A-2 group genes are regulated by salt and drought, but not by cold (Liu et al., 1998; Dubouzet et al., 2003). However, some reports have shown conflicts with respect to these trends. For example, number of *DREB1/CBFs* genes such as *BrCBF* from *Chinese cabbage* (Jiang et al., 1996), *MbDREB1* from apple (Yang et al., 2011), *OsDREB1F* from rice (Wang et al., 2008), *VviDREB1* from *Vaccinium vitis-idaea* (Wang et al., 2010) also respond to high-salt stress, drought and exogenous ABA treatment in addition to cold stress. The different abiotic stress signalling pathways are assumed to interact and share some common elements that formed as potential 'node' for crosstalk (Haake et al., 2002; Nakashima et al., 2009; Lata and Prasad, 2011). These *DREB1/CBF* genes may act as cross-point or node connecting several pathways and simultaneously regulate cold, salt, drought and ABA pathways. Interestingly, *OsDREB1B* showed high expression in cold stress as expected, but was also induced by high temperature (Qin et al., 2004). Different expression profiles are also reported within the same *CBF* family. For example, in *Vitis sp.*, the *CBF4* gene was mostly induced by cold while the *CBF1*, *CBF2* and *CBF3* genes showed better response to drought compared to cold (Xiao et al., 2008). Majority of CBF-encoding genes from different plant species are reported to be significantly upregulated in response to cold stress and perform the important function of cold adaptation. However some genes such as *CrCBF* from *Catharanthus roseus* and *OsDREB1C* from *Oryza sativa* are found to be constitutively expressed under cold stress (Dubouzet et al., 2003; Dutta et al., 2007). The common features of most *CBFs* are quick and durable response. *AtDREB1A* is induced within 10 min at 4◦C and transcript levels of *EguCBF* and *MbDREB1* are detectable within 15 min and 30 min, respectively, under cold stress condition (Liu et al., 1998; Navarro et al., 2011; Yang et al., 2011). *EguCBF1*, *OsDREB1A/CBF3* and *ZmDREB1A* are induced by cold within a period of 30min, 40min and 60min, respectively, and remain detectable even at 24 h after exposure to cold stress (Dubouzet et al. 2003; Qin et al. 2004; Navarro et

al. 2009). The different role that can be taken by *DREB1/CBFs* homolog gene in different species evidence the natural plasticity developed by plants to cope with stress.

1.5. Natural diversity and genetic variability in coffee

Genetic diversity of *Coffea* species has been elucidated based on DNA sequence data from plastid and nuclear regions (Lashermes et al., 1997; Davis et al., 2007; Anthony et al., 2010; Davis et al., 2011; Nowak et al., 2012) and refined by the use of microsatellites (Montagnon et al., 1992; Lashermes et al., 1999; Montagnon, 2000; Moncada and McCouch, 2004; Poncet et al., 2007; Cubry et al., 2008; Cubry et al., 2012; Montagnon et al., 2012; Razafinarivo et al., 2013). The results suggest a recent origin in Africa with a radial dispersion leading to speciation of the genus *Coffea*. In the most recent phylogenetic study, six main well-supported lineages with strong geographical correspondence were retrieved and recognized: (1) African *'Psilanthus'* clade (i.e. short-styled African *Coffea*); (2) Asian and Australasian *'Psilanthus'* clade (short-styled non-African *Coffea*); (3) the Lower Guinea/Congolian (LG/C) clade [species from West and Central Africa, west of the Great Rift Valley, with some species (*C. canephora* and *C. liberica*) also occurring in the Upper Guinea Region]; (4) the Upper Guinea (UG) clade (three Upper Guinea endemics); the East-Central Africa (E-CAfr) clade [species straddling the Great Rift Valley but with one species (*C. anthonyi*) in West and Central Africa]; (5) the Indian Ocean (Madagascar, Comoros and Mascarenes) (IO) clade, which includes the Mascarene (MAS) clade and the Madagascar (MAD) clade; (6) the dry-adapted Madagascan bara*Coffea* alliance; an East African (EA) clade was also consistently but with inconsistent levels of support (Figure 12).

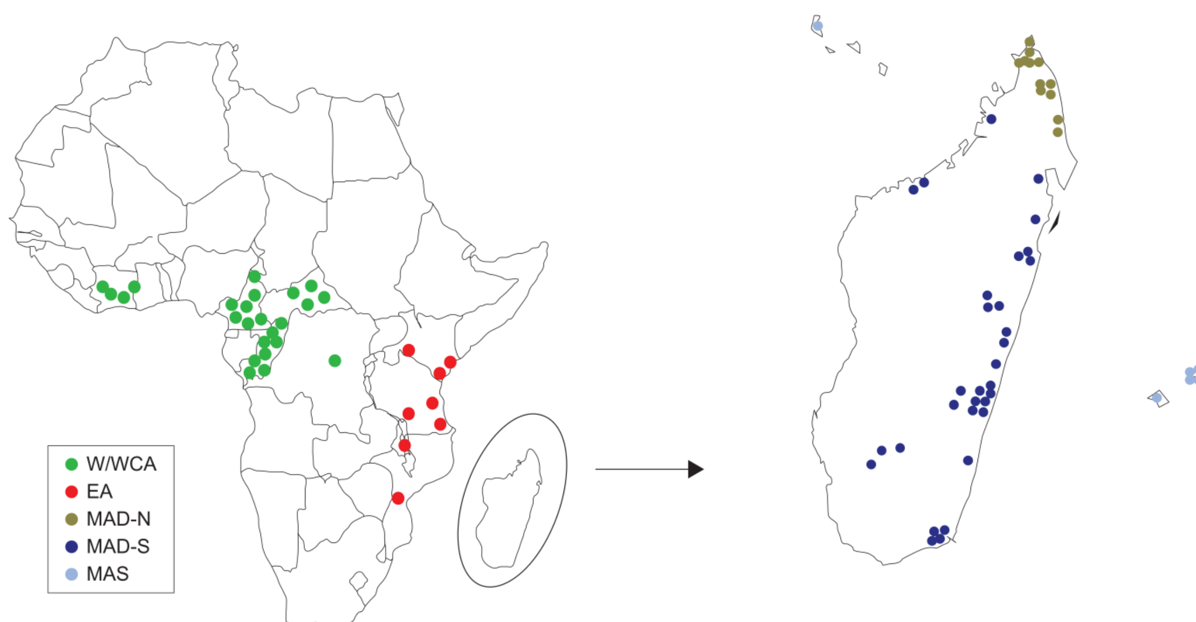


Figure 12 Location of African populations with strong geographical correspondence. Abbreviations: W/WCA, West and West-Central Africa; EA, East Africa; MAD-N, Madagascar north; MAD-S, Madagascar south; and MAS, the Mascarenes, and the Comoros Islands.
(Source: Razafinarivo et al., 2013).

1.5.1.Structuration of *Coffea* diversity

1.5.2. *Coffea arabica*

C. arabica primary center of origin was the highlands of Southwest Ethiopia and the Boma Plateau of Sudan, however, wild populations have also been reported in Sudan and Kenya (Thomas, 1942; Anthony et al., 1987). *C. arabica* was firstly cultivated in Southwest Ethiopia about 1,500 years ago. Historically, the genetic inheritance of modern *C. arabica* cultivars arise from two main populations, known as Typica and Bourbon, that were spread worldwide in the 18th century (Anthony et al., 2002). This is the case of the most commercial *C. arabica* cultivars, Mundo Novo, Catuai and Caturra. The Caturra cultivar is a dwarf mutant of the Bourbon group. Mundo Novo is a hybrid between Bourbon and Typica, while Catuai cultivar derived from a cross between Mundo Novo and Caturra (Figure 13). Each cultivar displays distinctive plant architecture and physiological properties.

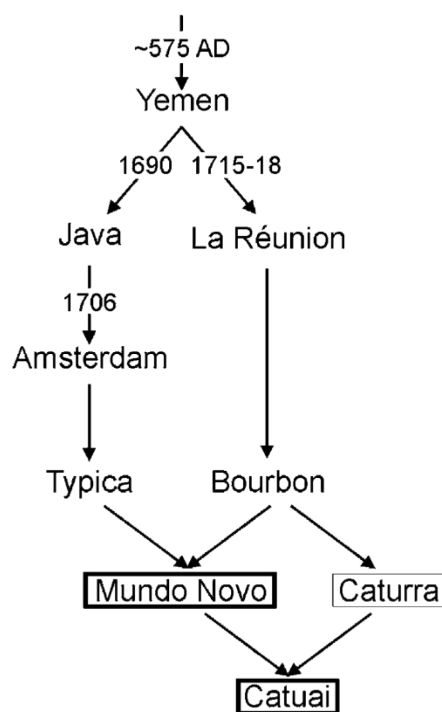


Figure 13. Origin of cultivated cultivars of *C. arabica*. The most commercial *C. arabica* cultivars, including Caturra, Mundo Novo and Catuai, were selected from only two base populations: Bourbon and Typica.
(Source: Vidal et al., 2010).

C. arabica narrow genetic diversity as result of its origin and reproductive behavior (autogamous) encouraged the search of new cultivars with improved traits in breeding programs, such as flowering time synchronicity, bean size, beverage (cup) quality, caffeine content, resistance to pests, and drought stress tolerance.

1.5.3. *Coffea canephora*

Therefore, during the 19th century in Africa, the interest to cultivate locally spontaneous forms of other species of coffee, like *C. canephora* increased. In particular, coffee plants from local forest populations Belgian Congo (now known as Democratic Republic of Congo) and Uganda were transferred to Java, a major breeding center from 1900 to 1930. At the same time, in Africa, the diversity of material cultivated was extended using local spontaneous forms: Conilon in Côte d'Ivoire, Niaouli in Togo and Benin, and Nana in the Central African Republic. The improved coffee

accessions selected in Java were reintroduced in the Belgian Congo around 1916 at INEAC (Institut National pour l'Etude Agronomique au Congo) research center which was the major breeding center of *C. canephora* from 1930 to 1960. Further, these accessions were largely distributed worldwide. Breeding programs have led to significant increase in the overall performance of coffee cultivated trees for the past century, howsoever a great genetic similarity has been maintained to individuals of the original natural populations (Razafinarivo et al., 2013). *C. canephora* genetic diversity can be divided in two major clades according to their geographical origins: the Guinean group (G) and the Congolese group. The Congolese group can be subdivided into SG2/B, C, SG1 and UW (Montagnon and Leroy, 1993) (Figure 14). Guinean genotypes are considered the most tolerant to drought and genotypes from the SG1 Congolese group are more tolerant to drought than those from the SG2 Congolese group (Montagnon and Leroy, 1993). The considerable genetic diversity observed in *C. canephora* is still largely unexploited. During the last decade, several breeding programs to development of new *C. canephora* clones have attempted to explore the genetic diversity of *C. canephora.* In Brazil, a genetic improvement program for the development of new cultivars, using SG1 genotypes as source of genetic variability, characterized a clonal variety of *Coffea canephora* var. Conilon highly productive under drought conditions (Ferrão et al., 2000).

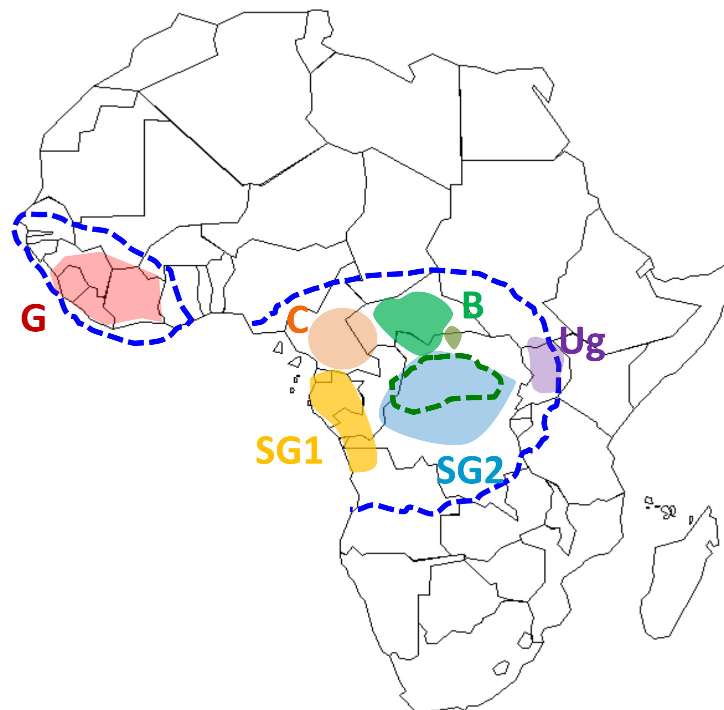


Figure 14 Geographic origins of the main genetic groups of *C. canephora*. Geographical origin of Guinean pool, red circle. Geographical origin of each genetic group of the Congolese pool: orange, C group; green, B group; purple, Uganda group; blue, SG2 group; and yellow SG1 group. Blue dashed lines indicate regions with short dry season and green dashed lines regions without dry season. Adapted from Montagnon et al. (2012).

1.5.4.Single Nucleotide Polymorphisms (SNP) as new molecular markers

DNA molecular markers have gained popularity in the last decade. Although a multiplicity of sequence polymorphism detection assays have been used for genetic analysis in plant species, single-nucleotide polymorphisms (SNPs) provide the ideal genotyping system (Kaur et al., 2012). SNPs rapidly became the markers of choice in genetic studies because of their high frequency in the genome and low mutation rates compared with microsatellites (Rafalski, 2002). Advances in sequence generation technology enabled highly automation to SNP detection and make the use of SNP and InDels (Insertion-Deletion length polymorphism) markers attractive for marker-assisted breeding, expressed sequence tags (EST) mapping and integration in high-density genetic map to genome wide association studies (GWS) (Rafalski, 2002). SNPs represent the most common type of sequence differences between alleles and may be identified in the vicinity of virtually every gene. Therefore, the use of SNPs in the detection of associations between gene allelic forms and phenotypes has a great

potential, especially for multifactorial traits where high-resolution discrimination of alleles is required. Moreover, it is suggested the analysis of SNP haplotypes, rather than of individual SNPs, provides a more effective way of associating alleles with traits, once, large islands of high LD extend over much larger distances than predicted previously (Goldstein, 2001). "SNPs haplotype" are a cluster of SNPs that are genetic linked and are inheritance together from one or multiple loci (Figure 15).

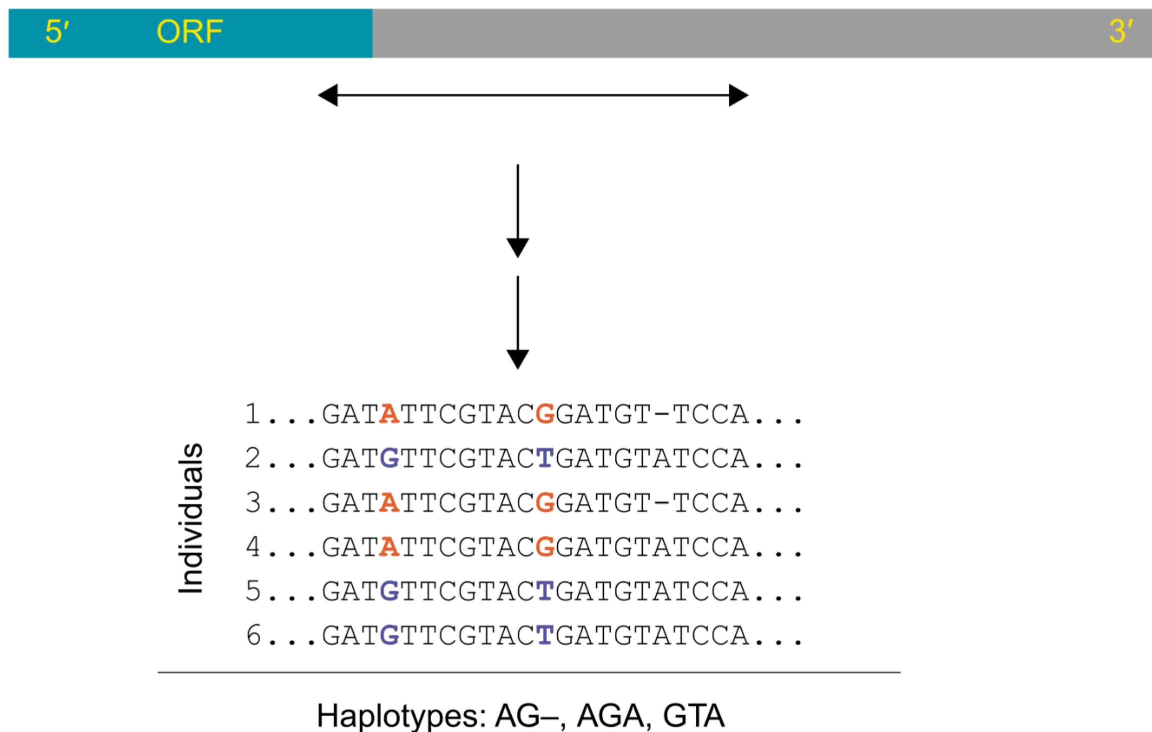


Figure 15 Schematic representation of SNPs haplotype in a hypothetical EST-related sequence. After PCR amplification from a set of genetically diverse individuals, the PCR products are cloned and sequenced. The resulting sequences are aligned and SNPs and InDel polymorphisms are identified. In this hypothetical example, three haplotypes could be distinguished. ORF, open reading frame. Adapted from Rafalski (2002).

1.5.5. Intra- e inter-genomic diversity

Methodologies for discovery and validation of predicted SNP variants have been optimized for both inbreeding and out-breeding diploid plant species (Cogan et al., 2006; Chagné et al., 2007). However, polyploidy is extremely common in plant lineages, accounting for 70% of living angiosperm species, including the majority of important crops. Polyploidy increase levels of genome complexity and bring additional challenges for SNP prediction, since sequence variation between subgenomes co-exists with allelic variation within subgenomes. (Kaur et al., 2012) Therefore, the comprehension of a number of biological factors, like the crop evolutionary context, is of great value for the resolution of this complexity. Polyploid species contain chromosomal complements in excess as a result of two conditions, either due to failures of homologous chromosome disjunction during meiotic prophase leading to reduplication of homologous chromosomes, autopolyploidy, or can result from the intergeneric or interspecific hybridization of related species with similar but not identical (homoeologous) chromosomes, allopolyploidy. *Coffea arabica* is a precise example of the latter condition. *C. arabica* is the only polyploid and self-compatible species in an essentially diploid and self-incompatible *Coffea* genus. The evolutionary history of *C. arabica* refers to a single polyploidization event from interspecific hybridization of two related diploid species, *C. canephora* and *C. eugenioides* (Figure 16). Therefore, allopolyploid *C. arabica* genome (2n=4x=44) is formed by two diploid subgenomes related to its progenitors, *C. canephora* (CaCc) and *C. eugenioides* (CaCe).

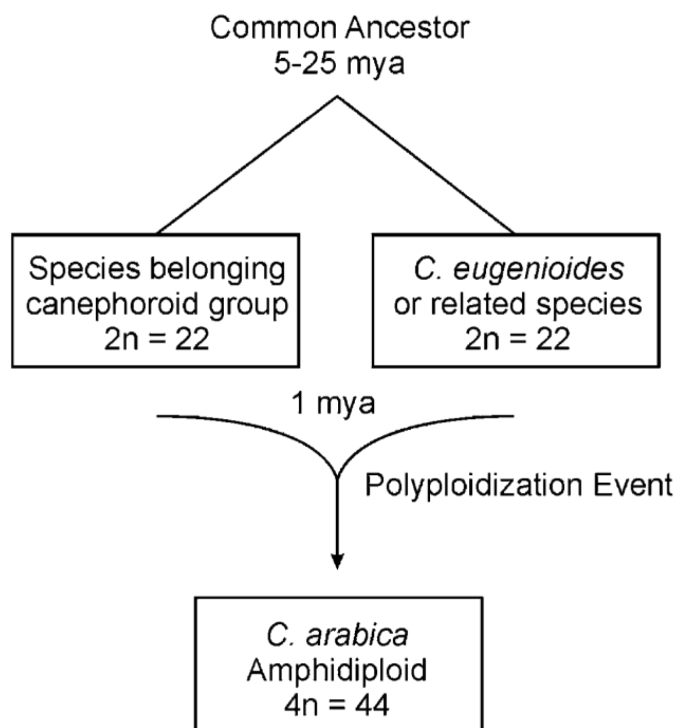


Figure 16 Evolutionary history of allotetraploid *C. arabica*. Origin of *C. arabica*. The progenitor genomes are represented by diploid *C. eugenioides* and *C. canephora*. *C. arabica* arose 1 to 2 million years ago (mya) from the fusion of *C. canephora* (or related species) and *C. eugenioides*.

With such complexity polyploid genomes harbor a series of nucleotide variants, for instance homologous single-nucleotide sequence variants (i.e. SNPs) can be found between chromosome pairs of either each subgenome (CaCc or CaCe), or between subgenome and its corresponding diploid progenitors (CaCc and Cc or CaCe and Ce), as well as homoeologous sequence variants (HSVs) can arise between genes that are present in both subgenomes (CaCc and CaCe) (Somers et al., 2003). Thus, the ideal method for SNP genotyping in an allopolyploid genetic system is to reduce complexity to the diploid level isolating subgenome-specific alleles. In this context, information on *C. arabica* progenitor's genome (Cc and Ce) may assist discrimination of nucleotide variant types through the ability to compare allopolyploid-derived (Ca) sequence haplotypes with those obtained from diploid counter-parts (Cc or Ce), permitting selective subtraction of homoeologous components (CaCc and CaCe) (Kaur et al., 2012). Assignment of HSVs and analyses of homeologous genes expression have revealed evidences of transcription profile divergence in *C. arabica* subgenomes (Eklund and Edqvist,

2003; Petitot et al., 2007; Vidal et al., 2010; Bardil et al., 2011; Marraccini et al., 2011; Mondego et al., 2011; Cotta et al., 2014). Indeed, the merging of divergent genomes in a single nucleus can lead to whole set of dynamic changes at genome, transcriptome and phenotype level to arising polyploid species (Doyle et al., 2008; Soltis et al., 2010). Vidal et al showed that *C. arabica* displays differential expression of homeologous genes and suggested that *C. arabica* ancestral subgenomes encode proteins to attend different physiological mechanisms. As a matter of fact, transcriptome divergence in *C. arabica*, in comparison with its two diploid parents, is modulated by the environment (Bardil et al., 2011). The plasticity of allopolyploid subgenome regulation allowed *C. arabica* to show a reduce divergence (9%) in transcriptome regulation to its related parental *C. canephora* at day-night temperature of 30-26°C compared to a 35% divergence at 22-26°C. Another example of dynamic changes resulting of polyploid formation is the genomic rearrangements involving homoeologous exchanges events (HEEs). In *C. arabica*, HEEs seems to be a major source of genetic diversity, evaluating five *C. arabica* accessions and the modern-day progenitors, *C. canephora* and *C. eugenioides*, it was inferred that at least 5% of the *C. arabica* genes have undergone homoeolog loss, and the HEEs specificity to one or a few accessions suggest a gradual accumulation of HEEs (Lashermes et al., 2014). Relative contributions of homeologues to locus specific expression have been used to increase the knowledge of subgenomes regulation in *C. arabica* transcriptome. Pierre *et al* (2011) highlighted the predominant expression of the homeologous CaCe form of small subunit ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) enconding gene over the CaCc form in the leaves of non-introgressed *C. arabica* cultivars. In the same work, the absence of Rubisco CaCc homeolog expression for the HT832/1 accession of Timor Hybrid was already postulated as HEEs. Moreover, differential expression of homeologous genes within coffee fruits were also described for *C. arabica* (Cotta et al., 2014). The exclusive expression of *nsLTP* CaCc homeolog form in the pericarp and co-expression of both CaCc and CaCe homeologs in the grain tissues (perisperm and endosperm) revealed for the first time spatiotemporal regulation of homeologous genes within different tissues of the same organ in *C. arabica*. Therefore, the identification of homeolocus-specific variants increments the correlation of allele and subgenome function with specific trait variation.

1.6. Associations between genetic diversity and adaptations

The recent advances in genome and transcriptome sequencing brought to daylight a remarkable variation in genes and their expression, however, a fundamental question remains, how this variation affects the function and fitness of whole organisms in populations (Mitchell-Olds et al., 2008). Thus, it is necessary to understand the functional basis of evolutionary forces shaping ecologically important traits in natural biological populations. This requires elucidating the biotic and abiotic challenges present in natural environments, the molecular and cellular mechanisms used to meet these challenges and the evolutionary forces and processes that influence these adaptations (Mitchell-Olds et al., 2008).

1.6.1.Molecular signatures of selection

Abiotic stresses such as drought, extreme temperatures, and salinity have a strong impact on plant adaptation acting as selective forces on plant physiology and morphology. These selective pressures leave characteristic footprints that can be detected at the DNA sequence level. One of the main interests in molecular population genetics is to distinguish molecular variation that is neutral, when a new mutant does not affect the fitness of the individual in which it arises, from variation that is subject to selection, when affect the fitness (Nielsen, 2005). Neutral theory presume that the majority of the polymorphisms seen within and among species are selectively neutral, or at least nearly so, what means that their evolutionary fate is not affect by selection. Neutrality makes mathematical modeling relatively easy, assuming neutrality means that genetic variability exists as a natural null model, and perturbations in this standard neutral model are consequences of phenomes such as natural selection, population structure (i.e. subdivision) and migration (Nordborg and Innan, 2002). When investigating molecular signatures of plant adaptation, we are concerned principally with two types of natural selection: positive selection, which represents the fixation of a favorable mutation, and balancing selection, which is the long-term selective maintenance of multiple alleles (Wright and Gaut, 2005). Both selection produce specific patterns of nucleotide diversity, which can be detected as departures from the expected pattern of natural null model. Statistical models design to test the neutrality of

evolution models and are so called neutrality test. Several neutrality tests can be used to infer, with reasonable considerations, the action of selection based on different features of sequence data.

In population genetic data, selection may affect levels of variability, allelic distribution in each nucleotide site (frequency spectrum), linkage disequilibrium and haplotype structure. The first feature is the amount of diversity or variability: strong positive selection causes a reduction in levels of nucleotide diversity (Smith and Haigh, 2007), whereas long-term balancing selection elevates diversity (Kaplan et al., 1988). However to reject neutral model based on diversity it is necessary to have explicit information on mutation rates, what generally requires multilocus data to detect diversity effects (Wright and Gaut, 2005). The second feature of sequence data is the frequency distribution of polymorphisms. In other words, it represents a summary of the allele frequencies of the various mutations. Selection deviate the population frequency of alleles from the neutral model expectations. Tajima's D statistic is one of several tests that measure this frequency spectrum (Tajima, 1989). In this test, the average number of nucleotide differences between pairs of sequences is compared with the total number of segregating sites (SNPs). If the difference between these two measures of variability is larger than what is expected on the standard neutral model, this model is rejected. Under neutral equilibrium, the mean Tajima's D statistic is expected to be zero. In a positive selection scenario, the frequency spectrum is skewed, and is expected an excess of rare polymorphisms as new variants accumulate after a selective sweep. This excess of rare variants relative to neutral theory expectations is indicated by a negative value of Tajima's D statistic, thus, consistent with the possibility of recent positive selection. Conversely, balancing selection retains genetic variants increasing the occurrence of intermediate frequency variants (Kaplan et al., 1988), and Tajima's D statistic can be elevated towards a positive value. A third feature of the data is the degree of association between polymorphisms, or linkage disequilibrium (LD). Regions under balancing selection will tend to reduce LD if the polymorphism is old, but may increase LD in a transient phase. In the presence of recombination between selected and neutral sites, a complete selective sweep can also increase LD because of the sampling of only a subset of the ancestral haplotypes.

While population genetic approaches focus on detecting ongoing selection in a population, comparative approaches uses data from multiple different species aim at detecting past selection. The

most usually statistic model used to detect selection from comparative data is the ratio of nonsynonymous differences to synonymous differences. The idea behind the test is that if synonymous mutations are essentially neutral because they do not result in a change in a protein, the rate of synonymous evolution (d_S: synonymous mutations per synonymous sites) will equal the mutation rate (Kimura, 1991). In contrast, nonsynonymous mutations are more likely to be subject to natural selection once they result in changes in protein composition. If most nonsynonymous mutations are deleterious, then the rate of nonsynonymous evolution (d_N: nonsynonymous mutations per nonsynonymous site) will be lower than neutral rate, resulting in $d_N / d_S < 1$. If a substantial fraction of nonsynonymous mutations are beneficial, however, the average rate of nonsynonymous evolution can be higher than the neutral rate, resulting in $d_N / d_S > 1$. The d_N / d_S ratio tests do not require any assumptions about population structure or equilibrium conditions, and therefore can provide particularly compelling evidence not only for selection, but also the directionality of selection (Ford, 2002).

1.6.2. Evolutionary significance of *cis*-regulatory mutations

During the past 5 to 10 years, empirical evidence has shown that mutations affecting the regulation of gene expression are a common source of evolutionary change. Expression divergence that correlates with phenotypic divergence, manipulations of gene expression that are sufficient for recreating phenotypic differences and genetic mapping identifying regulatory loci as being responsible for divergent phenotypes are evidences that changes in *cis*-regulatory sequences constitute an important part of the genetic basis for adaptation (Carroll, 2008; Stern and Orgogozo, 2008). These mutations underlie a variety of interesting and ecologically significant phenotypic differences in morphology, physiology and behavior (Carroll, 2008; Stern and Orgogozo, 2008).

However, it is difficult to comprehend the functional effects of the regulatory mutations. In part because the genetic code makes it easy to identify, accurately and comprehensively, mutations that alter protein sequences from DNA sequence comparison alone, whereas the same is not yet so evident for mutations that alter transcription, splicing, transcript stability and other regulatory processes

(Wray, 2007). This is because elucidating the genetic and molecular mechanisms that are responsible for *cis*- regulatory divergence is a complex empirical task that is mostly identified through functional or biochemical tests (Stern, 2000; Wray et al., 2003; Carroll, 2008). This difference in the ability to identify potentially functional mutations from coding and regulatory sequences makes it hard to estimate their relative contribution to the evolution of phenotypes. However, there are two hypotheses to sustain that regulatory mutations make a qualitatively distinct contribution to phenotypic evolution (Wray, 2007). The first hypothesis relies on the assumption that some kinds of phenotypic difference are easier to achieve through *cis*-regulatory mutations than through coding mutations. Phenotype aspects like reproduction, development, behavior, immune responses and resource utilization require rapid and dynamic changes in gene function, as dynamic process transcription can better “fine tune” meeting context-dependent functional demands, whereas structure is generally more static (Wray, 2007). Traits associated with dynamic processes such as these might be expected to evolve to some extent more readily through regulatory rather than coding mutations. Even if protein structures are not entirely static, and distinct isoforms can be obtain through alternative transcription start sites and splicing, and through post-translational modifications, the contribution of isoforms is relatively discrete compared to expression which is a continuous variable that can be adjusted in fine increments across a broad dynamic range. Therefore, *cis*-regulatory mutations might consequently play a disproportionate part in the evolution of quantitative traits and of responses to environmental factors that vary over time (Wray, 2007).

The second hypothesis proposes that natural selection acts differently on mutations in *cis*-regulatory sequences based on two properties of the organization and function of *cis*-regulatory regions (Ruvkun et al., 1991; Stern, 2000; Wray et al., 2003). Fist, the existence of allele-specific and independent regulation in either diploid or polyploid species suggests that mutations in *cis*-regulatory regions are often co-dominant (Chen, 2007; Leitch and Leitch, 2008; Hegarty and Hiscock, 2009; Buggs et al., 2010; Jackson and Chen, 2010; Higgins et al., 2012; Combes et al., 2013; Woodhouse et al., 2014). By contrast, many or most coding mutations are recessive. Natural selection operates far more efficiently on co-dominant mutations once they can have fitness consequences as heterozygotes, while recessive mutations requires genetic drift to raise allele frequencies until homozygotes begin to appears in

population (Ruvkun et al., 1991). The second property is the modular organization of some *cis*-regulatory regions. This structural organization means that a mutation in one module might affect only part of the overall transcription. For instance, the effects of a *cis*-regulatory mutation could be limited to a step of the spatiotemporal regulation, or to a single organ or tissue even when the gene is much more widely expressed. On the other hand, most non-synonymous coding mutations modify the resulting protein no matter where it is expressed (Wray, 2007).

The elucidation of the genetic and molecular mechanisms that are responsible for *cis*- regulatory divergence is a complex empirical task that require identifying the *cis*-regulatory elements (CREs) that controls the expression pattern of interest, finding functionally divergent sites within this region and determining how these changes in *cis*-regulatory sequence alter biochemical interactions with and among *trans*-acting factors that mediate gene expression (Wittkopp and Kalay, 2011). Candidates regions for divergent activity can be sought out of sites at which the species that has derived enhancer activity also has a derived allele. Such sites are the strongest candidates for contributing to *cis*-regulatory divergence.

1. Reference

Agarwal M, Hao Y, Kapoor A, Dong C-H, Fujii H, Zheng X, Zhu J-K (2006a) A R2R3 type MYB transcription factor is involved in the cold regulation of *CBF* genes and in acquired freezing tolerance. J Biol Chem **281**: 37636–37645

Agarwal PK, Agarwal P, Reddy MK, Sopory SK (2006b) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep **25**: 1263–1274

Agarwal PK, Jha B (2010a) Transcription factors in plants and ABA dependent and independent abiotic stress signalling. Biol Plant **54**: 201–212

Agarwal PK, Jha B (2010b) Transcription factors in plants and ABA dependent and independent abiotic stress signalling. Biol Plant **54**: 201–212

Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, Singh NK (2012) DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. J Genet **91**: 385–395

Alègre C (1959) Climates et caféiers d’Arabie. Agron Trop **14**: 23–58

Allen GJ, Chu SP, Harrington CL, Schumacher K, Hoffmann T, Tang YY, Grill E, Schroeder JI (2001) A defined range of guard cell calcium oscillation parameters encodes stomatal movements. Nature **411**: 1053–1057

Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M (1998) A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. EMBO J **17**: 5484–5496

Alvim P de T (1960) Moisture stress as a requirement for flowering of coffee. Science **132**: 354

Anthony F, Berthaud J, Guillaument JL, Lourd M (1987) Colecting wild *Coffea* species in Kenya and Tanzania. Plant Genet Resour Newsl **69**: 23–29

Anthony F, Combes C, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theor Appl Genet **104**: 894–900

Anthony F, Diniz LEC, Combes MC, Lashermes P (2010) Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. Plant Syst Evol **285**: 51–64

Bardil A, de Almeida JD, Combes MC, Lashermes P, Bertrand B (2011) Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytol **192**: 760–774

Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. CRC Crit Rev Plant Sci **24**: 23–58

Bohnert HJ, Ayoubi P, Borchert C, Bressan RA, Burnap RL, Cushman JC, Cushman MA, Deyholos M, Fischer R, Galbraith DW, et al (2001) A genomics approach towards salt stress tolerance. Plant Physiol Biochem **39**: 295–311

Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid Tragopogon miscellus by next-generation sequencing and Sequenom iPLEX Massarray genotyping. Mol Ecol **19 Suppl 1**: 132–146

Camargo AP (1985) O clima e a cafeicultura no Brasil. Inf Agropec **11**: 13–26

Camargo MBP de (2010) The impact of climatic variability and climate change on arabic coffee crop in Brazil. Bragantia **69**: 239–247

Canella D, Gilmour SJ, Kuhn LA, Thomashow MF (2010) DNA binding by the *Arabidopsis* CBF1 transcription factor requires the PKKP/RAGRxKFxETRHP signature sequence. Biochim Biophys Acta **1799**: 454–462

Carr MKV. (2001) The water relations and irrigation requirements of coffee. Exp Agric **37**: 1–36

Carroll SB (2008) Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134**: 25–36

Chagné D, Batley J, Edwards D, Forster JW (2007) Single nucleotide polymorphism genotyping in plants. In: Association Mapping in Plants (Eds. NC Oraguzie, EHA Rikkerink, SE Gardiner and HN De Silva) Springer NY, 77-94

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol **58**: 377–406

Chinnusamy V, Ohta M, Kanrar S, Lee B, Hong X, Agarwal M, Zhu J (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev **17**: 1043–1054

Cogan NOI, Ponting RC, Vecchies AC, Drayton MC, George J, Dracatos PM, Dobrowolski MP, Sawbridge TI, Smith KF, Spangenberg GC, et al (2006) Gene-associated single nucleotide polymorphism discovery in perennial ryegrass (*Lolium perenne* L.). Mol Genet Genomics **276**: 101–112

Combes MC, Dereeper A, Severac D, Bertrand B, Lashermes P (2013) Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. New Phytol **200**: 251–260

Cotta MG, Barros LMG, de Almeida JD, de Lamotte F, Barbosa E a, Vieira NG, Alves GSC, Vinecky F, Andrade AC, Marraccini P (2014) Lipid transfer proteins in coffee: isolation of *Coffea* orthologs, *Coffea arabica* homeologs, expression during coffee fruit development and promoter analysis in transgenic tobacco plants. Plant Mol Biol **85**: 11–31

Crisosto CH, Grantz DA, Meinzer FC (1992) Effects of water deficit on flower opening in coffee (*Coffea arabica* L.). Tree Physiol **10**: 127–139

Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2012) Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol **60**: 483–501

Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2008) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome **51**: 50–63

DaMatta FM (2003) Drought as a multidimensional stress affecting photosynthesis in tropical tree crops. In: A Hemantaranjan, ed, Adv. plant Physiol. Scientific Publishers, Jodhpur, pp 227–265

DaMatta FM (2004a) Ecophysiological constraints on the production of shaded and unshaded coffee: a review. F Crop Res **86**: 99–114

DaMatta FM (2004b) Exploring drought tolerance in coffee: a physiological approach with some insights for plant breeding. Brazilian J Plant Physiol **16**: 1–6

DaMatta FM, Maestri M, Barros RS (1998) Photosynthetic performance of two coffee species under drought. Photosynthetica **34**: 257–264

DaMatta FM, Ramalho JDC (2006) Impacts of drought and temperature stress on coffee physiology and production: a review. Brazilian J Plant Physiol **18**: 55–81

Davis AP, Chester M, Maurin O, Fay MF (2007) Searching for the relatives of *Coffea* (Rubiaceae, Ixoroideae): The circumscription and phylogeny of *Coffeeae* based on plastid sequence data and morphology. Am J Bot **94**: 313–329

Davis AP, Gole TW, Baena S, Moat J (2012) The impact of climate change on indigenous arabica coffee (*Coffea arabica*): Predicting future trends and identifying priorities. PLoS One **7**: e47981

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. Bot J Linn Soc **167**: 357–377

Descroix F, Snoeck J (2004) Environmental factors suitable for coffee cultivation. Coffee growing, process. Sustain. Prod. Wiley-VCH Verlag GmbH, 164–177

Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF (2009) Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. Plant Cell **21**: 972–984

Dong C-H, Agarwal M, Zhang Y, Xie Q, Zhu J-K (2006) The negative regulator of plant cold responses, HOS1, is a RING E3 ligase that mediates the ubiquitination and degradation of ICE1. Proc Natl Acad Sci USA **103**: 8281–8286

Dong C-J, Liu J-Y (2010) The *Arabidopsis* EAR-motif-containing protein RAP2.1 functions as an active transcriptional repressor to keep stress responses under tight control. BMC Plant Biol **10**: p47

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet **42**: 443–461

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J **33**: 751–763

Dutta A, Sen J, Deswal R (2007) Downregulation of terpenoid indole alkaloid biosynthetic pathway by low temperature and cloning of a AP2 type C-repeat binding factor (CBF) from *Catharanthus roseus* (L). G. Don. Plant Cell Rep **26**: 1869–1878

Eklund DM, Edqvist J (2003) Localization of nonspecific lipid transfer proteins correlate with programmed cell death responses during endosperm degradation in *Euphorbia lagascae* seedlings. PLANT Physiol **132**: 1249–1259

Ferrão R, Fonseca A, Silveira J, Ferrão M, Bragança S (2000) EMCAPA 8141 - Robustão Capixaba, variedade clonal de café conilon tolerante à seca, desenvolvida para o estado do Espírito Santo. Rev Ceres **47**: 555–560

Ford MJ (2002) Applications of selective neutrality tests to molecular ecology. Mol Ecol **11**: 1245–1262

Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell **14**: 1675–1690

Fursova O V, Pogorelko G V, Tarasov V a (2009) Identification of ICE2, a gene involved in cold acclimation which determines freezing tolerance in *Arabidopsis thaliana*. Gene **429**: 98–103

Gay C, Estrada F, Conde C, Eakin H, Villers L (2006) Potential impacts of climate change on agriculture: A case of study of coffee production in Veracruz, Mexico. Clim Change **79**: 259–288

Goldstein DB (2001) Islands of linkage disequilibrium. Nat Genet **29**: 109–111

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol **130**: 639–648

Harada A, Sakai T, Okada K (2003) Phot1 and phot2 mediate blue light-induced transient increases in cytosolic Ca^{2+} differently in *Arabidopsis* leaves. Proc Natl Acad Sci USA **100**: 8583–8588

Hegarty MJ, Hiscock SJ (2009) The complex nature of allopolyploid plant genomes. Heredity (Edinb) **103**: 100–101

Heidarvand L, Maali AR (2010) What happens in plant molecular responses to cold stress? Acta Physiol Plant **32**: 419–431

Hernandez-Garcia CM, Finer JJ (2014) Identification and validation of promoters and *cis*-acting regulatory elements. Plant Sci **217-218**: 109–119

Higgins J, Magusin A, Trick M, Fraser F, Bancroft I (2012) Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. BMC Genomics **13**: 247p

Huang G-T, Ma S-L, Bai L-P, Zhang L, Ma H, Jia P, Liu J, Zhong M, Guo Z-F (2012) Signal transduction during cold, salt, and drought stresses in plants. Mol Biol Rep **39**: 969–987

International Coffee Organization (ICO) (2014) Coffee market report. 1–6

IPCC (2013) Summary for Policymakers. Clim. Chang. 2013 Phys. Sci. basis

Jackson S, Chen ZJ (2010) Genomic and expression plasticity of polyploidy. Curr Opin Plant Biol **13**: 153–159

Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF (2001) Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol **127**: 910–917

Jiang C, Iu B, Singh J (1996) Requirement of a CCGAC *cis*-acting element for cold induction of the BN115 gene from winter *Brassica napus*. Plant Mol Biol **30**: 679–684

Jofuku KD (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. Plant Cell **6**: 1211–1225

Kaplan NL, Darden T, Hudson RR (1988) The coalescent process in models with selection. Genetics **120**: 819–829

Kaur S, Francki MG, Forster JW (2012) Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. Plant Biotechnol J **10**: 125–138

Kidokoro S, Maruyama K, Nakashima K, Imura Y, Narusaka Y, Shinwari ZK, Osakabe Y, Fujita Y, Mizoi J, Shinozaki K, et al. (2009) The phytochrome-interacting factor PIF7 negatively regulates *DREB1* expression under circadian control in *Arabidopsis*. Plant Physiol **151**: 2046–2057

Kim T-H, Böhmer M, Hu H, Nishimura N, Schroeder JI (2010) Guard cell signal transduction network: advances in understanding abscisic acid, CO_2, and Ca^{2+} signaling. Annu Rev Plant Biol **61**: 561–591

Kimura M (1991) The neutral theory of molecular evolution: a review of recent evidence. Jpn J Genet **66**: 367–386

Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, Surpin M, Lim J, Mittler R, Chory J (2007) Signals from chloroplasts converge to regulate nuclear gene expression. Science **316**: 715–719

Krishnaswamy S, Verma S, Rahman MH, Kav NV (2011) Functional characterization of four APETALA2-family genes (*RAP2.6, RAP2.6L, DREB19* and *DREB26*) in *Arabidopsis*. Plant Mol Biol **75**: 107–127

Lashermes P, Combes MC, Hueber Y, Severac D, Dereeper A (2014) Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. Plant J **78**: 674–685

Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. Mol Gen Genet **261**: 259–266

Lashermes P, Combes MC, Trouslot P, Charrier a. (1997) Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theor Appl Genet **94**: 947–955

Lata C, Prasad M (2011) Role of DREBs in regulation of abiotic stress responses in plants. J Exp Bot **62**: 4731–4748

Lata C, Yadav A, Prasad M (2007) Role of plant transcription factors in abiotic stress tolerance. In: A Shanker, ed, Abiotic stress response plants - physiol. biochem. Genet. Perspect. InTech, 269–296

Lee TI, Young RA (2000) Transcription of eukaryotic protein-coding genes. Annu Rev Genet **34**: 77–137

Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. Science **320**: 481–483

Li Y, Wang GX, Xin M, Yang HM, Wu XJ, Li T (2004) The parameters of guard cell calcium oscillation encodes stomatal oscillation and closure in *Vicia faba*. Plant Sci **166**: 415–421

Lin R-C, Park H-J, Wang H-Y (2008) Role of *Arabidopsis* RAP2.4 in regulating light- and ethylene-mediated developmental processes and drought stress tolerance. Mol Plant **1**: 42–57

Lindemose S, O'Shea C, Jensen MK, Skriver K (2013) Structure, function and networks of transcription factors involved in abiotic stress responses. Int J Mol Sci **14**: 5842–5878

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) 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*. Plant Cell **10**: 1391–1406

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys **444**: 139–158

MansWeld TJ, Atkinson CJ (1991) Stomatal behaviour in water stressed plants. In: RG Alscher, JR Cumming, eds, Stress responses plants adapt. acclim. Mech. Wiley-Liss, New York, 241–264

Marraccini P, Freire LP, Alves GS, Vieira NG, Vinecky F, Elbelt S, Ramos HJ, Montagnon C, Vieira LG, Leroy T, et al (2011) RBCS1 expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. BMC Plant Biol **11**: 85p

Matiello J (1998) Café Conillon: Como Plantar, Tratar, Colher, Preparar e Vender. 162p

Matsukura S, Mizoi J, Yoshida T, Todaka D, Ito Y, Maruyama K, Shinozaki K, Yamaguchi-Shinozaki K (2010) Comprehensive analysis of rice *DREB2*-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. Mol Genet Genomics **283**: 185–196

Mishra S, Shukla A, Upadhyay S, Sanchita, Sharma P, Singh S, Phukan UJ, Meena A, Khan F, Tripathi V, et al. (2014) Identification, occurrence, and validation of DRE and ABRE *Cis*-regulatory motifs in the promoter regions of genes of *Arabidopsis thaliana*. J Integr Plant Biol **56**: 388–399

Mitchell-Olds T, Feder M, Wray G (2008) Evolutionary and ecological functional genomics. Heredity (Edinb) **100**: 101–102

Miura K, Ohta M, Nakazawa M, Ono M, Hasegawa PM (2011) ICE1 Ser403 is necessary for protein stabilization and regulation of cold signaling and tolerance. Plant J **67**: 269–279

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta **1819**: 86–96

Molina C, Grotewold E (2005) Genome wide analysis of *Arabidopsis* core promoters. BMC Genomics **6**: 25p

Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. Genome **47**: 501–509

Mondego JM, Vidal RO, Carazzolle MF, Tokuda EK, Parizzi LP, Costa GG, Pereira LF, Andrade AC, Colombo C a, Vieira LG, et al (2011) An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. BMC Plant Biol **11**: 30p

Montagnon C (2000) Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de *Coffea canephora* Pierre. 133p

Montagnon C, Cubry P, Leroy T (2012) Amélioration génétique du caféier *Coffea canephora* Pierre: connaissances acquises, stratégies et perspectives. Cah Agric **21**: 143–153

Montagnon C, Leroy T (1993) Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de Côte d'Ivoire appartenant à différents groupes génétiques. Café Cacao Thé **37**: 179–190

Montagnon C, Leroy T, Yapo A (1992) Diversité génotypique et phénotypique de quelques groupes de caféiers (*Coffea canephora* Pierre) en collection. Conséquences sur leur utilisation en sélection. Café Cacao Thé **36**: 187–198

Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the *ERF* gene family in *Arabidopsis* and rice. Plant Physiol **140**: 411–432

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol **149**: 88–95

Navarro M, Ayax C, Martinez Y, Laur J, El Kayal W, Marque C, Teulières C (2011) Two *EguCBF1* genes overexpressed in *Eucalyptus* display a different impact on stress tolerance and plant development. Plant Biotechnol J **9**: 50–63

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet **39**: 197–218

Niu X, Helentjaris T, Bate NJ (2002) Maize ABI4 binds coupling element1 in abscisic acid and sugar response genes. Plant Cell **14**: 2565–2575

Nordborg M, Innan H (2002) Molecular population genetics. Curr Opin Plant Biol **5**: 69–73

Nowak MD, Davis AP, Yoder AD (2012) Sequence data from new plastid and nuclear COSII regions resolves early diverging lineages in *Coffea* (Rubiaceae). Syst Bot **37**: 995–1005

Ohme-Takagi M, Shinshi H (1995) Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell **7**: 173–182

Petitot A-S, Lecouls A-C, Fernandez D (2007) Sub-genomic origin and regulation patterns of a duplicated *WRKY* gene in the allotetraploid species *Coffea arabica*. Tree Genet Genomes **4**: 379–390

Pinheiro H a, Damatta FM, Chaves ARM, Loureiro ME, Ducatti C (2005) Drought tolerance is associated with rooting depth and stomatal control of water use in clones of *Coffea canephora*. Ann Bot **96**: 101–108

Poncet V, Dufour M, Hamon P, Hamon S, de Kochko A, Leroy T (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. Genome **50**: 1156–1161

Qin F, Sakuma Y, Li J, Liu Q, Li Y-Q, Shinozaki K, Yamaguchi-Shinozaki K (2004) Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. Plant Cell Physiol **45**: 1042–1052

Rae L, Lao NT, Kavanagh TA (2011) Regulation of multiple aquaporin genes in *Arabidopsis* by a pair of recently duplicated DREB transcription factors. Planta **234**: 429–444

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol **5**: 94–100

Razafinarivo NJ, Guyot R, Davis AP, Couturon E, Hamon S, Crouzillat D, Rigoreau M, Dubreuil-Tranchant C, Poncet V, De Kochko A, et al (2013) Genetic structure and diversity of coffee (*Coffea*) across Africa and the Indian Ocean islands revealed using microsatellites. Ann Bot **111**: 229–248

Reeves WM, Lynch TJ, Mobin R, Finkelstein RR (2011) Direct targets of the transcription factors ABA-Insensitive(ABI)4 and ABI5 reveal synergistic action by ABI4 and several bZIP ABA response factors. Plant Mol Biol **75**: 347–363

Roychoudhury A, Basu S (2012) Ascorbate-glutathione and plant tolerance to various abiotic stresses. In: N Anjum, S Umar, A Ahmad, eds, Oxidative Stress plants causes, consequences Toler. IK International Publishing House Pvt. Ltd., New Delhi, 177–258

Roychoudhury A, Paul S, Basu S (2013) Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. Plant Cell Rep **32**: 985–1006

Ruvkun G, Wightman B, Bürglin T, Arasu P (1991) Dominant gain-of-function mutations that lead to misregulation of the *C. elegans* heterochronic gene lin-14, and the evolutionary implications of dominant mutations in pattern-formation genes. Dev Suppl **1**: 47–54

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290**: 998–1009

Schwartz A, Ilan N, Grantz DA (1988) Calcium effects on stomatal movement in Commelina communis L.: use of EGTA to modulate stomatal response to light, KCl and CO_2. Plant Physiol **87**: 583–587

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, et al (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J **31**: 279–292

Shaikhali J, Heiber I, Seidel T, Ströher E, Hiltscher H, Birkmann S, Dietz K-J, Baier M (2008) The redox-sensitive transcription factor Rap2.4a controls nuclear expression of 2-Cys peroxiredoxin A and other chloroplast antioxidant enzymes. BMC Plant Biol **8**: 48p

Shanker A, Venkateswarlu B (2011) Abiotic stress response in plants-physiological, biochemical and genetic perspectives. 358p

Shao H-B, Guo Q-J, Chu L-Y, Zhao X-N, Su Z-L, Hu Y-C, Cheng J-F (2007) Understanding molecular mechanism of higher plant plasticity under abiotic stress. Colloids Surf B Biointerfaces **54**: 37–45

Shinwari ZK, Nakashima K, Miura S, Kasuga M, Seki M, Yamaguchi-Shinozaki K, Shinozaki K (1998) An *Arabidopsis* gene family encoding DRE/CRT binding proteins involved in low-temperature-responsive gene expression. Biochem Biophys Res Commun **250**: 161–170

Siegel RS, Xue S, Murata Y, Yang Y, Nishimura N, Wang A, Schroeder JI (2009) Calcium elevation-dependent and attenuated resting calcium-dependent abscisic acid induction of stomatal closure and abscisic acid-induced enhancement of calcium sensitivities of S-type anion and inward-rectifying K channels in *Arabidopsis* guard cells. Plant J **59**: 207–220

Smith JM, Haigh J (2007) The hitch-hiking effect of a favourable gene. Genet Res **89**: 391–403

Soltis DE, Buggs RJA, Doyle JJ, Soltis PS (2010) What we still don’t know about polyploidy. Taxon **59**: 1387–1403

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome **46**: 431–437

Stern DL (2000) Evolutionary developmental biology and the problem of variation. Evolution **54**: 1079–1091

Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? Evolution **62**: 2155–2177

Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE , a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Plant Biol **94**: 1035–1040

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–95

Thomas AS (1942) The wild arabica coffee on the Boma Plateau, Anglo-Egyptian Sudan. Emp J Exp Agric **10**: 207–212

Van Hilten HJ (2011) The coffee exporter's guide, 3rd ed. 247p

Vidal RO, Mondego JMC, Pot D, Ambrósio AB, Andrade AC, Pereira LFP, Colombo CA, Vieira LGE, Carazzolle MF, Pereira GAG (2010) A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. Plant Physiol **154**: 1053–1066

Vogel JT, Zarka DG, Van Buskirk HA, Fowler SG, Thomashow MF (2005) Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. Plant J **41**: 195–211

Wang Q, Guan Y, Wu Y, Chen H, Chen F, Chu C (2008) Overexpression of a rice *OsDREB1F* gene increases salt, drought, and low temperature tolerance in both *Arabidopsis* and rice. Plant Mol Biol **67**: 589–602

Wang Q-J, Xu K-Y, Tong Z-G, Wang S-H, Gao Z-H, Zhang J-Y, Zong C-W, Qiao Y-S, Zhang Z (2010) Characterization of a new dehydration responsive element binding factor in central arctic cowberry. Plant Cell, Tissue Organ Cult **101**: 211–219

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta **218**: 1–14

Webb AAR (2008) The chloroplast as a regulator of Ca^{2+} signalling. New Phytol **179**: 568–570

Weinl S, Held K, Schlücking K, Steinhorst L, Kuhlgert S, Hippler M, Kudla J (2008) A plastid protein crucial for Ca^{2+}-regulated stomatal responses. New Phytol **179**: 675–686

Willson K (1999) Coffee, cocoa and tea. 300p

Wittkopp PJ, Kalay G (2011) *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet **13**: 59–69

Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc Natl Acad Sci USA **111**: 5283–5288

Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet **8**: 206–216

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman M V, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol **20**: 1377–1419

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol **22**: 506–519

Wrigley G (1995) Coffee. In: J Smart, NW Simmonds, eds, Evol. Crop plants, 2nd ed. Longman Scientific and Technical, Harlow, 438–443

Xiao H, Tattersall EAR, Siddiqua MK, Cramer GR, Nassuth A (2008) CBF4 is a unique member of the CBF transcription factor family of *Vitis vinifera* and *Vitis riparia*. Plant Cell Environ **31**: 1–10

Yamaguchi-Shinozaki K, Shinozaki K (1994) A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell **6**: 251–264

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci **10**: 88–94

Yang HM, Zhang XY, Wang GX, Li Y, Wei XP (2003) Cytosolic calcium oscillation may induce stomatal oscillation in *Vicia faba*. Plant Sci **165**: 1117–1122

Yang W, Liu XD, Chi XJ, Wu CA, Li YZ, Song LL, Liu XM, Wang YF, Wang FW, Zhang C, et al. (2011) Dwarf apple *MbDREB1* enhances plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-independent pathways. Planta **233**: 219–229

Young JJ, Mehta S, Israelsson M, Godoski J, Grill E, Schroeder JI (2006) CO_2 signaling in guard cells: calcium sensitivity response modulation, a Ca^{2+}-independent phase, and CO_2 insensitivity of the *gca2* mutant. Proc Natl Acad Sci USA **103**: 7506–7511

Zou C, Sun K, Mackaluso JD, Seddon AE, Jin R, Thomashow MF, Shiu S-H (2011) *Cis*-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. Proc Natl Acad Sci USA **108**: 14992–14997

CHAPTER 1

An analysis of sequence variability in the *DREB1D* gene among different genotypes of the *Coffea* genus

1 **Article Submission: November 13, 2014**

2 **Article Accept with revisions: December 28, 2014**

3 **Promoter of *Coffea DREB1D* as a major source of diversity for this gene and potential implication in**

4 **local adaptation to drought tolerance.**

5 Running title: *DREB1D* in *Coffea*

6 Gabriel S. C. Alves[1,4], Luciana P. Freire[1,4], Natalia G. Vieira[1,4], Felipe Vinecký[1,4], Dominique This[2],

7 David Pot[3], Luciano V. Paiva[4], Pierre Marraccini[1,3] and Alan C. Andrade[1*]

8

9 [1] EMBRAPA Recursos Genéticos e Biotecnologia (LGM), Parque EB, CP 02372, 70770-917 Brasília,

10 DF, Brazil.

11 [2] Montpellier SupAgro, UMR AGAP, 34398 Montpellier, France.

12 [3] CIRAD, UMR AGAP, 34398 Montpellier, France.

13 [4] Departamento de Química, Laboratório Central de Biologia Molecular (LCBM), UFLA, Lavras,

14 Minas Gerais, Brazil.

15

16 *Corresponding author:

17 Alan Carvalho Andrade

18 alan.andrade@embrapa.br

19 Phone number: +55 61 34484795

20 Fax: +55 61 34483624

21 **Abstract**

22 Climate change is posing a major challenge to coffee worldwide production leading to a need for the
23 development of coffee cultivars with increased drought tolerance. In several plant species, the
24 application of *DREB* genes in crop improvement has achieved promising results to desiccation
25 tolerance engineering, however knowledge on natural diversity of these genes and its association to
26 phenotypic variability in coffee is largely unknown. Since the development of molecular markers is
27 one of the major goals for accelerating breeding programs, a study was done to evaluate the sequence
28 variability of the *DREB1D* gene in several *Coffea* genotypes. Nucleic acid variation in promoter and
29 coding regions of this gene were evaluated on a diverse population consisting of 38 accessions most of
30 them characterized by different phenotypes (tolerance *vs.* susceptibility) regarding to drought. Our
31 findings show several evidences of association between drought tolerance and the genetic variations
32 on *DREB1D* promoter region, but not with those from the coding region. Further analyses indicated
33 that the promoter region is evolving by the rearrangement of *cis*-regulatory elements, and as result the
34 alteration of expression patterns of *DREB1D* might consequently play a more significant part in the
35 role of this locus on drought adaptation than alteration in protein structure or function.

36 **1. Introduction**

37 The adaptation of crops to their environments and management systems is a key process for
38 overcoming the future challenges of food security and climate change. Among the adverse
39 environmental conditions, drought is a major challenge to coffee production and can lead to significant
40 production losses of up to approximately 80% [1]. Natural variability exists for drought tolerance
41 among *Coffea* species, even at the intraspecific level between and within genetic groups in *C.*
42 *canephora* [2,3]. Associations of either sequence or gene expression variability with physiological
43 traits are useful for revealing the genetic mechanism underlying drought adaptation and for identifying
44 favorable alleles for coffee improvement. These associations rely on finding statistical links between
45 genetic diversity (i.e. nucleotide diversity or expression variability) and phenotypic variability. The
46 classic biparental QTL mapping and advanced association mapping designs provide the framework for
47 performing such analyses. An alternative strategy is the identification of signatures of natural or

48 artificial selection at molecular level. Several tests intended to detect deviation from neutrality and
49 demographic equilibrium have been developed to tackle these challenges, not only at the nucleotide
50 diversity level [4,5] but also in the context of gene expression variability [6,7].

51 Single nucleotide polymorphisms (SNPs) are the most frequent form of sequence variation
52 within the genome and they make remarkable contributions to applications such as association
53 analysis, QTL mapping and identification of genetic correlations among individuals [8–10]. For
54 quantitative and qualitative traits, SNPs can have a direct impact on traits or can be related to a
55 phenotype as a result of linkage disequilibrium (LD). Therefore, the discovery of SNPs and insertions-
56 deletions (InDels) in gene sequences provides a source for the detection of associations between allelic
57 forms of a gene and phenotypes. Moreover, SNP haplotype-based analysis is more effective for
58 association mapping than analyses based on individual SNPs [11]. Next generation DNA sequencing
59 technologies (NGS) can provide large-scale genome coverage allowing a rapid and direct study of
60 diversity. As a complementary approach, the study of candidate genes identified based on
61 expressional, functional or positional evidence is of great interest for deciphering the genetic
62 determinism of traits of interest. In this last context the challenge is to provide evidence of the impact
63 of the genetic diversity of these genes on the variability of phenotypes. In comparison with the
64 relatively "blind" NGS approach, which does not always provide full genome coverage (GBS usually
65 explores only 10% of the whole genome sequence) it provides an "a priori" based analysis taking
66 advantage of already available information to focus efforts on candidate regions.

67 Candidate genes (CGs) for drought tolerance have been identified by different approaches
68 during drought acclimation in *Coffea canephora* genotypes and its molecular mechanisms have been
69 investigated [12–14]. Marraccini et al. (2012) showed that, among others, the *CcDREB1D* gene
70 displayed an increase in mRNA levels during drought stress acclimation. In addition, *CcDREB1D*
71 overexpression was observed in a drought-tolerant genotype (clone 14) of Conilon but not with a
72 sensitive genotype (clone 22). This indicates a distinct regulation for homologous gene expression in
73 the two genotypes and suggests that this gene might contribute to the diversity observed between
74 genotypes.

75 DREBs (Dehydration Responsive Element Binding) are transcription factor proteins that
76 compose abscisic acid (ABA)-dependent and -independent pathways of signal transduction in abiotic
77 stress response, and play a key role in regulating the expression of several stress-related genes [15]. In
78 view of their role, *DREB* genes are ideal candidates for crop improvement [16]. DREB proteins belong
79 to the APETALA2/Ethylene Responsive Factor (AP2/ERF) superfamily and present a unique
80 AP2/ERF domain. The AP2/ERF domain is composed of 60 amino acids and forms a secondary
81 structure of 3-stranded antiparallel β-sheets and an α helix of 18 amino acids [17]. Within the
82 AP2/ERF domain, the DREB subfamily presents the WLG motif located in the third β-sheet, the
83 amino acids valine (V14) and glutamic acid (E19) all being conserved [18]. DNA contact is mediated
84 by β-sheet residues Arg and Trp [17] to the dehydration response element (DRE) core sequence
85 CCGAC, though it is well known that DREB proteins can also bind to other *cis*-acting elements that
86 share the common DRE core, such as C-repeat and low-temperature-responsive elements [19–21].
87 DREB1/CBF (Dehydration Responsive Element Binding/C-repeat binding-factor) proteins
88 integrate the DREB subfamily in group III and present several other signature motifs in addition to the
89 conserved DNA-binding domain AP2/ERF [22]. In their work, Nakano et al. (2006) identified the
90 conserved motifs, CMIII-1 to -4, in all AtDREB1 proteins and proposed that they may play a role in
91 transcriptional activation and protein interaction. Furthermore, another particularity of this group is the
92 presence of a consensus sequence PKRPAGRTKFRETRHP that has been hypothesized as the nuclear
93 localization signal (NLS) motif [23]. The role of *DREB1/CBF* genes in abiotic stress has been studied
94 in several plants [15,24,25]. Initially, CBF genes were described as being responsive only to cold
95 stress [26–28], but some *DREB1* genes have been described as being induced by salinity, dehydration
96 and exogenous ABA treatment [29–32]. Gilmor et al. (1998) and Liu et al. (1998) concluded that
97 *Arabidopsis* genes *DREB1A/CBF3*, *DREB1B/CBF1* and *DREB1C/CBF2* were not expressed under
98 either drought or high-salinity stress, but exhibited high expression under cold stress. In contrast,
99 *DREB1D/CBF4* was reported to be induced by ABA treatment and drought stress, but not responsive
100 to cold stress [33]. Based on a phylogenetic analysis, sequence and structural similarity between CBF4
101 and other DREB1/CBF proteins, Haake et al. (2002) proposed that CBF factors evolved from a

102 common ancestral gene that, after a duplication event, gave rise to CBF1, 2, 3 ancestral genes and
103 CBF4. The high identity among CBF1, 2, and 3 promoter sequences, but the low identity between
104 these and the CBF4 promoter, indicates gene duplication followed by promoter evolution [33]. DNA-
105 binding specificity and amino acid substitutions of DREB1/CBF proteins have been studied [18,34–
106 37] but information on the promoter region is still lacking.

107 Some information on *DREB1/CBF* gene regulation has already been elucidated; *DREB1C* can
108 negatively regulate the expression of *DREB1B* and *DREB1A* [38,39] whereas MYB15 (a MYB type II
109 protein in *Arabidopsis*) was shown to repress DREB1/CBF expression by binding to its promoter
110 regions [40]. Unlike a MYC-like protein (basic helix-loop-helix, bHLH) encoded by the *Inducer of*
111 *CBF Expression 1* (*ICE1*) gene binds to MYC *cis*-acting elements in CBF3 promoters and can
112 upregulate its expression [40,41].

113 Nonetheless, only the presence of some *cis*-acting elements, such as DRE (A/GCCGACNT),
114 is not enough alone to trigger DREB1/CBF transcription factor regulation or to induce a stress
115 response. A clearer understanding of sequence structure around these *cis*-acting elements could be
116 important for the binding of transcription factors (TF) [42]. In *Arabidopsis*, 4% of all genes present in
117 their promoter sequence the DRE element but only a part of those are directly targeted by
118 *DREB1/CBF* genes [43]. Likewise, *DREB1A* and *DREB2A* target common and different genes,
119 indicating that sequences alongside the conserved motifs might also be important for binding
120 specificity [42].

121 The nucleotide diversity of the *DREB1D* gene in the coffee has not been described yet and, to
122 do so, in some *Coffea* species genome complexity has to be taken into account. The high levels of
123 genome complexity pose major additional challenges for SNP prediction. In allopolyploid plants, such
124 as *C. arabica* (amphidiploid; 2n = 4x = 44), sequence variation between subgenomes and allelic
125 variation within subgenomes coexist [44,45]. *C. arabica* is an allopolyploid arising from a natural
126 hybridization between *C. canephora* and *C. eugenioides* (2n = 2x = 22) [46], wherein SNPs within the
127 subgenome have been identified together with homologous single-nucleotide variants (HSVs) between
128 the *CaCc* and *CaCe* subgenomes (corresponding to ancestral genomes of *C. canephora* and *C.*
129 *eugenioides*, respectively) [44,45].

130 *C. arabica* has a narrow genetic base, mainly resulting from its origin (a single or a few
131 allopolyploidization events) and autogamous reproductive behavior. In contrast, *C. canephora* has
132 wide diversity, better phenotypic adaptation to some constraints and higher productivity. *C. canephora*
133 genetic diversity has been elucidated [47,48] and its population structure refined by use of
134 microsatellites [2,49,50]. Distinct groups were identified according to their geographical origins: the
135 Guinean group (G) and the Congolese group. Moreover, the Congolese group can be subdivided into
136 SG2/B, C, SG1 and UW. Among them, two major groups contrasting for drought response, tolerant
137 SG1 and susceptible SG2, were identified [3].

138 In this context the goals of this study were to (i) evaluate nucleotide diversity in the promoter
139 and coding region of *DREB1D*, (ii) link the occurrence of DNA *cis*-regulatory elements to the pattern
140 of nucleotide diversity, (iii) evaluate nucleotide diversity within *C. arabica* subgenomes and (iv) put
141 forward hypotheses regarding the potential role of the nucleotide diversity of *DREB1D* in the drought
142 tolerance of *C. canephora* and *C. arabica*.

143 **2. Materials and methods**

144 *Plant materials and DNA isolation*

145 The plant material used in this study was obtained from a selection of genotypes of *Coffea canephora*
146 and *C. arabica* species, and also included *C. eugenioides,* which is particularly relevant in the context
147 of *C. arabica* analysis. The genotypes were chosen in order to obtain a wide coverage of the coffee
148 genotypic variability based on criteria such as the diversity of geographical origin as well as
149 morphological and agronomic differences. *C. arabica* accessions encompass the diversity of Ethiopia,
150 the primary center of diversity for this species, with high quality cultivars and accessions introgressed
151 with *C. canephora* genotypes. As for *C. canephora*, the accessions mainly cover the diversity of the
152 Congolese group (SG1, SG2 and C, including accessions that have been shown to diverge regarding
153 their response to drought, clones 14 and 22), the UW subgroup and one genotype of the Guinean
154 group [50]. A total of 38 genotypes was used in this study (Table 1) corresponding to 11 *C.*
155 *canephora*, 26 *C. arabica* and 1 *C. eugenioides*. DNA was extracted from 100 mg of pulverized leaf

156 tissues according to Saghai-Maroof (1984). Slightly different sets of accessions were considered for
157 the analysis of the coding and promoter sequence (Table 1).

158 *Gene amplification and sequencing*

159 The *DREB1D* promoter and coding sequences were isolated by PCR on genomic DNAs obtained from
160 the panel of genotypes. The coding sequences and 3'-UTR were amplified using oligonucleotides
161 (Table 2) designed according to published DNA sequences for *C. canephora* (Accession number
162 GW463524). The promoter sequence of the *DREB1D* gene was amplified using the primers pairs
163 (Table 2) designed according to the *DREB1D* promoter reference sequence available in the Coffee
164 Genome [51]. PCR reactions were performed using 100 ng of gDNA as a template using Taq-DNA
165 polymerase (Invitrogen, São Paulo, SP, Brazil) in an amplification program with a denaturation step at
166 94°C for 5 min followed by 30 amplification cycles at 94°C for 45 s, 69.7°C for 30 s and 72°C for 2
167 min, and final extension at 72°C for 5 min. Then, to isolate the alleles, the amplified sequences were
168 cloned into the Topo 2.1 pCR vector system (Invitrogen). For each genotype 12 recombinant plasmids
169 with the PCR amplification products were sequenced in both directions with universal primers (M13F
170 and M13R) and internal primers (DrebS2F and DrebS1R). Similarly, the primer pair Dreb1AF and
171 Dreb1AR was used to amplify the entire full-length cDNA of the *DREB1D* gene. The PCR product
172 was then cloned and sequenced with the universal primers. Each cloned fragment was sequenced by
173 the dideoxy chain termination method using the BigDye Terminator V3.1 sequencing kit on an ABI
174 PRISM 3130xl (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's
175 instructions. Base calling and quality trimming of ABI chromatograms was done by Phred [52,53] at a
176 stringent quality setting of 40 (1 error in 10,000 bp) and manual edition was also carried out when
177 necessary. Standard chromatogram files of quality-trimmed sequences were used for assembly.

178 *Sequence analysis*

179 All the analyses were performed separately for the promoter and coding regions. Sequences were
180 aligned with the SEQMAN 9.1.1 (4). 418 program (DNASTAR Version 2.1.0.97, Copyright 2010-
181 2011, Inc.). The numbers of polymorphic sites S, the nucleotide diversity π [54] along with the number
182 of segregating sites θ [55] with their sampling variances were estimated for promoter and coding
183 nucleotide sequences. The average number of nucleotide substitutions per site between populations
184 (Dxy) was used to analyze sequence divergence among the species. Haplotype diversity estimations
185 and tests of deviation from neutrality and demographic equilibrium were performed when relevant.
186 Alignment gaps were excluded from comparisons. The synonymous and non-synonymous variations
187 within and between species were evaluated. To analyze the pattern of diversity along the gene we
188 applied the sliding window method with a window size of 100 bp and a step size of 25 bp. All
189 analyses were performed using the DnaSP program version 5 [56]. Identification and location of the
190 DNA *cis*-regulatory elements were searched using the PlantPAN (http://plantpan.mbc.nctu.edu.tw,
191 [57]) and the TSSP / Prediction of Plant Promoters (http://www.softberry.com, [58]) web interfaces. A
192 multiple alignment of the predicted proteins was performed in the CLUSTALW program [59].
193 Conserved protein motifs were identified using the SALAD database
194 (http://salad.dna.affrc.go.jp/salad, [60]. The functional effects of non-synonymous SNPs were
195 predicted with the SNAP method [61].

196 *Haplotype network analysis*

197 The median joining haplotype networks were built using Network 4.5.1 [62].

198

199 **3. Results**

200 *Gene amplification*

201 Promoter (1.5 Kb spanning the promoter / 5'UTR) and coding sequences (829 bp spanning the coding
202 sequence/3'-UTR) homologous to the *DREB1D* gene were isolated from 29 and 25 coffee genotypes,
203 respectively (Table 1).

204 *Sequence diversity of the DREB1D gene on the whole panel of Coffea genotypes*

205 A total of 960 sequences were assembled and only alleles present in at least two different sequences
206 were considered as polymorphisms (these two different sequences could originate either from the same
207 genotype or from different genotypes). Insertions and deletions (INDELs) were observed exclusively
208 in the promoter region and represented a total of 163 sites, including a microsatellite repeat (TCA) 2 to
209 (TCA) 16 (Figure 1). We identified 155 SNPs along the 2332 bp of the *DREB1D* gene, with 132 in the
210 non-coding sequence (including one tri-allelic site) and 23 in the coding sequence, resulting in an
211 average polymorphism density of one SNP every 11 bp and 30 bp, respectively (Table 3). Of the 154
212 polymorphic sites, 116 were parsimony-informative sites (the minor allele was present in at least two
213 different genotypes).
214 Eighteen protein sequences of 233 amino acids were deduced from a 699 bp of coding sequences
215 (Figure 2). As commonly observed for members of its subfamily, the *DREB1D* gene did not exhibit
216 any intron. Moreover, DREB1D proteins presented a high sequence identity at the AP2/ERF DNA-
217 binding domain and at the other 4 signature domains of the CBF/DREB1 group, CMIII-1 to 4 (Figure
218 2). Out of the 23 SNPs found in the coding sequence, 10 resulted in non-synonymous variation,
219 including 9 neutral modifications and only one non-neutral change on the sixth amino acid residue, an
220 Isoleucine to Threonine (Figure 2).
221 Several *cis*-acting elements were identified in the promoter region, including regulatory elements
222 commonly observed in *DREB* promoters and involved in osmotic-, cold- and drought-inducible gene

223 expression (Figure 1, Table 4). These included 1 *cis*-acting element belonging to ABA-independent
224 pathways (DRE), 17 ABA-dependent regulatory motifs (9 MYB-, 4 ABRE-, 7 MYC-like regulatory
225 sequences) and 1 cold responsive regulatory element (ICEr2). For the following analyses we
226 considered only those motifs related to drought response. We analyzed the polymorphism distribution
227 in a sliding window graphic to analyze the co-occurrence of polymorphic sites and *cis*-acting elements
228 along the 1503 bp of the *DREB1D* promoter sequence (Figure 3). The first 700 bp of the promoter
229 region harbored most of the larger invariant regions identified in the *DREB1D* promoter (Figure 1),
230 and within this region ABRE and DRE, the two major *cis*-elements responsive to osmotic stress, were
231 72 bp apart. Nevertheless, overlaps between the polymorphic site frequency and *cis*-acting elements
232 were more frequent in the last 840 bp, including a peak at -900 bp coinciding with one MYC and one
233 MYB element (Figure 3). Ten polymorphic sites promoted changes in the core sequence of nine *cis*-
234 regulatory elements, 1 ICEr2, 3 ABRE-, 2 MYC- and 3 MYB-like (Figure 1).
235 A total of 37 and 18 haplotypes were obtained from the analysis of the whole set of genotypes for the
236 promoter and coding regions respectively, when InDels were excluded from the analysis (Table 3).
237 When considered, the InDels increased the number of haplotypes to 45 for the promoter region. One to
238 two alleles were identified for diploid species and two to four alleles for allotetraploids (Table 5). Not
239 surprisingly, the number of polymorphic sites in the promoter region (S_P) was considerably higher
240 compared to the coding region (S_C) (Figure 3).

241 *DREB1D intra- and interspecies diversity in Coffea*

242 When considered separately, *C. canephora* presented 94 SNPs and a total of 27 InDels events (events
243 varying between 1 and 32 bp) (Table 3). We evaluated the promoter and coding regions separately to
244 investigate the evolutionary signatures on the *DREB1D* locus. The promoter region exhibited a slightly
245 greater nucleotide diversity (0,01238) compared to the coding region (0,01006) and a greater
246 haplotype diversity. The *C. canephora* population structure was highlighted by haplotype distribution
247 and frequency. In particular, the NJ network built for promoter haplotypes revealed the clustering of
248 *C. canephora* haplotypes from the same genetic groups. Five groups were obtained, indicated in red,

249 dark red, blue, orange and yellow, which contained haplotypes from genetic groups SG1, G, B, SG2
250 and UW, respectively (Figure 4B). Exceptionally, haplotype 17 of clone 22 did not cluster with the
251 SG1 group, as might have been expected. Differently, no specific cluster related to genetic groups
252 could be observed in the coding haplotype network (Figure 4A). Haplotype H_2, although specific to
253 *C. canephora,* was shared by the Congolese groups B, C, SG1 and SG2. Moreover, some *C.*
254 *canephora* and *C. arabica* genotypes shared a common haplotype, H-1. Tajima's D test (D) was
255 conducted to test departure from the neutral theory and demographic equilibrium. A negative D value
256 was obtained for the *CcDREB1D* promoter and a positive one for the coding region. However, both
257 results were not statistically significant (Table 3).
258 Further, we analyzed sequence divergence between *C. canephora* and *C. eugenioides* accessions. The
259 average number of nucleotide substitutions per site (Dxy) between *CcDREB1D* and *CeDREB1D* was
260 plotted against the nucleotide position in a slide window graphic (Figure 5). The main divergence
261 between the homologous gene *CcDREB1D* and *CeDREB1D* lay in three major regions, -125 to -325, -
262 600 to -1000 and -1230 to -1503 (Figure 5). Taking into account the presence of drought related *cis*-
263 elements in these regions, the sequence variations led to a loss or gain of some *cis*-elements in the last
264 two regions only. Exclusively *C. eugenioides* rendered a third MYB1AT repressor element at -648 and
265 the ABRELATERD1 motif at -1389. On the other hand, only *C. canephora* presented a *cis*-regulatory
266 element ABRE-Like at -731, adding a second ABA regulatory element to the first 733 bp of the
267 *CcDREB1D* promoter.
268 In *C. arabica*, we found a total of 91 SNPs and a total of 27 InDels events (events varying between 1
269 and 29 bp) (Table 3). As *C. arabica* originates from an allopolyploidization event between *C.*
270 *canephora* and *C. eugenioides*, the concomitant analysis of *C. canephora* and *C. eugenioides*
271 genotypes was used to infer *C. arabica* subgenome haplotypes. When considering arabica genotypes
272 (yellow) in the promoter NJ network (Figure 4B), a pair of well separated haplotype clusters
273 presenting similarity with the *C. eugenioides* and *C. canephora* accessions clearly indicated a
274 partitioning of the *C. arabica* genomes into the *CaCc* and *CaCe* subgenomes. All *C. arabica*
275 genotypes presented at least one *CaCc* and *CaCe* haplotype (Table 5). However, the coding sequence
276 NJ network did not reveal such a dissociation of the subgenomes (Figure 4A). Polymorphic sites were

277 more frequent in the regions -38 to -270, -825 to -1000 and -1265 to 1375 (Figure 6). Interestingly, the
278 difference in S pattern between the *CaDREB1D* and *CcDREB1D* promoters seemed to match the Dxy
279 found between *C. canephora* and *C. eugenioides*.
280 In order to determine the main source of *CaDREB1D* diversity, we evaluated π for the *CaCc* or *CaCe*
281 subgenomes. Each of the subgenomes, *CaCe* and *CaCc*, presented a similar low number of
282 segregating sites and nucleotide diversity for the coding region (Table 6). Considering the promoter
283 region, the observed S (35) and π (0,00385) values for the *CaCe* subgenome were higher than the
284 values of *CaCc*, S (13) and π (0,00096). Indeed, the main source of *CaDREB1D* diversity was derived
285 from the *CaCe* subgenome and relied on the regions -125 to -325 and -1235 to -1375, matching the
286 main two divergence regions in $S_{(Cc)}$ and $S_{(Ca)}$ frequency (Figure 7). For *CaCe*, a non-significant
287 Tajima's D value was observed whereas a significantly negative D value was observed in the promoter
288 region for *CaCc* (Table 6). A significantly negative D can indicate an excess of rare variants, which
289 was confirmed by the promoter haplotype structure of *C. arabica* (Figure 4B) where the *CaCc*
290 subgenome presented a star-like pattern (one high frequency haplotype with several low divergent and
291 low frequency haplotypes).
292 In addition, we compared the nucleotide diversity of *C. canephora* accessions and the *C. canephora*
293 subgenome (*CaCc*). It was found that π indicated much lower diversity in the *CaCc* subgenome
294 compared with the *C. canephora* genome (Figure 8).

295 **4. Discussion**

296 *DREB* genes have been the subject of an extensive amount of research aimed at revealing the role of
297 transcription factors in the regulation and adaptation of plant responses to abiotic stress [63].
298 Considerable attention has been paid to the identification and functional validation of *DREB* genes,
299 but knowledge of the natural diversity of *DREB* genes and how it contributes to natural variation in
300 plant responses to abiotic stress is limited. In particular, their promoter nucleotide diversity and their
301 potential roles in stress responses have yet to be thoroughly explored. Here, we analyzed the natural
302 diversity of the candidate gene for drought tolerance, *DREB1D,* in the two economically important
303 species of coffee, *C. arabica* and *C. canephora*. Unlike the majority of previous studies, which

304 focused on associating drought tolerance with sequence variations mostly present in the coding region
305 of *DREB* genes, this study evaluated the occurrence of sequence variations also in non-coding regions
306 such as promoter, 5'- and 3'- UTRs sequences.

307 *DREB1D polymorphisms in Coffea*

308 DNA polymorphism and π values for *DREB* genes have been previously reported in some model
309 species and crops such as *Arabidopsis* [64], maize [11], tomatoes [65], sunflower [66], and common
310 bean [67]. The nucleotide diversity estimates obtained in our study were consistent with previous
311 reports for *DREB* genes [11,64,66]. As expected, nucleotide diversity was higher for non-coding
312 regions than for coding regions. The average polymorphism for non-coding sequences, found to be
313 one SNP every 12 bp and 8 bp for the promoter and 3'UTR, respectively, were also similar to
314 previously described results [64]. Here, we showed that nucleotide substitutions are more frequent in
315 certain regions of the promoter and the coding region and that the regions are shared at intraspecific
316 and interspecific level. *C. canephora* presented a similar number of polymorphic sites, 94 (1 SNP/23
317 bp), to *C. arabica,* 91 (1 SNP/24 bp). *C. arabica* harbored more polymorphic sites in the promoter
318 region than *C. canephora,* while the latter was more polymorphic in the coding and 3'UTR regions
319 (Table 3). Although less polymorphic in the promoter region, *C. canephora* diversity was mostly
320 observed near the *cis*-acting elements that mediate drought-inducible gene expression. Changes in or
321 around *cis*-acting elements might either influence the binding stability between TF and DNA or
322 promote the rearrangement of *cis*-regulatory elements along the promoter region [42]. The
323 arrangement of *cis*-acting elements has great importance in transcriptional regulation, since
324 combinatorial interactions of multiple nearby regulatory elements are necessary for gene expression
325 under stress conditions [42].
326 A low level of diversity was observed at protein level among the homologous sequences. Synonymous
327 mutations amounted to 56% of sequence variations and all non-synonymous mutations were neutral
328 with the exception of one. All specific motifs (CMIII-1,-2, -3 and -4) were identified in DREB1D and
329 only four amino acid changes were observed within motifs. By considering the low level of diversity

330 at protein level, the high degree of sequence identity within motifs and the *in silico* prediction of SNP
331 effects, no significant change in protein activity should be expected. In support of this assumption, the
332 majority of the amino acid substitutions were in the COOH terminal region and the two substitutions
333 in the NH_3-terminal region were outside the binding DNA domain. In yeast and *Arabidopsis*, direct
334 induced mutation experiments showed that the COOH terminal region is remarkably tolerant of amino
335 acid substitutions harboring motifs with substantial functional redundancy that ensures transactivation,
336 and no single residue is critical for transactivation activity [37]. Similarly, amino acid residue changes
337 observed in haplotypes of the *ZmDREB2.7* gene in contrasting genotypes regarding drought response
338 had no effect on protein transactivation activity and were not considered as functional variations
339 conferring drought tolerance [11].

340 Five invariant regions were found in the homologous *DREB1D* promoters and two of them flanking
341 the most polymorphic region. Concomitant analyses of these conserved sequences on the PlantPAN
342 and Softberry sites assigned regulatory elements such as DRE (Dehydration Responsive Element),
343 ABRE (ABA Responsive Element) and an Inr (Initiator element). Conservation of these sequences
344 highlighted the importance of these regulatory elements in *DREB1D* regulation and supported the
345 reported studies on *DREB1D* gene-inducible expression and its predicted role in ABA and ABA-
346 independent pathways [15,24,33,42,68]. Additionally, the occurrence within the first 500 bp, the short
347 gap length of 72 bp and the orientation of the first ABRE and DRE *cis*-motifs – ABRE in direct (D)
348 and DRE in reverse (R) form – endorsed the implication of *Coffea* DREB1D protein in osmotic stress
349 responses in addition to its the role in cold response commonly observed for DREB1/CBF proteins. A
350 recent study revealed the orientation patterns and gap length occurrence in the predicted gene
351 promoters in *A. thaliana* [69]. As a result, the majority of the more upregulated genes under osmotic
352 stress and ABA treatment presented a DR form and a gap length of 100 bp for ABRE and DRE *cis*-
353 motifs present in the first 1000 bp [69]. However, the occurrence of regulatory elements in the
354 *DREB1D* promoter was not restricted to the conserved regions. More ABRE sites, as well as other *cis*-
355 acting sequences, were identified along the promoter. A MYB1AT, a binding domain of MYB15, a
356 MYB repressor type II, was found at the -649 position. The MYB15 repressor was shown to possess
357 greater binding activity to the MYB1AT regulatory elements present within -750/-1000 of the CBF1,

358 CBF2 and CBF3 genes [70]. Interestingly, some of the INDELs and SNPs identified overlapped with
359 the *cis*-regulatory DNA motifs. A comparison of predicted *cis*-acting elements for all the homologous
360 and orthologous promoter sequences signaled the loss or gain of some regulatory DNA elements. This
361 rearrangement of regulatory elements is a result of the nine modifications observed on the core
362 sequence of *cis*-acting elements which promoted changes on the recognition sequence and may have a
363 direct effect on transcriptional regulation control of the *DREB1D* gene. In addition, the majority of the
364 polymorphism led to sequence modifications surrounding the core of the regulatory elements, which
365 cannot impair site recognition but could influence the binding affinity of the transcription factors to
366 the DNA *cis*-regulatory elements and thus modulate gene expression.
367 Remarkably, the *C. canephora* frequency spectrum for nucleotide diversity (π) showed an increase in
368 sequence variations at the most active region of the MYB repressor [70]. If we consider that in most
369 cases only a small portion of the sites in a gene may be subjected to selection [71], we can assume a
370 possible effect of selective pressure within a specific region of the promoter sequence. In this case we
371 can hypothesize that some regions of *DREB1D* would be more a target of selection, while some
372 conserved regions assure the basic motif necessary for transcriptional control at basal level. These
373 hypotheses are consistent with previously proposed hypotheses that natural selection operates
374 differently on mutations in *cis*-regulatory regions. One reason is the structural organization of some
375 *cis*-regulatory regions in modules which can lead to a modular effect of mutations, meaning that a
376 mutation in one module might not affect the entire transcription profile [7]. In the following topic we
377 discuss the potential effects of haplotype variability and adaptive variability in the *DREB1D* locus of
378 *Coffea*.

379 *DREB1D diversity and drought adaptation in the Coffea panel*

380 In *CcDREB1D*, 7 haplotypes were identified in the coding region and 14 in the promoter sequences.
381 The greater nucleotide diversity in the non-coding region was accompanied by an increase in the
382 number of haplotypes in the promoter region compared to that obtained for the coding region.
383 Depending on the haplotype network based on the coding region, no clear association could be

384 established between its structure and the previously reported adaptation to drought of the different
385 genotypes or genetic groups. Nevertheless, in the promoter haplotype NJ network the combination of
386 population structure and estimated habitat drought tolerance of *C. canephora* accessions suggested a
387 possible relationship between adaptive variation and haplotype variability. Seven Congolese
388 accessions from five distinct genetic groups, SG1, G, B, SG2 and UW, matching the population
389 groups accessed by SSR markers [49] were analyzed in this study. The structure of diversity observed
390 for the DREB1D promoter did not correspond to the one expected based on neutral SSR information
391 (Cubry et al. 2012); with SSR markers we should likely expect one to be differentiated at neutral level.
392 Indeed, genotypes from geographical regions with equivalent mid-season dry periods showed greater
393 similarity in the promoter sequences, such as SG1 and G endemic to dry forests, and B, SG2 and UW
394 endemic to rain forests of Africa, which was not observed by Cubry et al. 2012. In addition, the SG2
395 group had been shown to be less adapted to drought stress than SG1 and G [3]. We also demonstrated
396 that *C. canephora* clones 14 and 22, belonging to the SG1 group, share a common haplotype (hp15)
397 and are distinguished by their second allele hp16 and hp17, respectively. Interestingly, the hp15 and
398 hp16 alleles were clustered in the SG1 group as expected, but hp17 was more akin to SG2 and B
399 groups, both less adapted to drought. The physiological characteristics of *C. canephora* clones 14
400 (tolerant) and 22 (susceptible) under drought stress, and the first insight of the molecular mechanism
401 underlying the adaptive variation, were previously reported [12]. Among the genes differentially
402 expressed by clones 14 and 22, the *DREB1D* gene was upregulated in response to drought and had a
403 significantly higher expression in the tolerant clone [12]. Taken together, all this evidence reveals a
404 potential relationship in haplotype variability of the DREB1D promoter and adaptive variation in *C.*
405 *canephora*.

406 *Subgenome diversity of CaDREB1D and departure from neutrality*

407 In some allopolyploid species, such as *C. arabica*, the subgenomes evolve independently with low
408 levels of inter-subgenome recombination being observed. Thus, evaluation of π and θ at species level
409 is not relevant. In order to evaluate π and θ values for *C. arabica* it was strictly necessary to separate

410 its two subgenomes. A NJ network was constructed using *C. arabica, C. canephora* and *C.*
411 *eugenioides* haplotypes from the coding and promoter regions to identify the subgenomes *CaCc* and
412 *CaCe* corresponding to its ancestral genomes. *C. arabica* accessions were shown here to harbor few
413 haplotypes with high frequency and shared two commons alleles with Canephora accessions in the
414 coding region. Under the neutral theory, random drift increases interspecific variability, leading to
415 nucleotide diversification and therefore one would expect divergence at haplotype level for the
416 *CcDREB1D* and *CaCcDREB1D* coding regions in Canephora and Arabica, respectively. As *C.*
417 *arabica* and *C. canephora* are derived from a common ancestor, some genetic similarity is expected.
418 Nevertheless, the diversity observed between *C. canephora* and its counterpart *CaCc* subgenome is a
419 result of evolutionary history 1 million years [72] and reproductive characteristics. However, the
420 reduced interspecific variability and πa/πs lower than one indicate a departure from neutral
421 expectations, reinforcing the assumption that the coding region is a target of negative selection. Under
422 purifying selection deleterious mutations are discarded to maintain improved structures already fixed
423 in a population.

424 For the promoter region, the significantly negative D, the low π value and the star-like pattern of
425 haplotype structure indicate an excess of rare variants for the *CaCc* subgenome. These could be
426 signatures of either negative selection or selective sweeps. If we consider the negative selection
427 perspective, given the values observed in the DREB1D promoter region would be under stronger
428 selection than those in its coding sequence. However, it could also be a consequence of a selective
429 sweep, in this case the reduced variation in the promoter region could be a result of negative selection
430 acting on the coding sequence. The $\pi_{(Cc)}$ and $\pi_{(CaCc)}$ frequency along the promoter region shows the
431 narrow integration of *C. canephora* diversity in the *CaCc* subgenome. *C. canephora* natural diversity
432 has been used over the years to increase *C. arabica* diversity and improve agronomic traits. Here, we
433 show that *CcDREB1D* diversity could probably be used to increase the diversity of this gene in the *C.*
434 *arabica* species. A practical application of this strategy is use of the Timor hybrid (TH), a spontaneous
435 interspecific hybrid of *C. canephora* and *C. arabica* [72], in breeding programs to improve *C. arabica*
436 resistance to various diseases. As previously discussed, we could not detect any introgression of
437 *CcDREB1D* diversity in the *CaCc* subgenome of Arabica accessions. We detected an increase in *CaCe*

438 nucleotide diversity (π) in the promoter region and the occurrence of two intermediate frequency
439 haplotypes in the promoter NJ network, which might indicate balancing selection. However, within the
440 *C. arabica* accessions evaluated in this study, 6 displayed previous or recent genetic introgression of
441 TH. Therefore, we believe that the frequency observed for Hp_19 might be a consequence of genetic
442 introgression to the *CaCe* subgenome through backcrossing to TH.

443 **5. Conclusions**

444 In this study, comparative analyses of non-coding and coding regions revealed the importance of these
445 concomitant analyses in searching for signatures of molecular selection. Numerous studies have
446 supported the idea that *cis*-regulatory mutations constitute an important part of the genetic basis of
447 adaptive variation to environmental changes. It is believed that some phenotypic changes are more
448 likely to result from *cis*-regulatory mutations than from coding mutations [7]. Our findings suggest
449 that the diversity of *DREB1D* loci in *Coffea* evolved through diversification of the DNA regulatory
450 elements and conservation of the protein function. A fine dissection of the promoter region of
451 genotypes differing in drought tolerance should give additional insights into the genetic mechanisms
452 underlying the natural variability of drought response in coffee.

453 **6. Accession numbers**

454 The sequences from all coffee genotypes used in this present study were deposited in the GenBank
455 database. The *DREB1D* cDNA sequences are under the accession numbers KM281210 to KM281285.
456 The promoter sequences were deposited under the accession numbers KM281286 to KM281383.

457 **7. Acknowledgments**

458 This work was carried out under the Embrapa-CIRAD scientific cooperation project entitled "Genetic
459 determinism of drought tolerance in coffee". The authors acknowledge the financial support from the
460 Brazilian Coffee R&D Consortium, FINEP, INCT-café (CNPq/FAPEMIG) and CAPES/COFECUB
461 (Project Sv738-12/University of Lavras and Montpellier SupAgro). The authors would like to thank
462 Drs Aymbiré Francisco Almeida da Fonseca and Romário Gava Ferrão from the INCAPER Institute,

463 the Centre National de la Recherche Agronomique (CNRA) -Divo (Ivory Coast) and the Coffee
464 Research Center (NARO-COREC) -Mukono (Uganda) for providing plant material. They also
465 acknowledge Dr. Thierry Leroy and Peter Biggins (CIRAD UMR AGAP) for their discussions and
466 critical reading of the manuscript.

1 **8. References**

2 1. DaMatta FM, Ramalho JDC (2006) Impacts of drought and temperature stress on coffee
3 physiology and production: a review. Braz J Plant Physiol 18: 55–81.
4 2. Montagnon C, Cubry P, Leroy T (2012) Amélioration génétique du caféier *Coffea canephora*
5 Pierre: connaissances acquises, stratégies et perspectives. Cah Agric 21: 143–153.
6 3. Montagnon C, Leroy T (1993) Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de
7 Côte d'Ivoire appartenant à différents groupes génétiques. Café Cacao Thé 37: 179–190.
8 4. Ford MJ (2002) Applications of selective neutrality tests to molecular ecology. Mol Ecol 11:
9 1245–1262.
10 5. Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39: 197–218.
11 6. Nuzhdin S V, Wayne ML, Harmon KL, McIntyre LM (2004) Common pattern of evolution of
12 gene expression level and protein sequence in Drosophila. Mol Biol Evol 21: 1308–1317.
13 7. Wray G a (2007) The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet 8:
14 206–216.
15 8. Garcés-Claver A, Fellman SM, Gil-Ortega R, Jahn M, Arnedo-Andrés MS (2007)
16 Identification, validation and survey of a single nucleotide polymorphism (SNP) associated
17 with pungency in *Capsicum spp*. Theor Appl Genet 115: 907–916.
18 9. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases.
19 Science 273: 1516–1517.
20 10. Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, et al. (2007) A soybean transcript
21 map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:
22 685–696.
23 11. Liu S, Wang X, Wang H, Xin H, Yang X, et al. (2013) Genome-wide analysis of *ZmDREB*
24 genes and their association with natural variation in drought tolerance at seedling stage of *Zea*
25 *mays* L. PLoS Genet 9: e1003790.
26 12. Marraccini P, Vinecky F, Alves GSC, Ramos HJO, Elbelt S, et al. (2012) Differentially
27 expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of
28 *Coffea canephora*. J Exp Bot 63: 4191–4212.
29 13. Vinecky F, Silva FR da, Andrade AC (2012) Análise in silico das bibliotecas de cDNA SH2 e
30 SH3 para a identificação de genes responsivos à seca em cafeeiro. Coffee Sci 7: 1–19.
31 14. Vieira NG, Carneiro FA, Sujii PS, Alekcevetch JC, Freire LP, et al. (2013) Different molecular
32 mechanisms account for drought tolerance in *Coffea canephora* var. Conilon. Trop Plant Biol
33 6: 181–190.
34 15. Lata C, Prasad M (2011) Role of DREBs in regulation of abiotic stress responses in plants. J
35 Exp Bot 62: 4731–4748.
36 16. Xu Z-S, Chen M, Li L-C, Ma Y-Z (2011) Functions and application of the AP2/ERF
37 transcription factor family in crop improvement. J Integr Plant Biol 53: 570–585.
38 17. Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M (1998) A novel mode of DNA
39 recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain
40 in complex with DNA. EMBO J 17: 5484–5496.
41 18. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, et al. (2002) DNA-binding specificity of
42 the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration-
43 and cold-inducible gene expression. Biochem Biophys Res Commun 290: 998–1009.
44 19. Baker SS, Wilhelm KS, Thomashow MF (1994) The 5'-region of *Arabidopsis thaliana* cor15a
45 has *cis*-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant
46 Mol Biol 24: 701–713.
47 20. Jiang C, Iu B, Singh J (1996) Requirement of a CCGAC *cis*-acting element for cold induction
48 of the BN115 gene from winter *Brassica napus*. Plant Mol Biol 30: 679–684.
49 21. Thomashow MF (1999) PLANT COLD ACCLIMATION: freezing tolerance genes and
50 regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 50: 571–599.
51 22. Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene
52 family in *Arabidopsis* and rice. Plant Physiol 140: 411–432.

53 23. Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, et al. (2001) Components of the
54 *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway
55 are conserved in *Brassica napus* and other plant species. Plant Physiol 127: 910–917.
56 24. Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, et al. (2012) DREB1/CBF transcription
57 factors: their structure, function and role in abiotic stress tolerance in plants. J Genet 91: 385–
58 395.
59 25. Agarwal PK, Agarwal P, Reddy MK, Sopory SK (2006) Role of DREB transcription factors in
60 abiotic and biotic stress tolerance in plants. Plant Cell Rep 25: 1263–1274.
61 26. Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, et al. (1998) Low
62 temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an
63 early step in cold-induced COR gene expression. Plant J 16: 433–442.
64 27. Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, et al. (1998) Two transcription factors, DREB1
65 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal
66 transduction pathways in drought-and low-temperature-responsive gene expression,
67 respectively, in *Arabidopsis*. Plant Cell 10: 1391–1406..
68 28. Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in
69 response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol 149: 88–95.
70 29. Jiang F, Wang F, Wu Z, Li Y, Shi G, et al. (2010) Components of the *Arabidopsis* CBF cold-
71 response pathway are conserved in non-heading chinese cabbage. Plant Mol Biol Rep 29: 525–
72 532.
73 30. Yang W, Liu X-D, Chi X-J, Wu C-A, Li Y-Z, et al. (2011) Dwarf apple *MbDREB1* enhances
74 plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-
75 independent pathways. Planta 233: 219–229.
76 31. Wang Q, Guan Y, Wu Y, Chen H, Chen F, et al. (2008) Overexpression of a rice *OsDREB1F*
77 gene increases salt, drought, and low temperature tolerance in both *Arabidopsis* and rice. Plant
78 Mol Biol 67: 589–602.
79 32. Wang Q-J, Xu K-Y, Tong Z-G, Wang S-H, Gao Z-H, et al. (2010) Characterization of a new
80 dehydration responsive element binding factor in central arctic cowberry. Plant Cell Tiss Org
81 101: 211–219.
82 33. Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, et al. (2002) Transcription
83 factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol 130: 639–648.
84 34. Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, et al. (2003) OsDREB genes in rice,
85 *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-
86 responsive gene expression. Plant J 33: 751–763.
87 35. Mondini L, Nachit M, Porceddu E, Pagnotta MA (2012) Identification of SNP mutations in
88 *DREB1*, *HKT1*, and *WRKY1* genes involved in drought and salt stress tolerance in durum wheat
89 (*Triticum turgidum* L. var durum). OMICS 16: 178–187.
90 36. Wei B, Jing R, Wang C, Chen J, Mao X, et al. (2008) Dreb1 genes in wheat (*Triticum aestivum*
91 L.): development of functional markers and gene mapping based on SNPs. Mol Breed 23: 13–
92 22.
93 37. Wang Z, Triezenberg SJ, Thomashow MF, Stockinger EJ (2005) Multiple hydrophobic motifs
94 in *Arabidopsis* CBF1 COOH-terminus provide functional redundancy in trans-activation. Plant
95 Mol Biol 58: 543–559.
96 38. Novillo F, Alonso JM, Ecker JR, Salinas J (2004) CBF2/DREB1C is a negative regulator of
97 CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in
98 *Arabidopsis*. P Natl Acad Sci USA 101: 3985–3990.
99 39. Chinnusamy V, Zhu J, Zhu J-K (2007) Cold stress regulation of gene expression in plants.
100 Trends Plant Sci 12: 444–451.
101 40. Chinnusamy V, Ohta M, Kanrar S, Lee B, Hong X, et al. (2003) ICE1: a regulator of cold-
102 induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev 17: 1043–1054.
103 41. Chen M, Xu Z, Xia L, Li L, Cheng X, et al. (2009) Cold-induced modulation and functional
104 analyses of the DRE-binding transcription factor gene, *GmDREB3*, in soybean (*Glycine max*
105 L.). J Exp Bot 60: 121–135.
106 42. Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of *cis*-acting regulatory elements in
107 osmotic- and cold-stress-responsive promoters. Trends Plant Sci 10: 88–94.

108 43. Maruyama K, Sakuma Y, Kasuga M, Ito Y, Seki M, et al. (2004) Identification of cold-
109 inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using
110 two microarray systems. Plant J 38: 982–993.
111 44. Mondego JM, Vidal RO, Carazzolle MF, Tokuda EK, Parizzi LP, et al. (2011) An EST-based
112 analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica*
113 and *Coffea canephora*. BMC Plant Biol 11: 30.
114 45. Vidal RO, Mondego JMC, Pot D, Ambrósio AB, Andrade AC, et al. (2010) A high-throughput
115 data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags
116 suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. Plant
117 Physiol 154: 1053–1066.
118 46. Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, et al. (1999) Molecular
119 characterisation and origin of the *Coffea arabica* L. genome. Mol Gen Genet 261: 259–266.
120 47. Montagnon C, Leroy T, Yapo A (1992) Diversité génotypique et phénotypique de quelques
121 groupes de caféiers (*Coffea canephora* Pierre) en collection. Café Cacao Thé 36: 187–198.
122 48. Montagnon C (2000) Optimisation des gains génétiques dans le schéma de sélection récurrente
123 réciproque de *Coffea canephora* Pierre. PhD Thesis. Montpellier University, France.
124 49. Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, et al. (2008) Diversity in coffee assessed
125 with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome 51:
126 50–63.
127 50. Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2013) Global analysis of *Coffea canephora*
128 Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from
129 climatic refuges and migration effects. Genet Resour Crop Evol 60: 483–501.
130 51. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, et al. (2014) The coffee genome
131 provides insight into the convergent evolution of caffeine biosynthesis. Science, in press.
132 52. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces
133 using phred. I. Accuracy assessment. Genome Res 8: 175–185.
134 53. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error
135 probabilities. Genome Res 8: 186–194.
136 54. Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press.
137 55. Watterson GA (1975) On the number of segregating sites in genetical models without
138 recombination. Theor Popul Biol 7: 256–276.
139 56. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
140 polymorphism data. Bioinformatics 25: 1451–1452.
141 57. Chang W-C, Lee T-Y, Huang H-D, Huang H-Y, Pan R-L (2008) PlantPAN: Plant promoter
142 analysis navigator, for identifying combinatorial *cis*-regulatory elements with distance
143 constraint in plant gene groups. BMC Genomics 9: 561.
144 58. Shahmuradov IA, Solovyev V V, Gammerman AJ (2005) Plant promoter prediction with
145 confidence estimation. Nucleic Acids Res 33: 1069–1076.
146 59. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of
147 progressive multiple sequence alignment through sequence weighting, position-specific gap
148 penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
149 60. Mihara M, Itoh T, Izawa T (2010) SALAD database: a motif-based database of protein
150 annotations for plant comparative genomics. Nucleic Acids Res 38: D835–842.
151 61. Bromberg Y, Rost B (2007) SNAP: predict effect of non-synonymous polymorphisms on
152 function. Nucleic Acids Res 35: 3823–3835.
153 62. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific
154 phylogenies. Mol Biol Evol 16: 37–48.
155 63. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in
156 plant abiotic stress responses. Biochim Biophys Acta 1819: 86–96.
157 64. McKhann HI, Gery C, Bérard A, Lévêque S, Zuther E, et al. (2008) Natural variation in CBF
158 gene sequence, gene expression and freezing tolerance in the Versailles core collection of
159 *Arabidopsis thaliana*. BMC Plant Biol 8: 105.
160 65. Mboup M, Fischer I, Lainer H, Stephan W (2012) Trans-species polymorphism and allele-
161 specific expression in the CBF gene family of wild tomatoes. Mol Biol Evol 29: 3641–3652.

162 66. Giordani T, Buti M, Natali L, Pugliesi C, Cattonaro F, et al. (2011) An analysis of sequence
163 variability in eight genes putatively involved in drought response in sunflower (*Helianthus*
164 *annuus* L.). Theor Appl Genet 122: 1039–1049.
165 67. Cortés AJ, This D, Chavarro C, Madriñán S, Blair MW (2012) Nucleotide diversity patterns at
166 the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus*
167 *vulgaris* L.). Theor Appl Genet 125: 1069–1085.
168 68. Nakashima K, Yamaguchi-Shinozaki K (2006) Regulons involved in osmotic stress-responsive
169 and cold stress-responsive gene expression in plants. Physiol Plant 126: 62–71.
170 69. Mishra S, Shukla A, Upadhyay S, Sanchita, Sharma P, et al. (2013) Identification, occurrence,
171 and validation of DRE and ABRE *cis*-regulatory motifs in the promoter regions of genes of
172 *Arabidopsis thaliana*. J Integr Plant Biol 56: 388–399.
173 70. Agarwal M, Hao Y, Kapoor A, Dong C-H, Fujii H, et al. (2006) A R2R3 type MYB
174 transcription factor is involved in the cold regulation of CBF genes and in acquired freezing
175 tolerance. J Biol Chem 281: 37636–37645.
176 71. Yang Z, Bielawski J (2000) Statistical methods for detecting molecular adaptation. Trends Ecol
177 Evol 15: 496–503.
178 72. Lashermes P, Andrzejewski S, Bertrand B, Combes MC, Dussert S, et al. (2000) Molecular
179 analysis of introgressive breeding in coffee (*Coffea arabica* L.). Theor Appl Genet 100: 139–
180 146.
181 73. Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, et al. (2008) Diversity in coffee assessed
182 with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome 51:
183 50–63.
184

9. Main Figures List

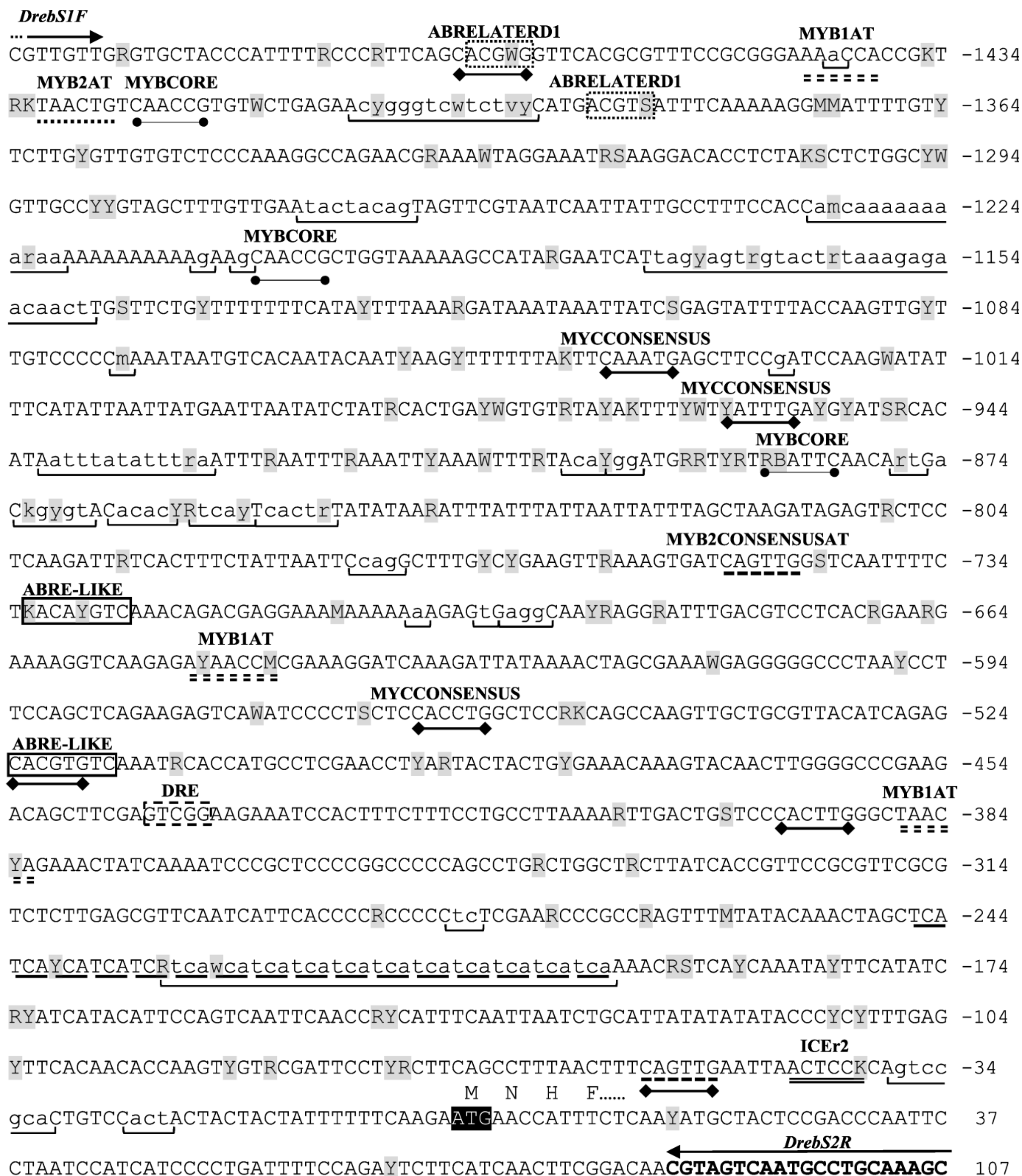


Figure 17 Promoter consensus sequence of *DREB1D* haplotypes and the observed *cis* regulatory elements acting in osmotic- and cold-stress response (see details in Table 4).
The promoter sequence represents a consensus sequence of all genotypes and the IUPAC code indicates the polymorphic sites (gray boxes). Black shading indicates the translation start site and the first base of this start codon corresponds to +1. Insertion sequences are indicated in lower case, deletions are represented by (˽) and (▁) indicates microsatellite core repeats. The ABRE-LIKE motif (white box), ABRELATERD1 (dotted box), ICEr2 motif (double underlined), MYBCORE (filled circled), MYB1AT motifs (double-dashed line), MYB2AT motif (dotted line), MYB2CONSENSUSAT motifs (dashed line), MYCCONSENSUS motifs (filled diamond) are also presented. This consensus sequence was deduced from the alignment of 98 coffee sequences (GB accession numbers from X to Y). The nucleotides are numbered (right) on each lane. Horizontal arrows indicate primers (see Table 2) used to amplify the coffee *DREB1D* promoter sequences.

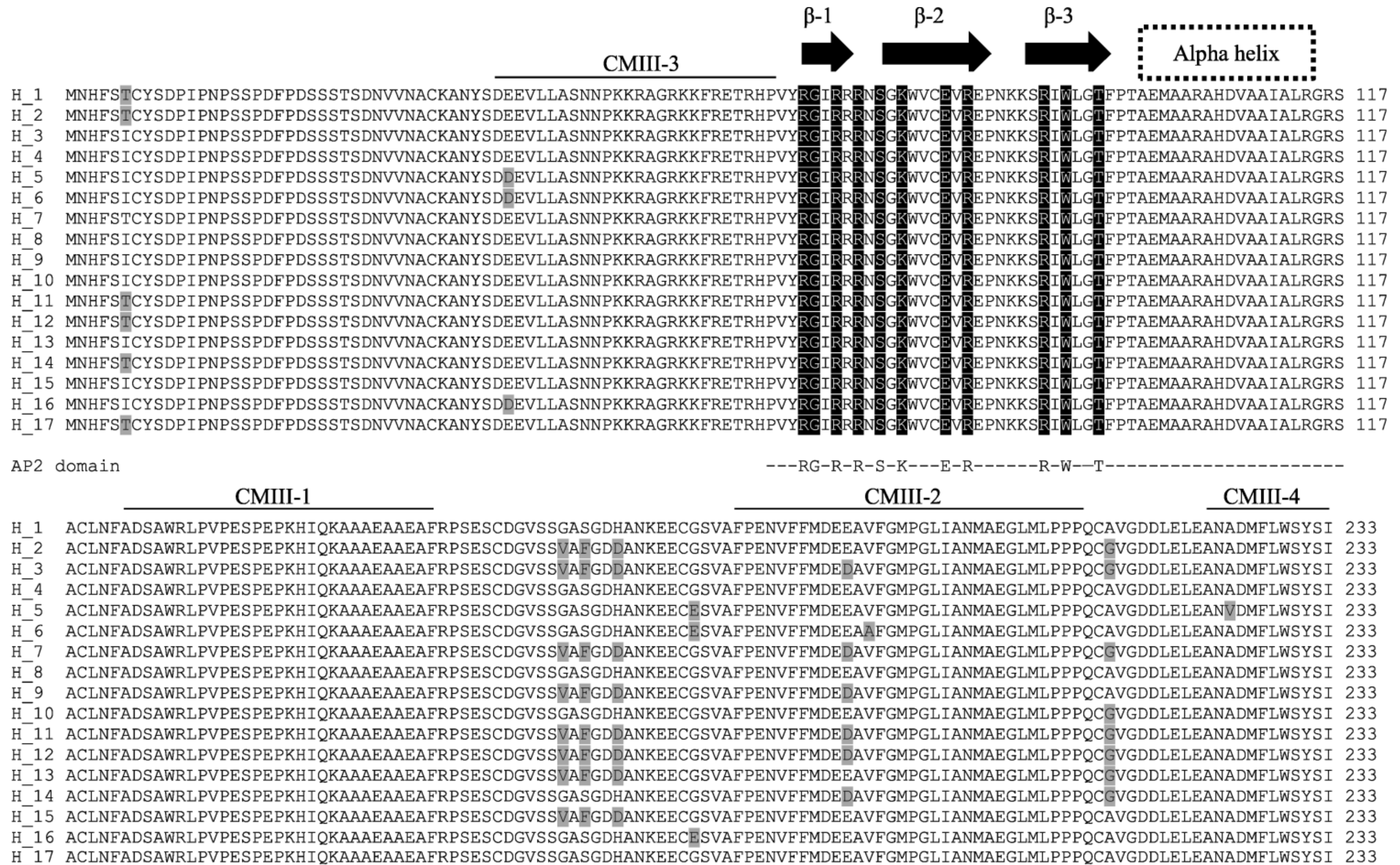


Figure 18 Alignment of the homologous predicted DREB1D protein sequences.

Divergent amino acids among DREB1D homologous protein sequences are box-shadowed in gray. Conserved amino acid positions of the AP2 domain are bolder, the arrows and dashed boxes indicate the secondary structure (β-sheet and α-helix). Motifs specific to DREB1 proteins are shown above the sequence.

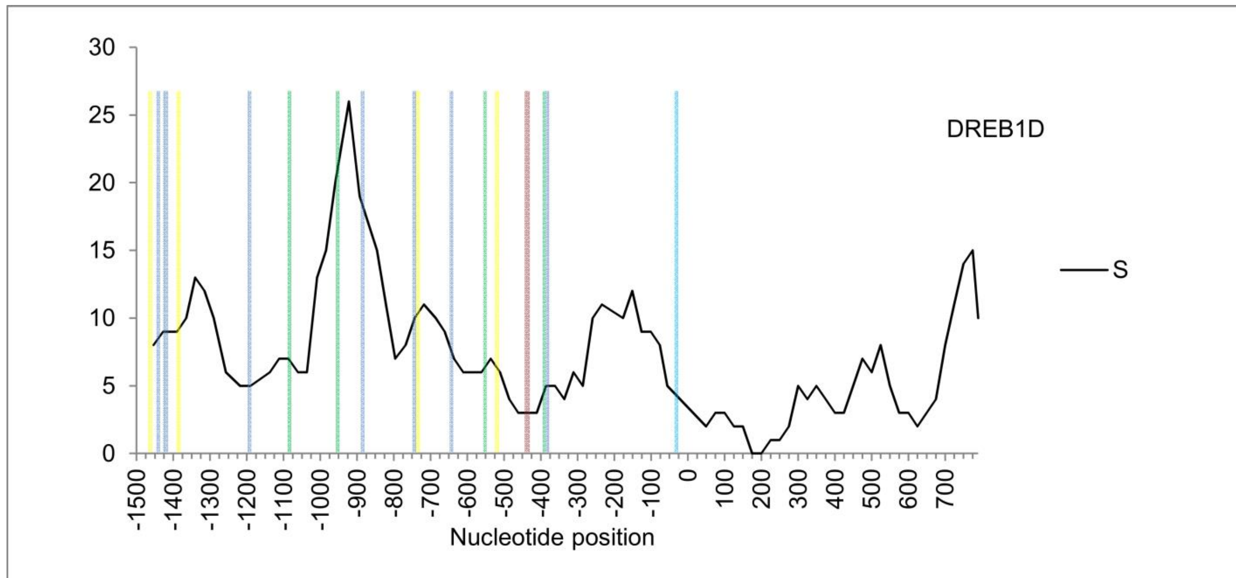


Figure 19 Polymorphic sites (S) frequency within the *DREB1D* gene among all *Coffea* genotypes analyzed.

The variability detected in the 1456 bp of the promoter region (-1 to -1456) and 738 bp of the coding region (+1 to +738) are presented based on a sliding window graphic (window size of 100 bp and steps of 25 bp). The shaded bars indicate the *cis*-regulatory elements in the promoter region of the *DREB1D* gene involved in drought-inducible gene expression (Table 4). The *Cis*-elements indicated are: ABRE (yellow), MYB (blue), MYC (green), DRE (red) and ICEr2 (light blue).

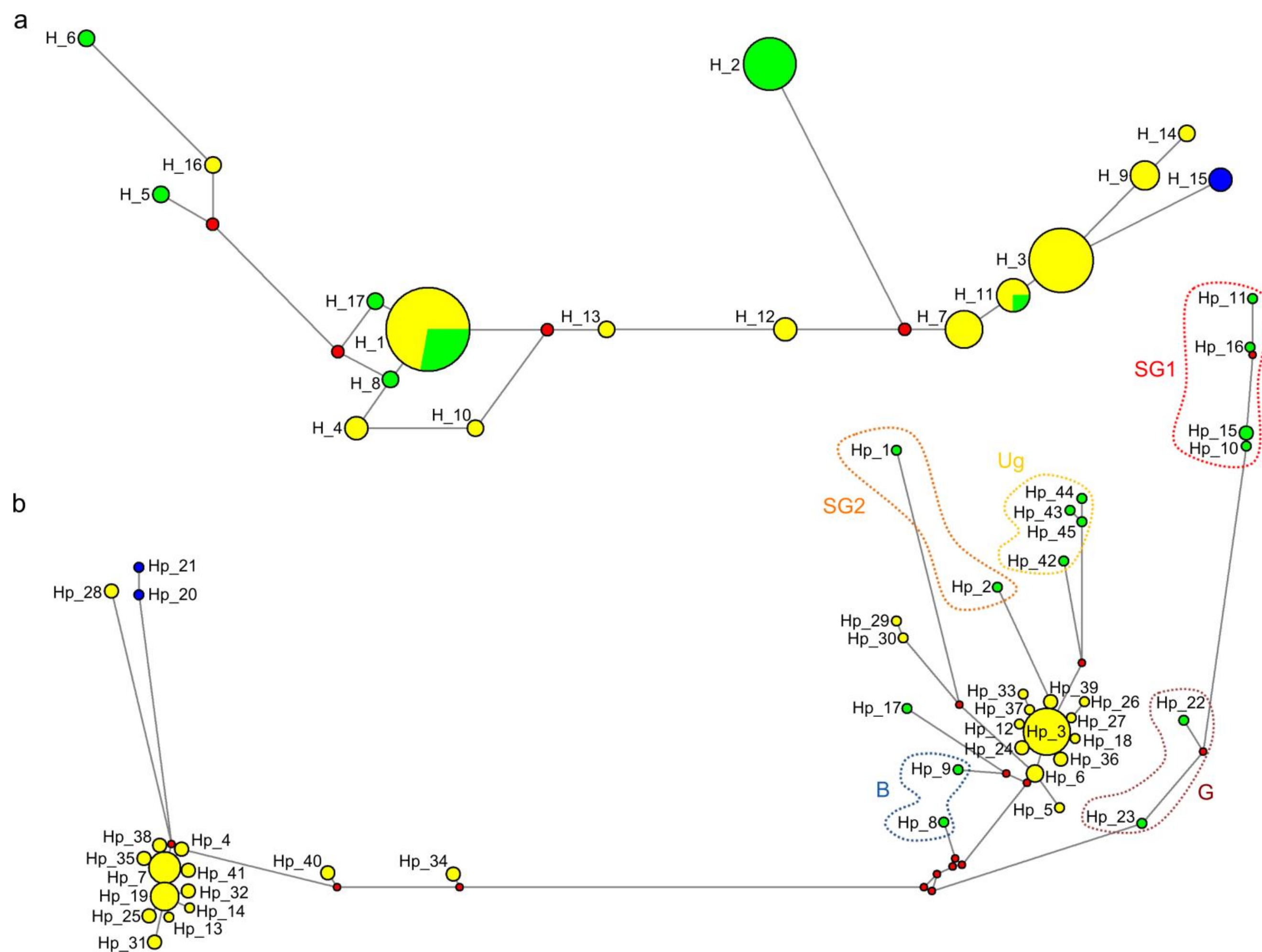


Figure 20 Haplotype network based on neighbor-joining analysis of the *DREB1D* coding (A) and promoter (B) sequences in *Coffea* accessions.

Haplotypes are represented by nodes and their size is proportional to the frequency. Lengths of node links are proportional to nucleotide distances between haplotypes. The accessions are colored according to species, *C. arabica* (yellow), *C. canephora* (green) and *C. eugenioides* (blue). Dashed circles indicate *C. canephora* groups and subgroups according to previous classification [2,50,73].

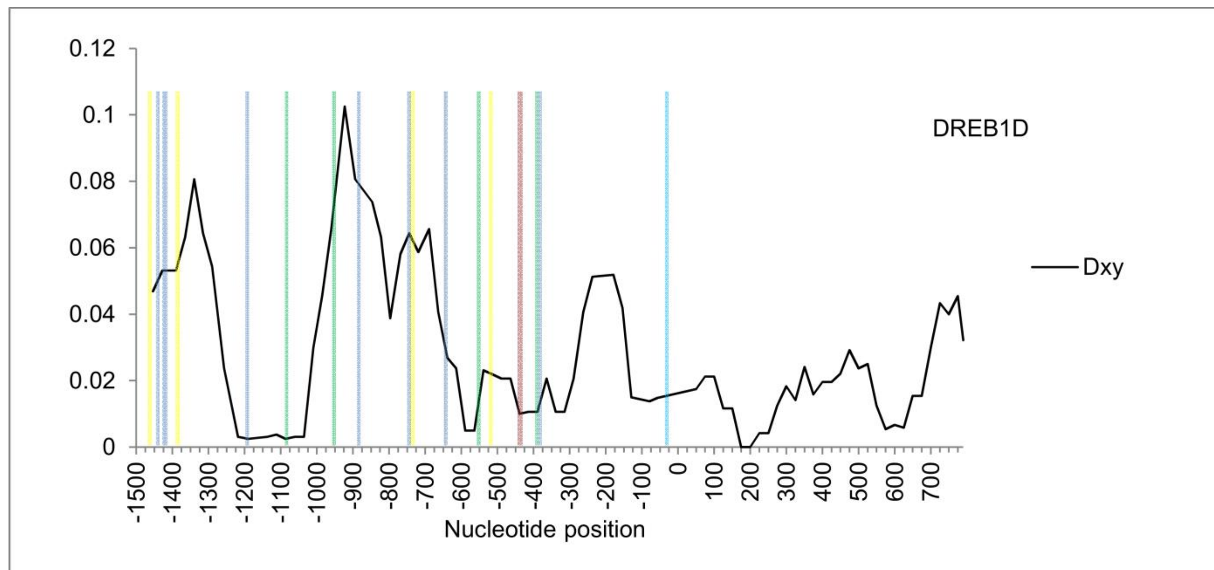


Figure 21 The average number of nucleotide substitutions per site (Dxy) of the *DREB1D* locus between *C. canephora* and *C. eugenioides*.

The variability detected in the 1456 bp of the promoter region (-1 to -1456) and 738 bp of the coding region (+1 to +738) are presented based on a sliding window graphic (window size of 100 bp and steps of 25 bp. The shaded bars indicate the *cis*-regulatory elements in the promoter region of the *DREB1D* gene involved in drought-inducible gene expression (see Table 4). The *Cis*-elements indicated are: ABRE (yellow), MYB (blue), MYC (green), DRE (red) and ICEr2 (light blue).

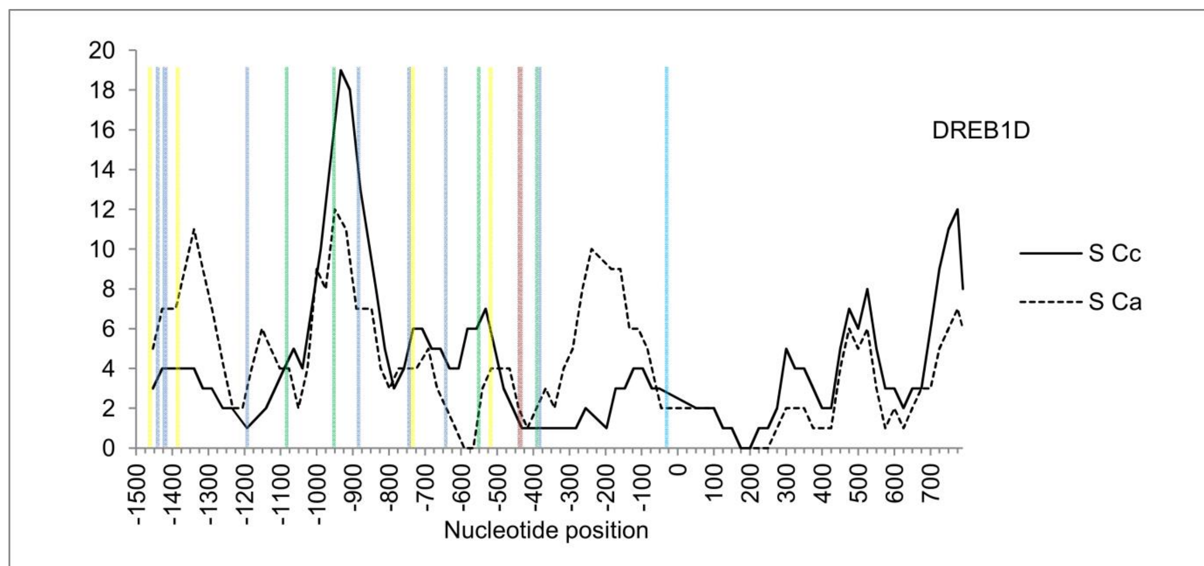


Figure 22 Polymorphic site frequency of the *C. canephora* (S_{Cc}) and *C. arabica* (S_{Ca}) *DREB1D* locus.

The variability detected in the 1456 bp of the promoter region (-1 to -1456) and 738 bp of the coding region (+1 to +738) are presented based on a sliding window graphic. The shaded bars indicate the *cis*-regulatory elements in the promoter region of the *DREB1D* gene involved in drought-inducible gene expression. The *Cis*-elements indicated are: ABRE-Like (yellow); MYB (blue) and MYC (green).

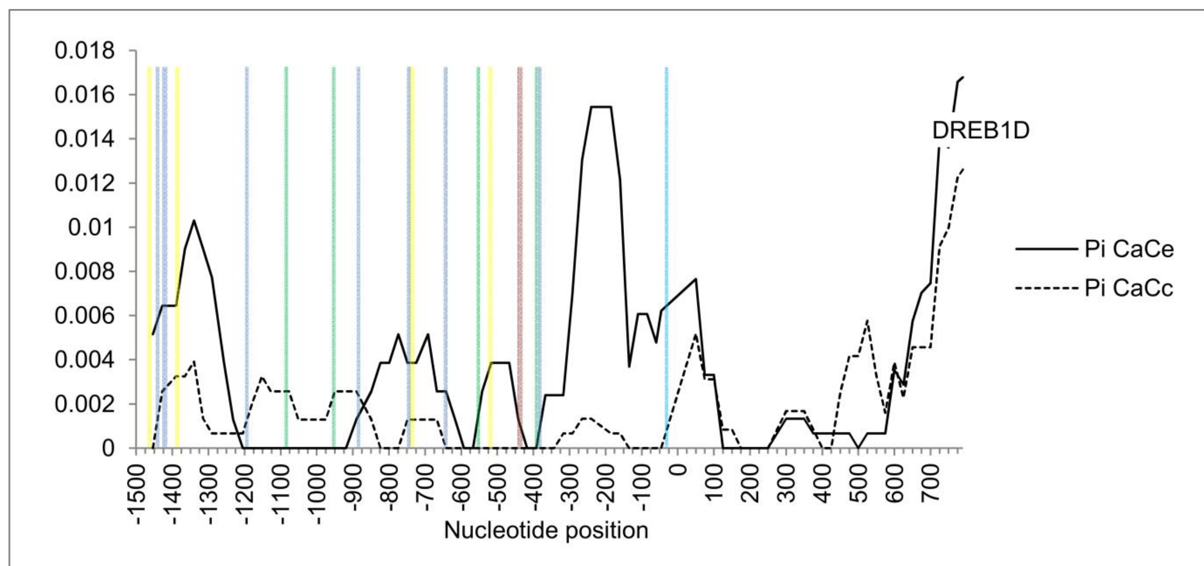


Figure 23 Nucleotide diversity (π) frequency within the *CaDREB1D* locus of the *CaCc* and *CaCe* subgenomes.

The variability detected in the 1456 bp of the promoter region (-1 to -1456) and 738 bp of the coding region (+1 to +738) are presented based on a sliding window graphic. The shaded bars indicate the *cis*-regulatory elements in the promoter region of the *DREB1D* gene involved in drought-inducible gene expression. The *Cis*-elements indicated are: ABRE (yellow), MYB (blue), MYC (green), DRE (red) and ICEr2 (light blue).

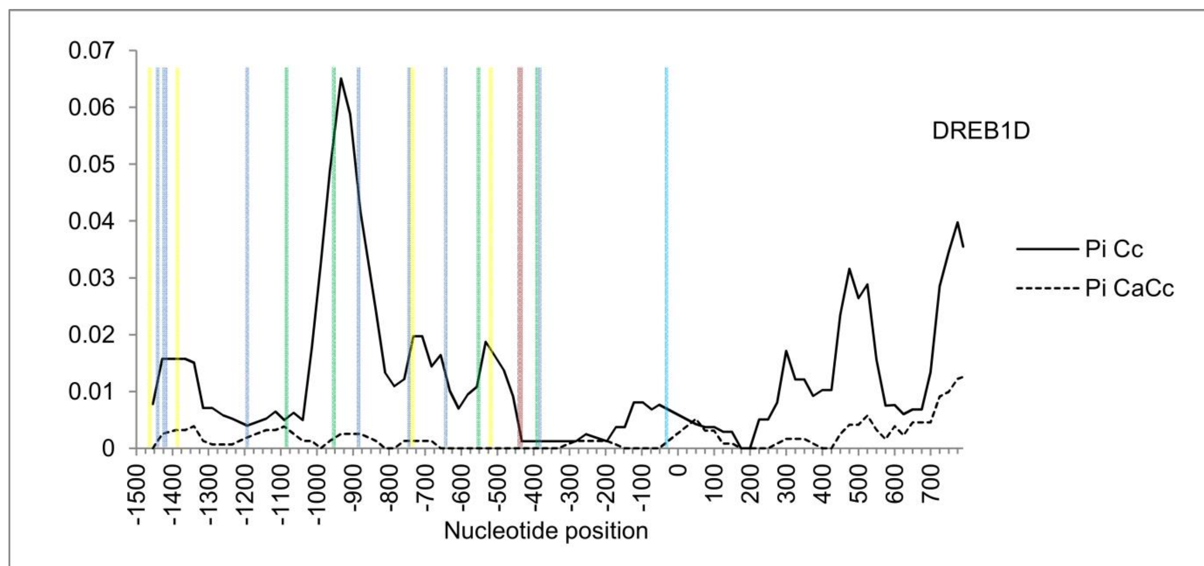


Figure 24 Nucleotide diversity (π) frequency within the *DREB1D* locus of the *CaCc* and Cc subgenomes.

The variability detected in the 1456 bp of the promoter region (-1 to -1456) and 738 bp of the coding region (+1 to +738) are presented based on a sliding window graphic. The shaded bars indicate the *cis*-regulatory elements in the promoter region of the *DREB1D* gene involved in drought-inducible gene expression. The *Cis*-elements indicated are: ABRE (yellow), MYB (blue), MYC (green), DRE (red) and ICEr2 (light blue).

10. Tabels

Table 1 Genotypes from the *Coffea* genus used for nucleotide diversity analyses of *DREB1D* promoter and protein coding sequences.

Genotype	P	Promoter	CDS	Species	Diversity Group / pedigree	I*
1. Acaiá 47119	4x	√		*C. arabica*	[2]	
2. Bourbon	4x	√	√	*C. arabica*	[2]	
3. C1007	2x	√	√	*C. canephora*	Congolese B[4]	
4. C2011	2x	√	√	*C. canephora*	Congolese SG2[4]	
5. C3001	2x	√	√	*C. canephora*	Congolese SG1[4]	
6. C4001	2x		√	*C. canephora*	Congolese C[4]	
7. Canephora	2x		√	*C. canephora*	Conilon (Congolese SG1)[2]	
8. Catuai 25	4x	√	√	*C. arabica*	Caturra x Mundo Novo[2]	
9. Catuai 144	4x	√		*C. arabica*	Caturra x Mundo Novo[2]	
10. Clone 14	2x	√	√	*C. canephora*	Conilon (Congolese SG1)[1]	
11. Clone 22	2x	√	√	*C. canephora*	Conilon (Congolese SG1)[1]	
12. Clone 46	2x		√	*C. canephora*	Conilon (Congolese SG1)[1]	
13. Eugenioides	2x	√	√	*C. eugenioides*	Not defined[3]	
14. E007	4x		√	*C. arabica*	Ethiopia[3]	
15. E017	4x		√	*C. arabica*	Ethiopia[3]	
16. E123A	4x		√	*C. arabica*	Ethiopia[3]	
17. E237	4x	√	√	*C. arabica*	Ethiopia[3]	
18. E238	4x		√	*C. arabica*	Ethiopia[3]	
19. E464	4x		√	*C. arabica*	Ethiopia[3]	
20. E516	4x		√	*C. arabica*	Ethiopia[3]	
21. G2020	2x	√	√	*C. canephora*	Guinean[4]	
22. Guatemalense Baixo	4x	√		*C. arabica*	[2]	
23. Guatemalense Alto	4x	√		*C. arabica*	[2]	
24. Iapar 59	4x	√	√	*C. arabica*	Villa Sarchi x Timor Hybrid (H 361/4)[3]	√
25. Icatú Colombiano	4x	√		*C. arabica*	*C. canephora* cv Robusta (4x) x *C. arabica* Bourbon Vermelho[2]	√
26. Mundo novo	4x	√	√	*C. arabica*	Sumatra x Bourbon vermelho[2]	
27. Obatã	4x	√		*C. arabica*	(H 361/4) x Catuai vermelho[2]	√
28. Palma 02	4x	√		*C. arabica*	Catuaí vermelho x Catimor[2]	√
29. Purpurenses	4x	√		*C. arabica*	[2]	
30. Rubi	4x	√	√	*C. arabica*	Catuai x Mundo Novo[2]	√
31. Sabiá	4x	√		*C. arabica*	Catimor x Acaia[2]	√
32. San Bernardo	4x	√		*C. arabica*	*C. arabica* var. Typica (Brazil [2]	
33. San Ramon Baixo	4x	√		*C. arabica*	*C. arabica* var. Typica (Brazil)[2]	
34. CPAC	4x	√		*C. arabica*	[2]	
35. Tupi	4x	√		*C. arabica*	Villa Sarchi x Timor Hybrid[2]	√
36. Typica	4x	√	√	*C. arabica*	*C. canephora* x *C. eugenioides*[2]	
37. UW002	2x	√	√	*C. canephora*	Congolese UW[5]	
38. UW099	2x	√	√	*C. canephora*	Congolese UW[5]	

P defines ploidy levels of the coffee plants analyzed
CDS corresponds to the nucleotide sequence of the gene coding the DREB1D protein.

* *C. arabica* hybrid introgressed with *C. canephora*.
[1] Clones of *C. canephora* provided by INCAPER.
[2] Clone of *C. canephora* provided by EMBRAPA Cerrados
[3] Clone of *C. arabica* provided by IAPAR
[4] Clones of *C. canephora* provided by CNRA
[5] Clones of *C. canephora* provided by NARO-COREC

Table 2 List of selected primers used for *DREB1D* diversity analysis.

Primer	Sequence
DrebS1F [(a,c)]	5’–ACTCCTAGTAAGCGGCACGTTGTT–3’
DrebS1R [(a)]	5’–CCTTTCGTGGTTGTCTCTTGACCT–3’
DrebS2F [(b)]	5’–TCGTGCATTCAACAGCACCGTCA–3’
DrebS2R [(b,c)]	5’–TGGCTTTGCAGGCATTGACTACG–3’
Dreb1AF [(d)]	5’–GTTGAATTAACTCCTCACTGTCCACTA–3’
Dreb1AR [(d)]	5’–CCAAAAACTGCAGTACGGAATAGA–3’

[a] Primers used to amplify the 5’-UTR and the first 660 bp of the promoter region.
[b] Primers used to amplify the last 884 bp of the promoter region.
[c] Primers used to amplify the complete promoter region.
[d] Primers used to amplify the coding region with 3’-UTR.

Table 3 Patterns of nucleotide variability in the coffee *DREB1D* locus

Sample	Nr. of genotypes	Region	Nr. of haplotypes	Sequence Length	S	π	θ	Tajima's D	Syn. Subs.	Nonsyn. Subs.	πa/πs
Total	29	Promoter	37	1503	115	nc	nc	nc	na	na	na
	25	CDS	17	699	23	nc	nc	nc	13	10	0.282
	25	3' UTR	9	130	16	nc	nc	nc	na	na	na
C. arabica	20	Promoter	23	1401	69	nc	nc	nc	na	na	na
	13	CDS	11	699	14	nc	nc	nc	6	8	0.361
	13	3' UTR	5	130	8	nc	nc	nc	na	na	na
C. canephora	8	Promoter	14	1364	61	0.01238	0.0137	-0.4112	na	na	na
	11	CDS	7	699	21	0.01006	0.00805	0.9155	11	10	0.268
	11	3' UTR	5	130	12	0.03057	0.02472	0.8126	na	na	na

na : not applicable
nc : not calculated
Promoter corresponds to the nucleotide sequence upstream of first translated nucleotide +1, including the 5' UTR
CDS corresponds to the nucleotide sequence of the gene coding to DREB1D protein.

Table 4 *Cis*-regulatory elements involved in osmotic- and cold-stress responsive gene expression identified using PlantPAN analysis and experimentally supported results from previously published reports.

	cis element	Sequence	TFs that bind to *cis* elements	Gene	Stress condition	Reference
ABRE	ABRELATERD1	ACGTG	bZIP	*erd1*	Water deficit, ABA	PLACE
	ABRE-like	BACGTGKM	bZIP	*rd29*	Water deficit, ABA	AGRIS
	ACGTABRE	ACGTGKC	bZIP	*osem*	ABA	PLACE
DRE	DRE	CCGNC	ERF/AP2	*rd29A*	Water deficit, cold	[18]
	LTRECOREATCOR15	CCGAC	ERF/AP2	*cor15A*	Water deficit, cold	PLACE
MYB	MYBCORE	CNGTTR	bHLH	*rd22*	Water deficit, ABA	PLACE
	MYB1AT	WAACCA	bHLH (MYB15)	*rd22*	Water deficit, ABA	PLACE
	MYB2AT	TAACTG	bHLH	*rd22*	Water deficit, ABA	PLACE
	MYB2CONSENSUSAT	YAACKG	bHLH	*rd22*	Water deficit, ABA	PLACE
MYC	MYCCONSENSUSAT	CANNTG	bHLH	*rd22*	Water deficit	PLACE
ICE	ICEr2	ACTCCG	Not known	*CBF2/ DREB1C*	Cold	[42]

Table 5 Haplotype information for the *DREB1D* locus across the 39 genotypes analyzed. The haplotypes detected in the coding sequence and promoter regions are indicated respectively as H and Hp. Based on haplotype clustering, putative *CaCe* haplotypes are shaded.

Genotype	Ploidy	Haplotype	
		CDS	Promoter
1. Acaiá 47119	4x		Hp_3; Hp_4
2. Bourbon	4x	H_1; H_3; H_4	Hp_5; Hp_6; Hp_7
3. C1007	2x	H_2; H_5	Hp_8; Hp_9
4. C2011	2x	H_1; H_2	Hp_1; Hp_2
5. C3001	2x	H_2	Hp_10; Hp_11
6. C4001	2x	H_2; H_6	
7. Canephora	2x	H_1; H_2	
8. Catuai 25	4x	H_1; H_4; H_7	Hp_3; Hp_7
9. Catuai 144	4x		Hp_3; Hp_12; Hp_13; Hp_14
10. Clone 14	2x	H_2;	Hp_15; Hp_16
11. Clone 22	2x	H_2; H_8	Hp_15; Hp_17
12. Clone 46	2x	H_1; H_2	
13. Eugenioides	2x	H_15	Hp_20; Hp_21
14. E007	4x	H_1; H_3; H_7; H_9	
15. E017	4x	H_1; H_10; H_11; H_12	
16. E123A	4x	H_1; H_3	
17. E237	4x	H_1; H_11	Hp_3; Hp_18; Hp_19
18. E238	4x	H_1; H_3	
19. E464	4x	H_1; H_3	
20. E516	4x	H_3; H_12; H_13; H_14	
21. G2020	2x	H_1; H_11	Hp_22; Hp_23
22. Guatemalense Baixo	4x		Hp_26; Hp_27; Hp_28
23. Guatemalense Alto	4x		Hp_24; Hp_25
24. Iapar 59	4x	H_1; H_3; H_7; H_9	Hp_3; Hp_19
25. Icatú Colombiano	4x		Hp_29; Hp_30; Hp_31
26. Mundo novo	4x	H_1; H_3	Hp_3; Hp_32
27. Obatã	4x		Hp_3; Hp_19
28. Palma 02	4x		Hp_3; Hp_33; Hp_34
29. Purpurenses	4x		Hp_3; Hp_7
30. Rubi	4x	H_1; H_3; H_9; H_16	Hp_3; Hp_35
31. Sabiá	4x		Hp_7; Hp_36
32. San Bernardo	4x		Hp_3; Hp_7; Hp_37
33. San Ramon Baixo	4x		Hp_3; Hp_38
34. CPAC	4x		Hp_39; Hp_40
35. Tupi	4x		Hp_6; Hp_19
36. Typica	4x	H_1; H_3; H_7	Hp_3; Hp_41
37. UW002	2x	H_1; H_17	Hp_42; Hp_43
38. UW099	2x	H_1	Hp_44; Hp_45

Table 6 Summary of variability in the two *Coffea arabica* subgenomes.

Sample	Nr. of genotypes	Region	Nr. of haplotypes	Sequence Length	S	π	θ	Tajima's D	Syn. Subs.	Nonsyn. Subs.	πa/πs
CaCe	30	Promoter	6	1425	35	0.00385	0.0062	-1.3869	na	na	na
	31	CDS	7	699	10	0.00251	0.00358	-0.94	7	3	0.27
	31	3' UTR	5	130	8	0.01386	0.0154	-0.3017	na	na	na
CaCc	30	Promoter	4	1420	13	0.00096	0.00231	-1.9304*	na	na	na
	24	CDS	6	699	13	0.00258	0.00498	-1.6747	5	8	0.469
	24	3' UTR	3	130	6	0.00942	0.01236	-0.7192	na	na	na

Significant results: *P > 0.05.
na (not applicable)

CHAPTER 2

CcDREB1D as part of the coffee *DREB* gene family: difficulties in extrapolating functional information from *Arabidopsis* to coffee.

CcDREB1D as part of the coffee DREB gene family: difficulties in extrapolating functional information from Arabidopsis to coffee.

1. Introduction

CcDREB1D, presenting a high similarity with *AtDREB1D* (At5g51990), has been identified as a potential candidate for drought tolerance in coffee and its diversity studied specifically in the previous chapter. Several models have been proposed in *Arabidopsis* in order to explain the relative involvement of *DREB1D* and other members of this gene family in the response to several abiotic stresses (Nakashima et al., 2009; Mizoi et al., 2012), showing some signs of subfunctionalization between the different members of this family. Extrapolating functional information from such a model plant to coffee is therefore tempting. However, orthology relationships need to be clearly established before extrapolating any functional clues from one species to the other one (Koonin, 2005; Kuzniar et al., 2008). Indeed, duplications arising during plant genomes evolution and speciation lead to paralogous sequences which can rapidly evolve and either degenerate, acquire some sort of subfunctionalization or neofunctionalization (Force et al., 1999). Therefore, orthologs between species (i.e. genes arising from a speciation event) have more chance to conserve the same function than paralogs.

The recent completion of the coffee genome (Denoeud et al., 2014) and availability of the integrative genome information system, Coffee Genome Hub (http://coffee-genome.org), that allows access to genomics and genetics data and analysis tools, has provided us with some opportunity to identify, manually check and retrieve an almost complete set of *DREB* gene family in coffee. As an inevitable step, the determination of specific biological function of each gene will require a great deal of experimental work. On the basis of phylogenetic analyses, an assessment of the structural and orthology relationships with other species, for which functional information is available, can allow us to infer relationships and provide a guide for predicting the functions of these genes. In this chapter, the establishment of a complete overview of coffee *DREB* gene family in *Coffea canephora* was attempted. To accomplish this, genes in the DREB family in the *C canephora* genome were surveyed.

Phylogenetic analyses were performed, as well as exon/intron and protein motif structural analyses of the DREB family genes. Obtained results were used to accurate the automatic annotation. Genes encoding proteins of DREB family in orthologue species were sought from genomic databases, and comparative analyses of the phylogeny between *C. canephora* and related species, along with comparative analyses of conserved motifs in *C. canephora* and *Arabidopsis* DREB families were performed.

More precisely the objectives of this chapter were to:

(i) identify the (almost) complete set of coffee *DREB* gene family, and its homologs in different species with complete genomes and eventually functional information;

(ii) correct the automatic annotation generated by each potential gene sequence, since small intronless genes are often misannotated and present spurious introns (Yang and Cho, 2012);

(iii) align the whole set of DREB protein sequences from coffee, *Arabidopsis* and other species and draw a robust phylogenetic tree for this family;

(iv) Try to infer orthology relationships for coffee DREBs and putative functional information, in particular for our candidate CcDREB1D protein.

2. Materials and Methods

Sequence data retrieval

Multiple database searches were performed to identify members of the *C. canephora* DREB family.

First, clusters of orthologous DREB protein sequences from *A. thaliana* were identified using the GreenPhyl v4 database (http://www.greenphyl.org /cgi-bin/index.cgi, (Rouard et al., 2011). GreenPhyl is a web resource for comparative and functional genomics which provides information on gene families and gene homology-relationships (orthologs and paralogs) using a tree reconciliation approach. Reconciliation consists in rooting the phylogenetic tree and the species tree to seek for evolutionary incongruence and infer the processes that might lead to it, such as gene duplication and loss, lateral transfer, recombination, and incomplete lineage sorting. In GreenPhyl, gene families are

clustered in different levels relative to sequence similarity from the less stringent (Level 1) to the most stringent (Level 4). Coffee proteins were extracted from DREB subfamily cluster GP000017 (Level 1) and aligned with MAFFT software (Katoh et al., 2002). Protein sequence from 39 *CcDREB* genes were further manually curated in the Coffee Genome Hub (http://coffee-genome.org/).

In a second phase, in an attempt to extend the set of CcDREB subfamily candidates, a homology-search based on profile Hidden Markov Model (HMM) was carried out on coffee proteomic data using a Galaxy platform (Giardine et al., 2005) available on the South Green bioinformatics platform (http://www.southgreen.fr). HMMs apply a statistical method to estimate the true frequency of a residue at a given position in the alignment from its observed frequency while standard profiles use the observed frequency itself to assign the score for that residue. This means that a profile HMM derived from only 10 to 20 aligned sequences can be of equivalent quality to a standard profile created from 40 to 50 aligned sequences. Protein sequences of AtDREB subfamily genes were downloaded from The *Arabidopsis* Information Resource (TAIR-www.*Arabidopsis*.org) and separated into four subgroups (I-IV) according to Nakano et al 2006 classification. The sequences were further aligned by subgroup using MAFFT and the reliable parts were filtered with Gblocks (Castresana 2000). This strategy allowed building a high stringent HMM subgroup-specific profile which retained not only the AP2/ERF motif but also the outside motifs conserved in each subgroup. The subgroup-specific HMM profiles were then used to retrieve homologous protein sequences in the coffee proteomic data. Further, we generate a second HMM profile with the retrieved coffee sequences that was used as seed model for homology-based search in *A. thaliana, S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays* databases. The coffee HMM profile increase the stringency of homologous search between coffee and homologous sequences compared with the *Arabidopsis* HMM profile. All steps were conducted and combined in a workflow generated in Galaxy and using existing tools implemented in the Southgreen Galaxy platform (http://gohelle.cirad.fr/galaxy/root).

The Greenphyl and HMM profile sets were combined and redundancy removed to produce a final set of 407 homologous sequences.

Phylogenetic analysis of the AP2 domain from putative AP2/ERF genes

All 407 homologous protein sequences were aligned using MAFFT and curated by Gblocks, in order to keep only the reliable part of alignments. Phylogenetic trees were constructed with PhyML software version 3.0 (Guindon et al., 2009) implementing a maximum-likelihood algorithm under the LG evolution model (Le and Gascuel, 2008). The gene tree was reconciliated with *Viridiplantea* tree (http://www.ncbi.nlm.nih.gov /sites/entrez?db=taxonomy) using RAP-Green software (Dufayard et al., 2005). Branch supports were rated by the approximate likelihood ratio test with a Shimodaira–Hasegawa- like procedure (Guindon et al., 2010). The tree was scratched in Dendroscope (Huson et al., 2007).

3. Results

Phylogenetic analysis of CcDREB Genes

a. Cluster analysis through GreePhylDB

In order to identify genes encoding DREBs in *C. canephora* reference database (Coffee Genome Hub), multiple homology-based searches were first performed for coffee genes encoding AP2/ERF TFs using *Arabidopsis* homologous sequences in GreenPhyl data base (Rouard et al., 2011). As a result, 78 coffee sequences harbouring the conserved AP2/ERF domain were retrieved in the level 1 cluster GP000017 (Figure 1). Within the 78 sequences, we identified 20 coffee DREB proteins, based upon the phylogenetic classification of this superfamily in *Arabidopsis* which present the conserved residue V(14), E(19) and a WLG motif. Additionally, 19 potential DREB sequences were found harbouring a semi-conserved motif – only V(14) and WLG motif conserved – already reported in some DREB protein. Our candidate gene (*CcDREB1D*) was identified as *Cc02g034430*, present on chromosome 2 in tandem with another *DREB* gene, *Cc02g034320*.

To improve the alignment, the 39 coffee sequences were manually curated in the Coffee Genome Hub, using gnpannot standard rules (Sidibe-Bocs, personal communication).

Compared to the initial automatic annotation, spurious introns were removed in eight *DREB* sequences out of 39, whereas 2 introns were maintained, in the genes *Cc04g04250* and *Cc04g04880*. Two putative gene sequences *Cc10g14150* and *Cc10g14160* were fused, two amino acids sequences were slightly modified in the putative protein sequences compared to the original ones, adding or deleting a few aminoacids. UTR limits were also extended by gene / mRNA comparison for much of the genes. All data were integrated into the coffee genome browser (http://coffee-genome.org/gbrowse).

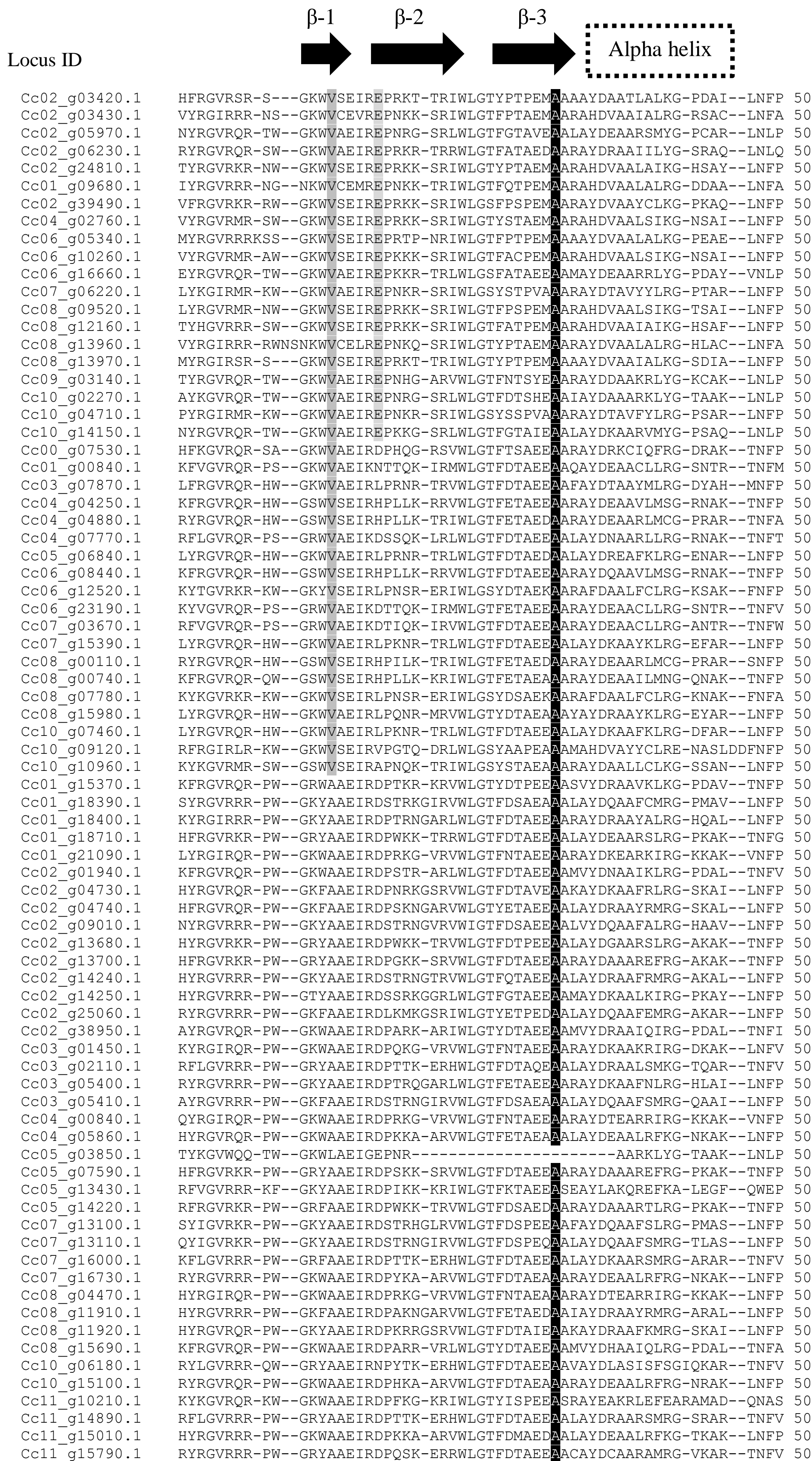


Figure 1 Alignment of the AP2/ERF domains from coffee protein sequences belonging to ERF family retrieved from GreenPhyl. Black and light gray shading indicate identical and conserved amino acid residues, respectively. The dotted box and black arrows Figure 1 (Cont.) represent predicted α-helix and β-sheet regions, respectively, within AP2/ERF domain. DREB family members have the amino acids V (14) and A (38) conserved.

b. HMM search for DREB homologs in coffee and other species

To better characterize the *CcDREB* subfamily members a second homology-based search by Hidden Markov model (HMM) was conducted in the Coffee Genome Hub. Moreover, *DREB* orthologous genes from *Arabidopsis*, grape, maize, potato, rice and tomato were also recovered.

59 putative *ERF/DREB* genes were identified by HMM search in the coffee genome. Only one sequence identified here was not identified by cluster analysis. In total, 407 protein sequences were retrieved from all homology-based search, of those 59 from coffee, 64, 46, 56, 61, 62 and 58 from *Arabidopsis*, grape, maize, potato, rice and tomato, respectively.

We next constructed a phylogenetic tree to access the homology relationship of DREB subfamily homologous members and infer gene function. Gene function inference can be greatly improved by reconciliation of gene tree and species tree to search for gene duplication. In this process, an overlap of gene tree and species tree is generated and topological disagreements are presumed to be a result of gene duplication and losses. This approach allows a better interpretation of the evolutionary relationship of homologous genes. The reconciled tree showed a clear separation of the DREB subgroups I, II, III and IV, except a small subnetwork from subgroup II showed close similarity to subgroup III (Figure 2). In accordance with the classification described by Nakano *et al.* (2006) all *Arabidopsis* DREB subfamily members maintained phylogenetic clustering characteristics. The absence of disruption on phylogenetic tree structure suggested that the classification of the *Arabidopsis* DREB family is similar and applicable to coffee DREB family. Based on phylogenetic analyses, we described 31 orthologous coffee genes as members of *DREB* subfamily (Figure 2) of which five belongs to subgroup I (*Cc03g07870*, *Cc05g06840*, *Cc07g15390*, *Cc08g15980*, *Cc10g07460*), six to subgroup II (*Cc06g12520*, *Cc07g06220*, *Cc08g07780*, *Cc10g04710*, *Cc10g09120*, *Cc10g10960*), twelve are clustered in subgroup III (*Cc01g09680*, *Cc02g03420*, *Cc02g03430*, *Cc02g24810*, *Cc02g39490*, *Cc04g02760*, *Cc06g05340*, *Cc08g09520*, *Cc06g10260*, *Cc08g12160*, *Cc08g13960*, *Cc08g13970*), and eight in subgroup IV (*Cc02g05970*, *Cc02g06230*, *Cc06g00780*, *Cc06g16660*, *Cc09g03140*, *Cc10g02270*, *Cc10g14150-Cc10g14160 (fused)*). Cc02g034430 was confirmed as a member of group III.

Orthology relationships of Cc DREBs with other species

Considering the evolutionary context in these analyses, 13 *DREBs* from *C. canephora* displayed pairwise correspondences with at least two dicot species (green circle node) (Figure 2). A closer relation was observed specially between *C. canephora*, *S. lycopersicum* and *S. tuberosum* DREB proteins, suggesting that *DREB* subfamily members could be phylogenetically conserved. Indeed, homologous that are most closely related through a speciation (orthologues) are more suitable to have similar function. It is thus likely that *Solanum* orthologs could provide an interesting information source of *CcDREB* functionality. However, the limited accurate information available on crop species turns inevitable the use of reviewed descriptions of model plants. Thus, we focused mainly on orthologous genes from *Arabidopsis* to infer gene function and provide a more accurate annotation of *CcDREB* genes.

Out of the members that composed the *DREB* subgroup IV, four (*Cc06g16660*, *Cc10g02270*, *Cc09g03140* and *Cc02g06230*) are closely pairwise orthologs of *Arabidopsis ABI4, DREB2D DREB2G* and *DREB2F* respectively, with high bootstrap support, and therefore one could expect them to present a conserved function (Figure 3).

Belonging to subgroup II, three ethylene-responsive transcription factors coding genes were identified with a clear orthology relationship with *Arabidopsis DREB* sequences: *AtERF13* vs *Cc10g09120 AtERF016* vs *Cc08g07780* and *AtRAP2.1* vs *Cc10g04710*. (Figure 4)

Subgroup I also contains two predicted ERFs with a clear orthology relationship with *Arabidopsis* DREB sequences, At*ERF062* vs *Cc03g07870* and *AtERF060* vs *Cc07g15390* (Figure 5).

Compared to the subgroup IV, proteins in the subgroup III are more phylogenetically divergent, with considerable number of recent duplication (Figure 6). The high occurrence of paralogs in subgroup III, reduce the accuracy of gene function prediction by homology. Even if *Cc02g03430* is more closely related to *AtDREB1D/CBF4* than to *AtDREB1A/CBF3*, *AtDREB1B/CBF1*, *AtDREB1C/CBF2*, and was therefore named *CcDREB1D*, the orthology relationship between *Arabidopsis* and coffee sequences cannot be clearly established, the 4 *Arabidopsis* sequences being considered as co-orthologs to

Cc02g034430. Another *CcDREB1/CBF* member identified was the *Cc08g13960* which has high similarity to *AtDREB1E/DDF1* and *AtDREB1F/DDF2*.

Apart from the *CcDREB1\CBF* group, we also described *Cc01g09680*, (*Cc06g10260*, *Cc08g09520*, *CcERF1 Cc08g13970* and *Cc02g39490* as the closely related orthologs of *SlERF10*, *SlERF4*, *AtTINY*, *SlERF1* and *AtERF023*, respectively.

Conserved Motifs outside of the AP2/ERF Domain

Conserved regions outside the DNA-binding domain can play a major role in transcriptional activity, protein-protein interactions, and nuclear localization (Liu et al., 1999). This functional domains, or amino acid sequence motifs, are likely to be shared between members of a subgroup in large families of transcription factors in plants (Kranz et al., 1998; Lijavetzky et al., 2003; Ooka et al., 2003; Reyes et al., 2004; Tian et al., 2004). Therefore, proteins in the same subgroup sharing these motifs have an increased probability to present equivalent functions. Indeed, multiple alignment analysis with ClustalW identified conserved motifs outside the DNA-binding domain in *CcDREB* predicted proteins sequences from related subgroups. The selectively occurrence of motifs within the specific clades in *CcDREB* gene family evidenced structural similarities among proteins within the same group. To confirm classifications and to analyse the phylogenetic relationships between *CcDREBs* and *AtDREBs*, similar analysis were performed in *Arabidopsis* and coffee DREB proteins from related subgroups to investigate the occurrence of these motifs. Multiple alignment analysis revealed sequence similarity among *CcDREB* and *AtDREB* genes not restricted to AP2/ERF motif. *CcDREB* and *AtDREB* genes shared in several motifs previously identified in *Arabidopsis* DREB family (Nakano et al., 2006), thus to facilitate the comprehension, we use the previously assigned names in *Arabidopsis*.

Despite the limited knowledge on the molecular function of these conserved motifs, some are expected to be main actors in proteins activities. The presence of ERF-associated amphiphilic repression (EAR: LxLxLx or DLNxxP) motif which has been described as a c-terminal functional domain in repressor-type ERF proteins and also of TFIIIA-type C2H2 zinc-finger proteins (Ohta et al., 2001) strongly suggest the repressor activity . Interestingly, this motif is similar to CMII-2 motif found in members of

group II (Figure 4B). In addition, we found several other conserved motifs in high homolog genes between *Arabidopsis* and coffee, however they still uncharacterized.

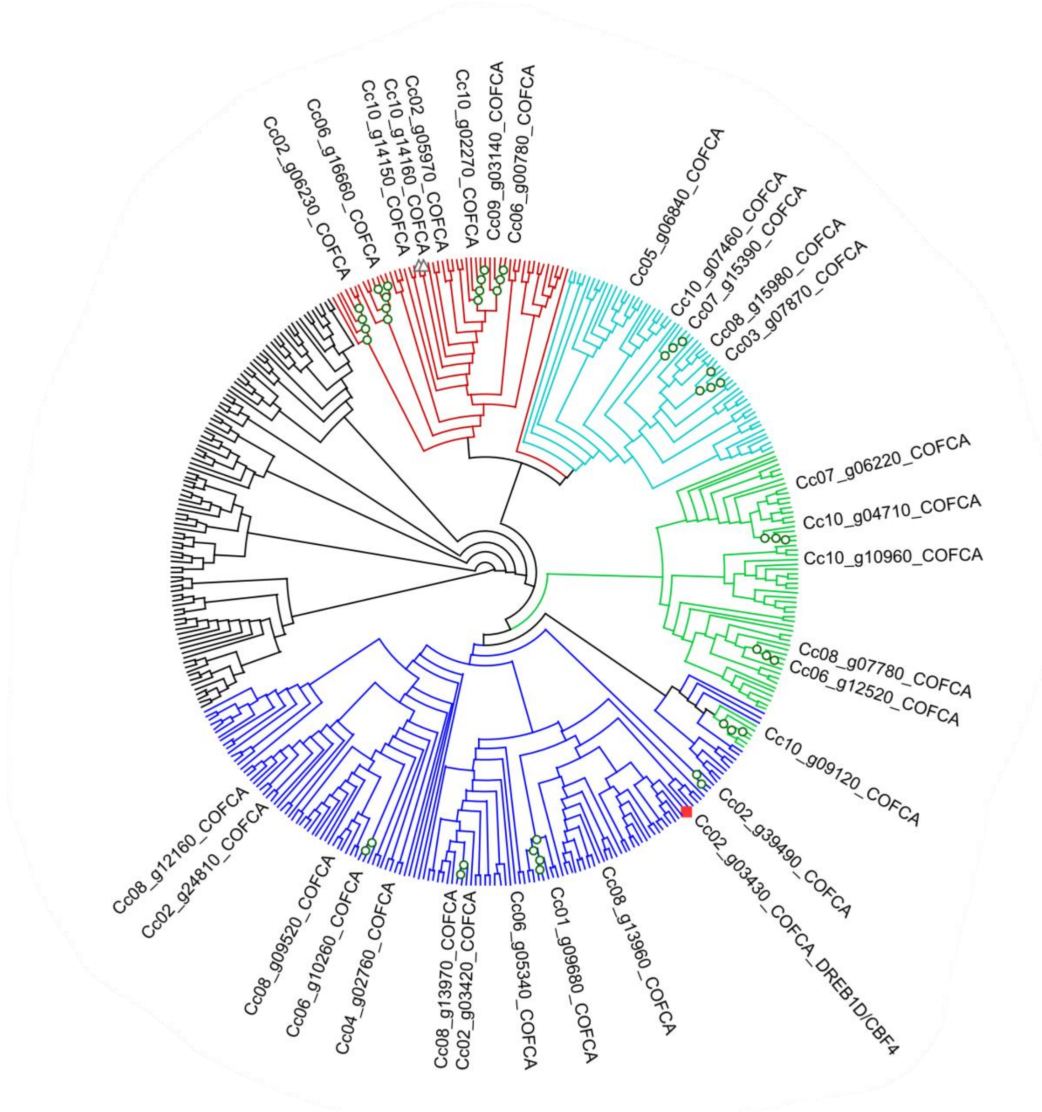


Figure 2 The reconciled taxonomic (species) tree with *DREB* gene tree. *DREB* tree harboring homologs sequences from *A. thaliana, Coffea canephora*, *S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays* was rooted with viridiplantae phylogenetic tree. *DREB* subgroups I, II, III and IV are highlighted in light blue, green, dark blue, and red, respectively. *CcDREB* members are in bold, green circle nodes indicate speciation with high bootstrap confidence, blue triangle marks the *CcDREB* genes undergone duplication, and the red square shows *CcDREB1D*.

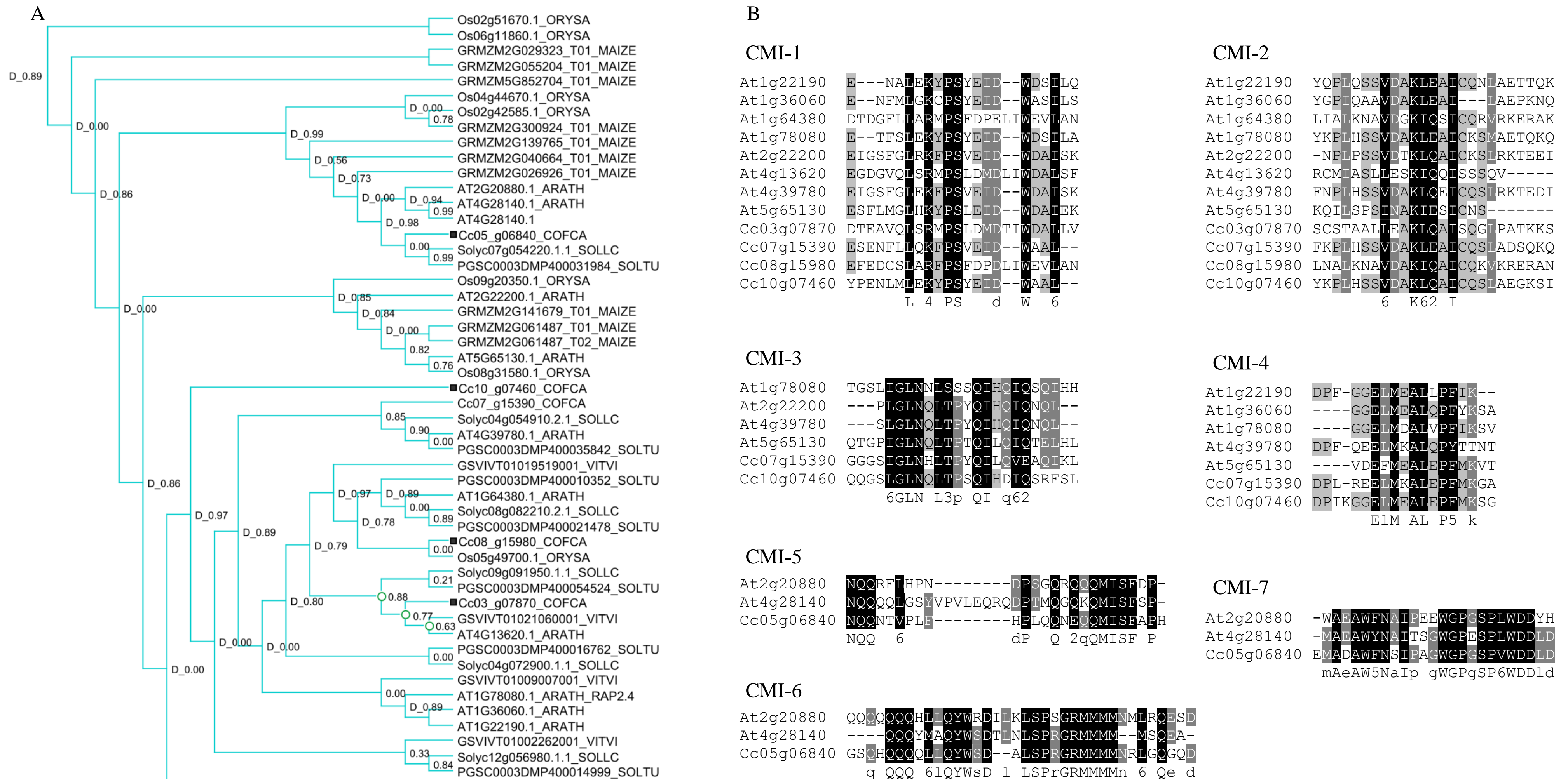


Figure 3 Subnetwork of DREB subfamily subgroup I. (A) Close-up view of *CcDREB* members that compose subgroup I of DREB subfamily and homologs from *A. thaliana*, *S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays.* Coffee sequences are marked by black squares. (B) Alignment of deduced protein sequence from *Arabidopsis* and coffee indicates conserved motifs outside of AP2 domain. An alignment of the N-terminal regions revealed the specific domains of subgroup I proteins CMI-3, CMI-4, CMI-5 and CMI-6.The motif CMII-3 is similar to CMIII-4. The conserved motifs are underlined. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. Consensus amino acid residues are given below the alignment.

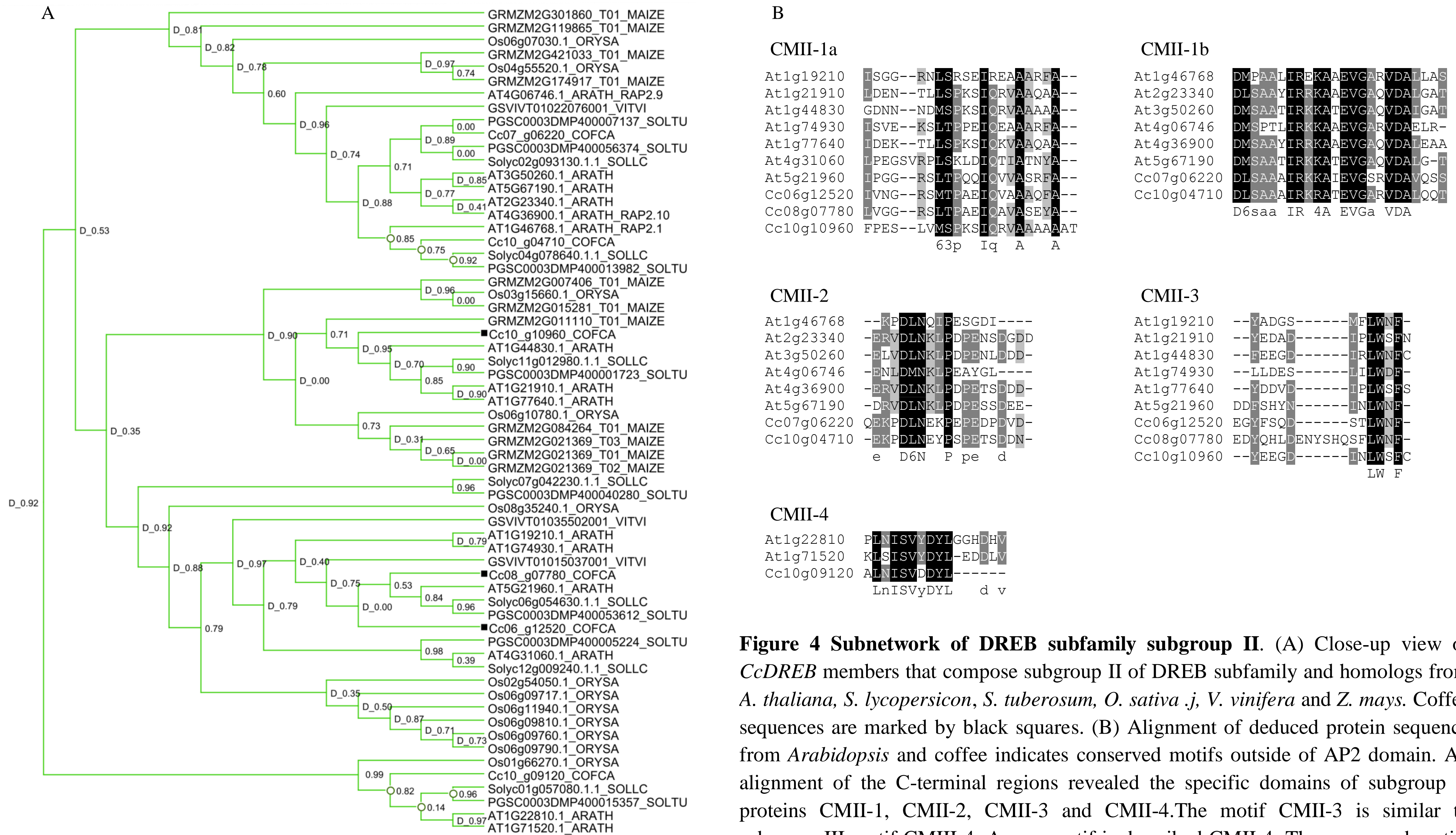


Figure 4 Subnetwork of DREB subfamily subgroup II. (A) Close-up view of *CcDREB* members that compose subgroup II of DREB subfamily and homologs from *A. thaliana, S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays.* Coffee sequences are marked by black squares. (B) Alignment of deduced protein sequence from *Arabidopsis* and coffee indicates conserved motifs outside of AP2 domain. An alignment of the C-terminal regions revealed the specific domains of subgroup II proteins CMII-1, CMII-2, CMII-3 and CMII-4.The motif CMII-3 is similar to subgroup III motif CMIII-4. A new motif is described CMII-4. The conserved motifs are underlined. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. Consensus amino acid residues are given below the alignment.

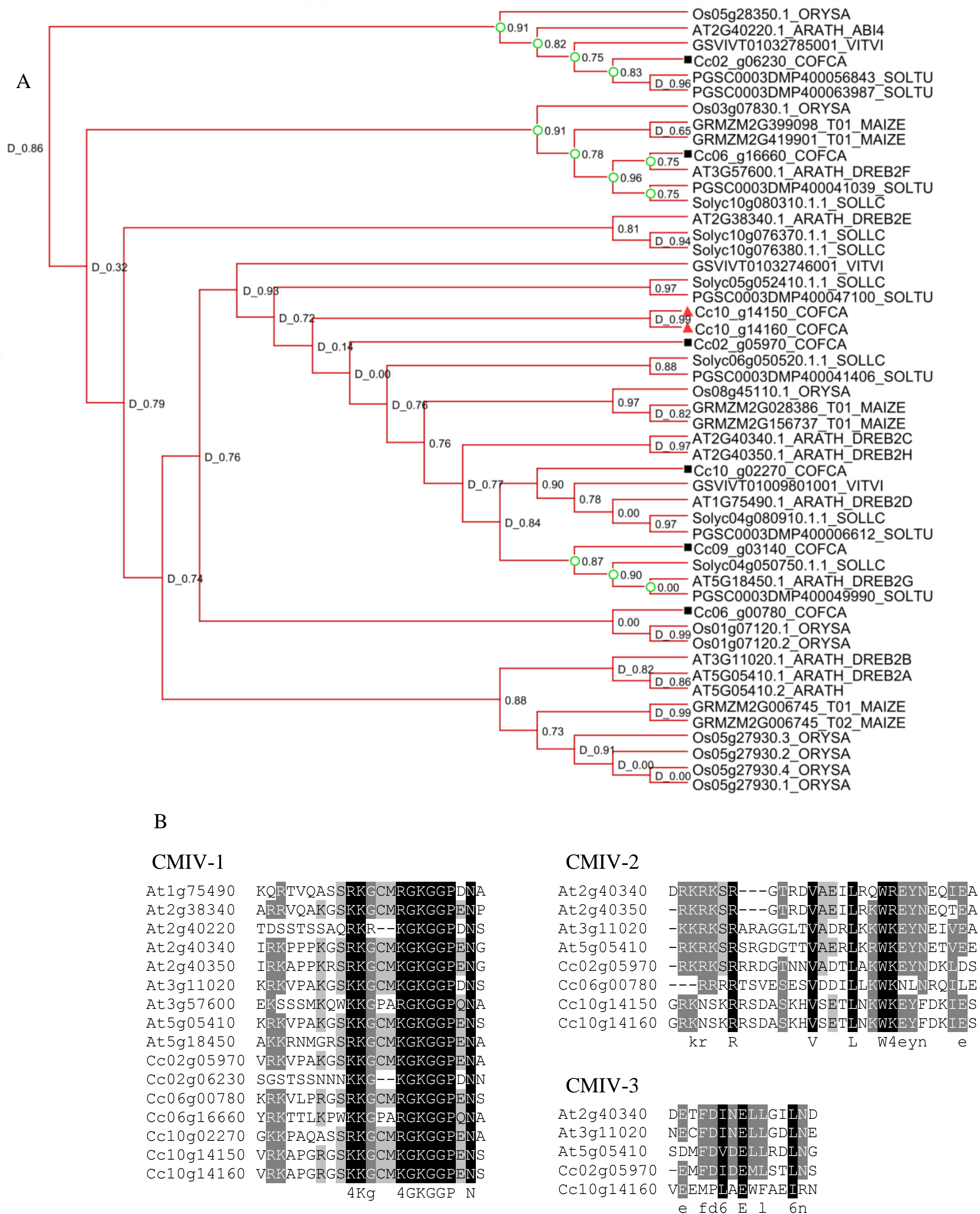
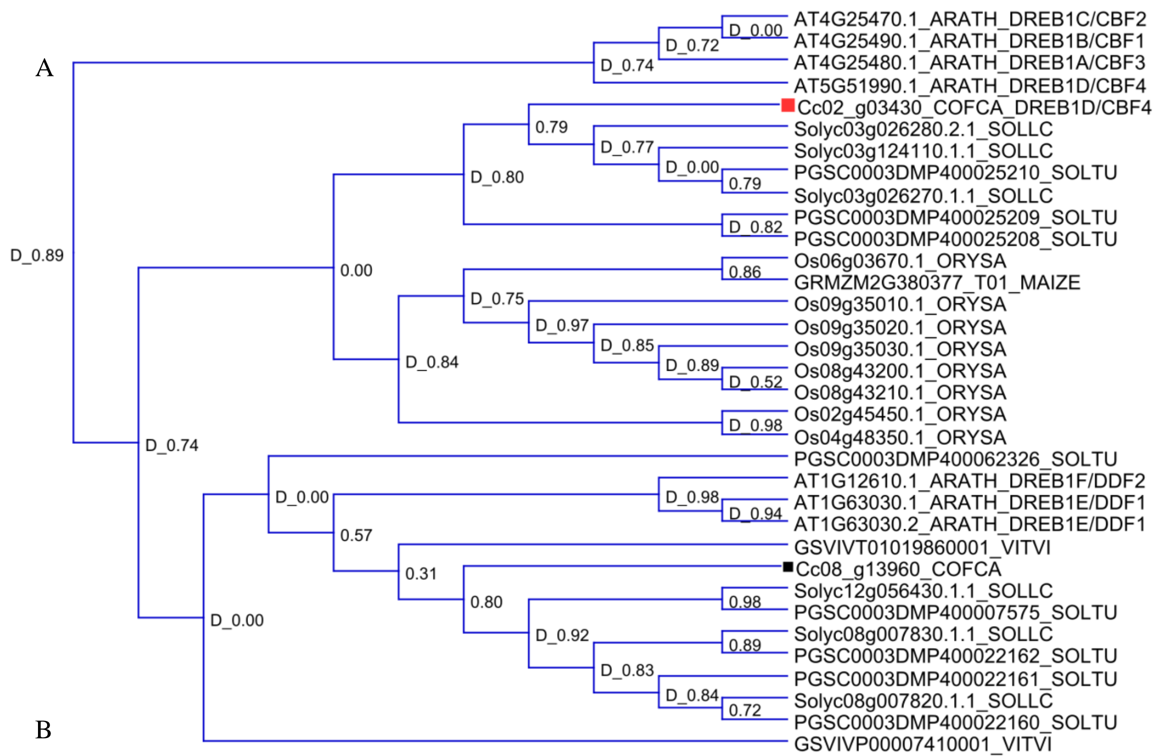


Figure 5 Subnetwork of DREB subfamily subgroup IV. (A) Close-up view of *CcDREB* members that compose subgroup IV of DREB subfamily and homologs from *A. thaliana, S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays.* Coffee sequences are marked by black squares. (B) Alignment of deduced protein sequence from *Arabidopsis* and coffee indicates conserved motifs outside of AP2 domain. An alignment of the N-terminal regions revealed the specific domains of subgroup IV proteins CMIV-1 and CMIV-2. Alignment of the C-terminal regions revealed high homology of Cc02g05970 protein sequence to three *Arabidopsis* orthologs in the domain CMIV-3. The conserved motifs are underlined. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. Consensus amino acid residues are given below the alignment.

Figure 6 Subnetwork of DREB subfamily subgroup III. Overview of *CcDREB* members that compose subgroup III of *DREB* subfamily and homologs from *A. thaliana, S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays.* Coffee sequences are marked by black squares. The candidate gene *CcDREB1D* is marked by a red square.



CMIII-3

```
At1g12610  -NDDIILAEMRPKKRAGRRVFKETRHP[AP2/ERF Domain]DSAWRLPVPESNDPDVIRRVAAEAAEMFRPVD
At1g63030  -NDDITVAEMKPKKRAGRRIFKETRHP[AP2/ERF Domain]DSAWRLPVPASTDPDTIRRTAAEAAEMFRPPE
At4g25470  GDYSPKLATSCPKKPAGRKKFRETRHP[AP2/ERF Domain]DSAWRLRIPESTCAKEIQKAAAEAALNFQDEM
At4g25480  GDYIPTLASSCPKKPAGRKKFRETRHP[AP2/ERF Domain]DSAWRLRIPESTCAKDIQKAAAEAALAFQDEM
At4g25490  GDYCPTLATSCPKKPAGRKKFRETRHP[AP2/ERF Domain]DSAWRLRIPESTCAKDIQKAAAEAALAFQDET
At5g51990  SECSPKLASSCPKKRAGRKKFRETRHP[AP2/ERF Domain]DSAWRLRIPETTCPKEIQKAASEAAMAFQNET
Cc02g03430 SDEEVLLASNNPKKRAGRKKFRETRHP[AP2/ERF Domain]DSAWRLPVPESPEPKHIQKAAAEAAEAFRPSE
Cc08g13960 SDEEIILSSSRPKKRAGRKKFKETRHP[AP2/ERF Domain]DSVWRLPVPESRDAKDIRKAAAEAAEMFRSQE
Cc01g09680 ---------------------------[AP2/ERF Domain]DSAWRVARAQSSLASDIQIAALQAAQAFRPSL
                6a    PKK AGR4kF4ETRHP[AP2/ERF Domain]DSaWR6  p 3      I  aAa2AA   F
```

CMIII-2

```
At1g12630  TEFMDEEEVLNMPNLLANMAEGMMVAPPS
At2g36450  NGFMDEDLVFDMPNVLMNMAEGMLLSPPR
At4g25470  AFYMDEEAMLGMSSLLDNMAEGMLLPSPS
At4g25480  AFYMDEEAMFEMPSLLANMAEGMLLPLPS
At4g25490  AFYMDEETMFGMPTLLDNMAEGMLLPPPS
At5g51990  VFYMDDEALLGMPNFFENMAEGMLLPPPE
At5g52020  LEFMDEEAMLNMPTLLTEMAEGMLMSPPR
Cc01g09680 TMFVDEEAVFNMPALIDNMAEGMLLTPPA
Cc02g03420 DQFVDEEELFDMPNLLVDMAEGMLVSPPR
Cc02g03430 VFFMDEEAVFGMPGLIANMAEGLMLPPPQ
Cc06g05340 HEFIDEDLIFDMPNVLTNMAEGMLLSPPR
Cc08g13960 SMFMDDEALFDVHGLMAHMAEGLLLSPPH
Cc08g13970 DEFIDVEALIDMPNLLVDMAEGMLVSPPR
             56d e 6   6p      MAEG666 pP
```

CMIII-4

```
At1g12610  MSLWSYR
At1g12630  EDLWGY-
At1g63030  LSLWSY-
At2g25820  SFLWNYE
At2g36450  DYLWNFP
At3g16280  VELWSFN
At4g25470  VSLWSY-
At4g25480  VSLWSY-
At4g25490  VSLWSY-
At5g51990  VSLWSFD
At5g52020  DNLWSYK
Cc02g03420 DNLWSYP
Cc02g03430 MFLWSYS
Cc08g09520 SLLWNYY
Cc08g13960 VPLWSYT
Cc08g13970 DSLWSYF
             LW 5
```

CMIII-7

```
At2g35700  SFPYEEPFLS
At2g44940  GFRLEEPFFL
At3g60490  GFRFEEPFNW
At4g16750  SFPYEDPFFL
Cc02g24810 VFRLEEPFLW
            F  EePF
```

Figure 7 Subnetwork of DREB subfamily subgroup III. (A) Close up view of *CcDREB* members that compose subgroup III of DREB subfamily and homologs from *A. thaliana, S. lycopersicon*, *S. tuberosum, O. sativa .j, V. vinifera* and *Z. mays.* Coffee sequences are marked by black squares. The candidate gene *CcDREB1D* is marked by a red square. (B) Alignment of deduced protein sequence from *Arabidopsis* and Coffee indicates conserved motifs outside of AP2 domain. An alignment of the

N-terminal regions revealed the specific domains of subgroup III proteins CMIII-3. Alignment of the C-terminal regions also revealed high homology of protein sequence between *Arabidopsis* and coffee in the additional domains described by Nakano et al (2006), CMIII-2, CMIII-4 (also referred as LWSF) and CMIII-7. The conserved motifs are underlined. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. Consensus amino acid residues are given below the alignment.

4. Discussion

The completion of the *C. canephora* genome sequence allowed us to identify for the first times the *DREB* gene families in this species, and to implement the corresponding coffee genome database. At least 31 canonical *DREB* genes were identified. Compared to other crops, no specific amplification of the *DREB* family could be noticed. Gene annotation could be enhanced by a thorough manual annotation of the genes using the genome browser and annotation software Artemis. Spurious introns in intronless genes were also identified on other automatic annotation programs (ie the banana genome sequence, Droc et al., [2013]), and stress out the necessity to manually check a gene family before interpreting phylogenetic data.

The reconciled tree showed a clear separation of the *DREB* subgroups I, II, III and IV, in reference to previously observed in *Arabidopsis* (Nakano et al., 2006). Not surprisingly, a small subnetwork from subgroup II showed close similarity to subgroup III (Figure 2A). This apparent contradiction has already been subject of discordance in previous classifications in *Arabidopsis* whereas the members of this subnetwork had already been clustered in either subgroup (Sakuma et al., 2002; Nakano et al., 2006).

Cc02g034430 was confirmed as a member of group III, and was present in tandem with another *DREB* gene. It would be interesting to analyze synteny relationship between different species and check if this tandem duplication was conserved in other species, particularly with tomato and vitis genomes as has been already done by other groups (Lefebvre-Pautigny et al., 2010; Guyot et al., 2012).

For some *DREB* genes, a clear orthology relationship with *Arabidopsis* suggests some conservation of function. For example, *Cc06g16660* presents some clear orthologie with ABI4, involved in ABA

sensitivity (Finkelstein et al., 1998). This suggests the conservation of this function in Coffee, but this should be tested experimentally.

Paralogous genes are less likely to have preserved the same function as two orthologs (Dolinski and Botstein, 2007). As an example, the *AtDREB1/CBF* group diverged from a recent duplication and it nearest orthologue in *C. canephora* is the *Cc02g03430*. Therefore, the function of *CcDREB1D*, like many other DREB genes, cannot be inferred directly from the physiological data obtained in *Arabidopsis*. The same pattern was recently observed for *C. canephora* coding for N-methyltransferases involved in the synthesis of caffeine (Denoeud et al., 2014).

A recent review from Gabaldón and Koonin (2013) highlights the fact that genes that are the most similar to each other in compared genomes might not be orthologous. Synteny analysis may therefore offer complementary tools to infer orthologie relationship, besides this phylogenetic analysis. Platforms like COGE (https://genomevolution.org/CoGe/) offer such possibility. At the same time, orthology does not necessary imply a conservation of gene function (Gabaldon and Koonin, 2013). This reinforces the necessity to characterize specifically these genes in Coffee.

5. References

Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science (80-) **345**: 1181–1184

Dolinski K, Botstein D (2007) Orthology and functional conservation in eukaryotes. Annu Rev Genet **41**: 465–507

Droc G, Larivière D, Guignon V, Yahiaoui N, This D, Garsmeur O, Dereeper A, Hamelin C, Argout X, Dufayard JF, et al (2013) The banana genome hub. Database **2013**: bat035

Dufayard JF, Duret L, Penel S, Gouy M, Rechenmann F, Perrière G (2005) Tree pattern matching in phylogenetic trees: Automatic search for orthologs or paralogs in homologous gene sequence databases. Bioinformatics **21**: 2596–2603

Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998) The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA 2 domain protein. Plant Cell **10**: 1043–1054

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**: 1531–1545

Gabaldon T, Koonin E V (2013) Functional and evolutionary implications of gene orthology. Nat Rev Genet **14**: 360–366

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, et al (2005) Galaxy: A platform for interactive large-scale genome analysis. Genome Res **15**: 1451–1455

Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol **537**: 113–137

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol **59**: 307–321

Guyot R, Lefebvre-Pautigny F, Tranchant-Dubreuil C, Rigoreau M, Hamon P, Leroy T, Hamon S, Poncet V, Crouzillat D, de Kochko A (2012) Ancestral synteny shared between distantly-related plant species from the asterid (Coffea canephora and Solanum Sp.) and rosid (Vitis vinifera) clades. BMC Genomics **13**: 103

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R (2007) Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics **8**: 460

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30**: 3059–3066

Koonin E V (2005) Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet **39**: 309–338

Kuzniar A, van Ham RCHJ, Pongor S, Leunissen JAM (2008) The quest for orthologs: finding the corresponding gene across genomes. Trends Genet **24**: 539–551

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol **25**: 1307–1320

Lefebvre-Pautigny F, Wu F, Philippot M, Rigoreau M, Zouine M, Frasse P, Bouzayen M, Broun P, Pétiard V, Tanksley SD, et al (2010) High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. Tree Genet Genomes **6**: 565–577

Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants: Functional domains, evolution and regulation. Eur J Biochem **262**: 247–257

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta **1819**: 86–96

Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol **140**: 411–432

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol **149**: 88–95

Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. Plant Cell **13**: 1959–1968

Rouard M, Guignon V, Aluome C, Laporte M-A, Droc G, Walde C, Zmasek CM, Périn C, Conte MG (2011) GreenPhylDB v2.0: comparative and functional genomics in plants. Nucleic Acids Res **39**: D1095–102

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290**: 998–1009

Yang L, Cho H-G (2012) Comparative evaluation of intron prediction methods and detection of plant genome annotation using intron length distributions. Genomics Inform **10**: 58–64

CHAPTER 3

Characterization of *CcDREB1D* promoter region from *Coffea canephora* and functional analysis to assess the responsiveness of the *CcDREB1D* alleles promoter Hp_15, Hp_16 and Hp_17 under drought stress.

Characterization of CcDREB1D promoter region from Coffea canephora and functional analysis to assess the responsiveness of the CcDREB1D alleles promoter Hp_15, Hp_16 and Hp_17 under drought stress.

1. Background

Studies of gene promoters are essential to understand the regulation of gene expression in plants. Up to now, a limited number of coffee promoters have been studied and functionally characterized by transgenic approaches. The first one was the grain-specific promoter isolated from *csp1* gene encoding an 11S-globulin seed storage protein from *Coffea arabica* that showed strong seed-specific expression in transgenic tabacco plants (Marraccini et al., 1999). In another work, Marraccini et al.(2003) reported the cloning of the rubisco small subunit (*CaRBCS1)* promoter of *C. arabica* that was tested in transgenic tobacco and shown to be leaf-specific. Satyanarayana et al. (2005) reported the cloning of the promoter of N-methyl transferase (NMT) gene involved in caffeine biosynthesis pathway. The promoter region was tested in tobacco, exhibiting expression of the *GUS* reporter gene in leaves. A recent study identified the promoter of the *C. arabica nsLTP* (non-specific lipid-transfer proteins) as a sequence able to target the expression of the *GUS* reporter gene specifically in the an endosperm of transformed *Nicotiana tabacum* plants (Cotta et al., 2014). Even if these results support the idea that the mechanisms implicated in the transcriptional control of these promoters are highly conserved between tobacco and coffee species, the absence of (1) certain anatomical structures in model plants (e.g. *A. thaliana* and *N. tabacum*) and/or (2) transcription factor involved in promoter regulation impeding the study of promoters such as those spatio-temporally regulated and inducible by biotic and abiotic stress, cannot be excluded (Potenza et al., 2004). Therefore, and whenever possible, it is better to perform functional analyses of these promoters in their own (homologous) biological context. Up to now, the only coffee promoter that was functionally tested in transgenic coffee plants was that of the *CaWRKY1a* gene involved in plant response to leaf rust fungus *Hemileia vastatrix* (Petitot et al., 2013). In that case, stable transgenic *C. arabica* lines expressing a *pW1a::GUS* construct were

obtained by *Agrobacterium*-mediated transformation and high GUS activity was observed in leaves subjected to mechanical wounding.

Regarding DREB genes, and even if several studies have been focusing in functional characterization of *DREBs* genes in multiple plant species (reviewed in Agarwal et al., 2006b; Lata and Prasad, 2011; Akhtar et al., 2012), less attention was paid to study their corresponding promoters. When fused to the *hGFP* (green fluorescent protein), the promoter of *DREB2* gene from *Zea mays* (*ZmDBP2*) targeted the expression of reporter gene in the onion epidermic cell (Wang et al., 2011). Also members of *DREB* subfamily, the promoter sequences of *AtRAP2.6*, *AtRAP2.6L*, *AtDREB26* and *AtDREB19* were able to drive the expression of *GUS* reporter gene mainly in flowers but also in immature siliques (Krishnaswamy et al., 2011). Moreover, *AtRAP2.6* and *AtRAP2.6L* were responsive to stress hormones (jasmonic acid, salicylic acid, abscisic acid and ethylene), salt and drought, while *AtDREB19* was highly responsive to salt, heat and drought but less responsive to stress hormones alike *DREB26* (Krishnaswamy et al., 2011). By performing a 5'-deletion analysis of *AtDREB2C* promoter, it was shown that proximal region of this promoter is sufficient to induce *GUS* expression by heat stress in young vascular tissue of transformed *A. thaliana* (Chen et al., 2012). The promoter from *GmDREB3* gene from *Glycine max* was upregulated in calli of wheat after 1 hour of cold stress (Chen et al., 2009). When analyzed in transgenic lines of *A. thaliana* ABA-deficient and ABA-insensitive, the transcription of the *GUS* reporter gene appeared significantly impaired to osmotic stress treatments, for *AtDREB2A:GUS* construct, indicating the contribution of both ABA-dependent and -independent response networks in the transcription control of the *AtDREB2A* promoter transcriptional under osmotic stress conditions (Kim et al., 2011). Point mutation in the 27 kb promoter region of *AtCBF2* gene of *A. thaliana* also revealed the importance of calmodulin binding transcription factor (CAMTA) in cold regulation of *AtCBFs* (*AtCBF1*, *AtCBF2* and *AtCBF3*) gene expression (Doherty et al., 2009). *AtDREB2A* and *AtDREB2B* promoters were shown to be induced by dehydration and high-salt stress in leaves and roots of transgenic plants of *Arabidopsis* (Nakashima et al., 2000).

In heterologous system, the activity of *OsDREB1B* promoter from rice was observed in leaves, stem and roots of transgenic plants of *Arabidopsis* in response to both cold and ABA stresses (Gutha and

Reddy, 2008a). In that case, the expression of the *GUS* reporter gene was highly induced in young tissues, whereas, individual osmotic stress (PEG, NaCl and mannitol), and oxidative stress (methyl viologen) strongly induced this promoter in both stem and roots but not in leaves of *Arabidopsis* GUS activities driven by the *MbDREB1* promoter from dwarf apple (*Malus baccata*) in transgenic *Arabidopsis* increased in response to ABA, cold, temperature, drought, and salt treatments (Yang et al., 2011). Despite the valuable information on promoter regulation obtain from these studies, almost of them were performed with *DREB* promoters from plant models and analyzed in transgenic model system. Therefore, it is of great interest the functional characterization of *DREB* promoters in crop plants to assist genetic improvement of crop plants. In the chapter 3, we present for the first time the results regarding the functional characterization of *DREB* promoter from non-model and perennial specie, *C. canephora* by genetic transformation of *C. arabica*.

This chapter is written in the form of an article draft which regroups (1) the results of kanamycin selection optimization for genetic transformation in coffee mediated by *Agrobacterium tumefaciens*, and (2) the functional characterization of *CcDREB1D* promoters in *C. arabica* transgenic plants. However, it is planned to publish the results of the first part as a technical note and those of functional characterization of *CcDREB1D* as a full-length research paper.

2. Introduction

Climatic variability has always been the main factor responsible for the fluctuation of coffee yields worldwide, and the climate change, as a result of global warming, is expected to present a major challenge for the coffee industry. Despite of the uncertainty surrounds the effect of climate changes on individual producing regions and overall coffee production, significant variations of temperature and rainfall are expected to occur in coffee producing regions. Furthermore, the potential impact will not only vary between countries but also within producing areas in individual countries (van Hilten, 2011). Climatic changes leads to increased extreme temperatures and drought periods which are the major abiotic factors affecting production in most coffee-growing countries (DaMatta and Ramalho, 2006). Natural variability exists for drought tolerance among *Coffea* species, even at the intraspecific level between and within genetic groups (Montagnon and Leroy, 1993; Montagnon et al., 2012). In *C. canephora*, two major groups contrasting for drought response originated from distinct geographical origins, a tolerant SG1 and sensitive SG2, were identified by genetic diversity and population structure analyses (Cubry et al., 2008; Montagnon et al., 2012; Cubry et al., 2013). In this context, several efforts were made to identify clones or cultivars better adapted to drought maintaining high productivity. In Brazil, a genetic improvement program for the development of new cultivars, using SG1 genotypes as source of genetic variability, identified a clonal variety of *Coffea canephora* var. Conilon EMCAPA 8141 highly productive under drought conditions (Ferrão et al., 2000). Considering that coffee retains high leaf relative water content (RWC) under dehydrating conditions, being considered a water-saving rather than a dehydrating tolerant species (Nunes, 1976; Josis et al., 1982; DaMatta et al., 1993), a complementary approach to improve plant performance for drought stress conditions involves the identification and selection of traits that enhance drought tolerance or water-use efficiency (WUE) (DaMatta, 2004). As a matter of fact, the rates of water use or efficiency of extraction of soil water appears to be a major component of differential adaptation to drought among *Coffea canephora* genotypes under water scarcity (DaMatta et al., 2003; Pinheiro et al., 2004). Regarding drought adaptation, these clones are distinguished by phenotypical and morphological

traits, and physiological parameters (Ferrão et al., 2000; Pinheiro et al., 2004; Pinheiro et al., 2005; Praxedes et al., 2006; Marraccini et al., 2012; Vieira et al., 2013). Both clones produce a good crop when grown under irrigation, unlike, under limited soil water survival, productivity and maintenance of tissue water status are considerably defect in 22, therefore classified as drought-sensitive (drought^{S}), compared to the drought-tolerant (drought^{T}) clone 14 (Ferrão et al., 2000). Moreover, drought adaptation in clone 14 was associated to root elongation and water use efficiency through stomatal control (Pinheiro et al., 2005). Evidence of the contrasting response to drought was also revealed at molecular level. Several candidate genes (CGs) for drought tolerance have been identified by different approaches during drought acclimation of *C. canephora* clones EMCAPA 8141 and its expression profiles have been investigated (Marraccini et al., 2012; Vinecky et al., 2012; Vieira et al., 2013). Marraccini et al. (2012) showed that clone 14 modulates more efficiently drought-related pathways and the expression of osmotic responsive genes under drought stress conditions. Patterns of gene expression in leaves from irrigated and non-irrigated plants of clones 14 and 22, but also of the drought^{T} clones 73 and 120, suggest that *CcDREB1D* could have a potential involvement in drought acclimation being highly expressed during water deprivation mostly in drought tolerant clones (Marraccini et al., 2012; Vieira et al., 2013). Alike, analyses of genetic diversity on promoter region of *CcDREB1D* showed that *C. canephora* genetic groups from geographical regions with equivalent mid-season dry periods have greater similarity in the promoter sequences, such as SG1 and G endemic to dry forests, and B, SG2 and UW endemic to rain forests of Africa (Alves et al, submitted [see Chapter 1]). Interestingly, *C. canephora* clones 14 and 22, belonging to the SG1 group (Lambot et al., 2008), share a common *CcDREB1D* haplotype (hp15) and are distinguishable by their second allele, hp16 exclusive to clone 14 and hp17 present only in clone 22 (Figure 4 from Chapter 1). Moreover, the hp15 and hp16 alleles clustered in the SG1 group, more adapted to drought, but hp17 was more akin to SG2 and B groups, both groups less adapted to drought.

DREB genes are key transcription factors implicated in drought, salt and cold adaptation and regulate many essential stress-responsive genes which modulate physiological adaptation of plants to abiotic stress. *DREBs* compose the abscisic acid (ABA) -dependent and -independent pathways of signal

transduction in abiotic stress response, and regulate the expression of several stress-related genes. Indeed, the overexpression of *DREB* genes in several genetic engineered plants lead to up regulation of cold-regulated genes (CORE) and osmotic-stress responsive genes (OR), resulting in increased abiotic stress tolerance (Mahajan and Tuteja, 2005; Tuberosa and Salvi, 2006; Ashraf, 2010; Hardy, 2010; Shanker and Venkateswarlu, 2011). A diversity of *DREBs* genes has been functionally characterized in model plants by homologous transformation. In *Arabidopsis* and rice, as an example, members of subfamilies groups *DREB1* and *DREB2* have shown promising results in inducing cold, salt and drought tolerance (Haake et al., 2002; Sakuma et al., 2002; Dubouzet et al., 2003; Novillo et al., 2004; Tran et al., 2004; Ito et al., 2006; Lim et al., 2007; Takasaki et al., 2010; Kim et al., 2011; Chen et al., 2012). On the other hand, *DREBs* genes have been less extensively characterized in non-model plants in parts due to inefficient protocols for genetic transformation or quite laborious *in vitro* regeneration selection. To cope with this limitation, heterologous transformation in model plants has been used to functional characterization of crop *DREBs* genes and results have been demonstrating a certain degree of functional conservation across plants species (Kasuga et al., 2004; Qin et al., 2004; Gutha and Reddy, 2008; Tong et al., 2009; Bihani et al., 2010). However, a complete knowledge of spatiotemporal, tissue and cell-specific regulation of these genes will likely be achieved mainly by characterization in homologous systems. In attempt to have a better comprehension of spatiotemporal, tissue, cell even as allelic-specific regulation of *CcDREB1D* transcriptional activation during drought stress, herein we studied the transcriptional activity of allelic and homolog forms of *CcDREB1D* promoter from *C. canephora* isolated from the drought^{T} clone 14 and drought^{S} clone 22 by genetic transformation of *C. arabica*.

3. Materials and Methods

Plant materials

Coffee leaves of *C. canephora* Pierre var. Conilon clones 14 and 22 were used for genomic DNA extraction and isolation of homologs *CcDREB1D* promoter. The clones were grown in individual 12

liters pots containing a mixture of soil, sand, and manure (3:1:1, v/v/v) under greenhouse conditions at the Federal University of Vicosa-UFV, Minas Gerais, Brazil. These clones are part of the clonal variety EMCAPA 8141 developed by the Institute for Research and Rural Assistance (Incaper, Vitoria, Espirito Santo, Brazil) during the genetic improvement program for the development of new cultivars for Espirito Santo (Ferrão et al., 2000). *CcDREB1D* promoter region from clones 14 and 22 of *C. canephora* were characterized by homologous transformation in *C. arabica* var. Caturra. Cryopreserved twelve months-old callus cultures from *C. arabica* var. Caturra were used for *Agrobacterium tumefaciens*–mediated genetic transformation experiments (Ribas et al., 2011). Embryogenic calli were induced as previously described (Etienne, 2005). Briefly, primary callus comprising meristematic cells were produced on explants of young leaves (1 cm^2 piece) after one month of cultivation on half strength MS (Murashige and Skoog, 1962) callogenesis (C) medium (Etienne, 2005). Explants containing primary callus were further transferred to half strength MS embryogenic callus production (ECP) medium in which remained for 6-8 months until the regeneration of embryogenic callus. Long-term embryogenic cultures were established cultivating embryogenic callus with monthly subcultures on ECP medium for an additional period of 12 months. Finally, the embryogenic cultures were cryopreserved until its re-establishment during 2 months after heating the cryo-tubes for genetic transformation experiments in view of *CcDREB1D* promoter functional analyse (Figure 1).

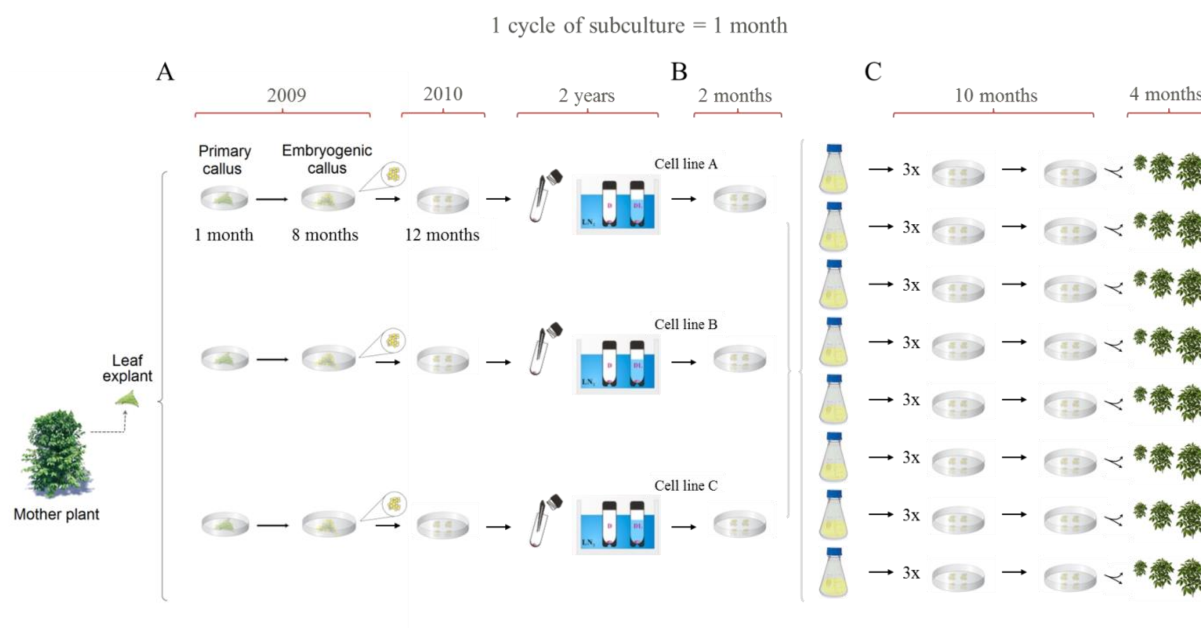


Figure 1 Schematic map of the steps composing the culture *in vitro* and genetic transformation process of *C. arabica* var. Caturra used in this study. (A) Production of dedifferentiated meristematic cells (primary callus) on leaf explants and appearance of the first embryogenic cells. Twelve months cultivation of embryogenic callus to generate long-term embryogenic cultures highly competent for genetic transformaiton and subsequent cryopreservation. (B) Re-establishment of embryogenic cultures from cryopreserved samples and genetic transformation experiments in Erlenmeyer flasks. (C) Selection of transformants on kanamycin enriched media and regeneration of stable transgenic coffee plants.

DNA extraction and isolation of CcDREB1D promoter

Coffee leaves from *C. canephora* var. Pierre clones 14 and 22 were collected and immediately frozen in liquid nitrogen until storage at -80 °C. Genomic DNA was extracted from 20 mg of ground coffee leaves according to Saghai-Maroof (1984). The quality and concentration of DNA samples were verified using a NanoDrop™ 1000 Spectrophotometer (Waltham, MA, USA). *CcDREB1D* promoter regions were amplified by PCR using primers designed on the sequence under the GenBank accession numbers (KM281308- KM281311) with PrimerQuest^{SM}® software (Integrated DNA Technologies - IDT). PCR amplification and primers, cloning and sequencing were performed as described by Alves et al. (submitted [see Chapter 1]). Briefly, PCR reactions were carried out using 100 ng of genomic DNA with DrebS1F/DrebS2R primers and Taq Platinum DNA polymerase (Invitrogen) according to the supplier's instructions. To isolate the alleles, the amplified sequences were then cloned into the pCR®2.1–TOPO® vector (Invitrogen) and transferred into *E. coli* DH5α

competent cells. Considering *C. canephora* ploidy (2n=22), 12 independent colonies were sequenced for each genotype and in both directions with universal primers (M13F and M13R) and internal primers (DrebS2F and DrebS1R) in order to increase the probability to sequence all alleles.

Bioinformatics analyses

Clones sequences were aligned with GenBank reference sequences using SEQMAN 12.0.0 (233). 421 program (DNASTAR Version 2.1.0.97, Copyright 2010-2011, Inc.) to confirm the isolation of sequences with complete homology. The PlantPAN (http://plantpan.mbc.nctu.edu.tw, Chang et al. 2008) and the TSSP / Prediction of Plant Promoters (http://www.softberry.com, Shahmuradov et al. 2005) web interfaces were used for scanning the *cis*-regulatory elements in the *CcDREB1D* promoter sequences. The significance of candidate *cis*-elements was evaluated through a maximum threshold of core similarity (equal to 1.0) and matrix similarity of 0.75.

Plasmid constructs for homologous transformation in Coffea arabica

Recombinant binary vectors were generated by PCR amplification of pCR®2.1–TOPO recombinant vectors harboring the allelic sequence of clone 14 and 22 using different forward primers and a single reverse primer. Restriction sites *Hind*III and *Bgl*II were included in 5'-end of forward and reverse primers, respectively. Two series of 5'-deletions of *CcDREB1D* promoter were amplified for hp15 (–1466/+1 and –762/+1, relative to ATG), hp16 (–1466/+1 and –762/+1) and hp17 (–1466/+1 and –1103/+1) using the common reverse primer RproDREB1D*Bgl*II and three different forward primers 1.4proDREB1D*Hind*III, 1.1proDREB1D*Hind*III and 0.7proDREB1D*Hind*III. All amplified regions covered the upstream sequence from the TSS and the 5'-UTR until the first amino acid.

Table 2 List of selected primers used for the construction of *CcDREB1Dpro* vectors.

Primer	Sequence
0.7proDREB1DHindIII (a)	5'– TAATTCCAAGCTTTGTCTGAAGT –3'
1.1proDREB1DHindIII (b)	5'– AAGAGAACAACAAGCTTCTTGT –3'
1.4proDREB1DHindIII (c)	5'– TCCTAGTAAGCTTCACGTTGT –3'
RproDREB1DBglII (a,b,c)	5'– TGTTGAGAAATGGTTAGATCTTGAA –3''
GUS F (d)	5'– GCACTAGCGGGACTTTGCAA –3'
GUS R (d)	5'– CGCGAAGCGGGTAGATATCA –3'

[a] Primers pair used to amplify the position -762/+1 of the promoter region.
[b]Primers pair used to amplify the position -1103/+1 of the promoter region.
[c]Primers pair used to amplify the position -1466/+1 of the promoter region.
[d]Primers pair used to amplify the *GUS* (formerly *GUS*) reporter gene.

The resulting PCR fragments were subcloned into the pGEM-T Easy vector (Promega, Madison, WI, USA) and then digested with *Hind*III and *Bgl*II restriction enzymes. To generate binary vectors harboring *CcDREB1Dpro*:*GUS* constructs, the constitutive promoter fragment *CaMV35S* (previously excised with *Hind*III/*Bam*HI restriction enzymes) of pBI121 vector (Clontech, Palo Alto, CA, USA, Accession: AF485783) was replaced by the homologous sequence of *CcDREB1D* promoter. These recombinant constructs were named pD14-hp15*pro*1.4, pD14-hp15*pro*0.7, pD14-hp16*pro*1.4, pD14-hp16*pro*1.1, pD22-hp17*pro*1.4, pD22-hp17*pro*0.7, respectively, according to genotype-haplotype proceeding and the nucleotide positions at the 5'-end (Figure 2). The pBI121 vector with *GUS* gene under the regulation of the constitutive promoter *CaMV35S* together with the *GUS*-promoterless construction pBI101 (pBI121 vector without *CaMV35S* promoter) were used as negative and positive control, respectively. Each recombinant vector was then transferred independently via electroporation into the disarmed *Agrobacterium tumefaciens* strain LBA1119. For all cloning steps, plasmidial DNA was extracted with the Wizard® Plus SV Minipreps DNA Purification System and recombinant vectors were verified by sequencing (Genome Express, France) to confirm that no mutations has been introduced into the PCR products. PCR reactions were carried out using 6.25 ng of genomic DNA, Taq Platinum DNA polymerase and primers at final concentration of 0.2 µM according to the supplier's instructions (Invitrogen) under the following conditions: initial denaturation (94 °C-2 min) followed by 30 cycles (94 °C-30 s, 68 °C-30 s, 72 °C-3 min) and a final extension step (72 °C-10 min.). For all reactions, a first step of 10 amplification cycles were carried out with lower annealing

temperature values to ensure primers mismatch pairing with target sequence. The mismatch annealing temperatures were calculated with PrimerQuest^{SM}® software (Integrated DNA Technologies - IDT).

Figure 2 Schematic representation of the *CcDREB1D* promoter region constructions used to transform *C. arabica*. (A) *CcDREB1D* haplotypes found in clone 14 (drought tolerant) and 22 (drought susceptible). Note that the clones 14 and 22 share a common allele (hp15: white square) but also specific alleles (hp16 [black square] for clone 14 and hp17 [black circle] for clone 22). (B) Schematic map of the *CcDREB1D* and *CaMV35S* promoter-*GUS* constructs used in this study. The left panel shows the name of each construct used for stable transformation of *C. arabica* plants and on the right are the relative locations and lengths of the promoter fragments that were fused with the *GUS* (formerly *GUS*) reporter gene. The pBI121 vector, harboring the *CaMV35S:GUS* cassette, and the pBI101 vector, containing the *GUS*-promoterless cassette, were also used as positive and negative controls of GUS enzymatic activities, respectively, during the transformation of *C. arabica*.

Agrobacterium tumefaciens-mediated transformation of C. arabica.

Transgenic lines of *C. arabica* var. Caturra were produced as described by Ribas et al. (2011). *A. tumefaciens* strain LBA1119 harboring the recombinant vectors were grown at 28°C for 24 hours in YMB selective medium (Yeast Extract Mannitol Agar) with rifampicin (25 mg.l^{-1}) and kanamycin (50

mg.l^{-1}). Highly efficient embryogenic callus cultures of 12 month-old (previously tested with transformation efficiencies superior to 90%; Ribas et al. 2011) were successfully restored from three different cryopreserved embryogenic cultures named A, B and C (Figure 3A) and then subcultured again for 2 months. Coffee restored embryogenic cultures were then placed on baby food jars and immersed with 10 ml of *A. tumefaciens* suspension (OD_{600} = 0.6) for 10 min. Ensuing, bacterial suspension was removed and the inoculated calli were co-cultivated at 20 °C for five days in the dark (Figure 3B). For decontamination, the co-cultivated calli were rinsed twice with 20 ml sterile water and further placed on a rotary shaker at 30 rpm for three hours with ECP medium containing 1.2 g.l^{-1} cefotaxime (Figure 3C). The liquid was removed and the calli rinsed one last time with ECP medium for 15 min. Finally, calli were blotted on dry filter paper to remove excess bacterial solution and were subsequently placed in Petri dishes containing ECP medium with 500 mg.l^{-1} cefotaxime (Figures 3D and 3E). For each one of the eight transgenic lines (pD14-hp15*pro*1.4, pD14-hp15*pro*0.7, pD14-hp16*pro*1.4, pD14-hp16*pro*1.1, pD22-hp17*pro*1.4, pD22-hp17*pro*0.7, pBI121 and pBI101) the inoculated embryogenic calli were subdivided in three Petri dishes (see Figure 1). Under the same conditions, a non-transformed control (WT) was generated by replacing bacterial solution for ultra-pure water during co-cultivation period.

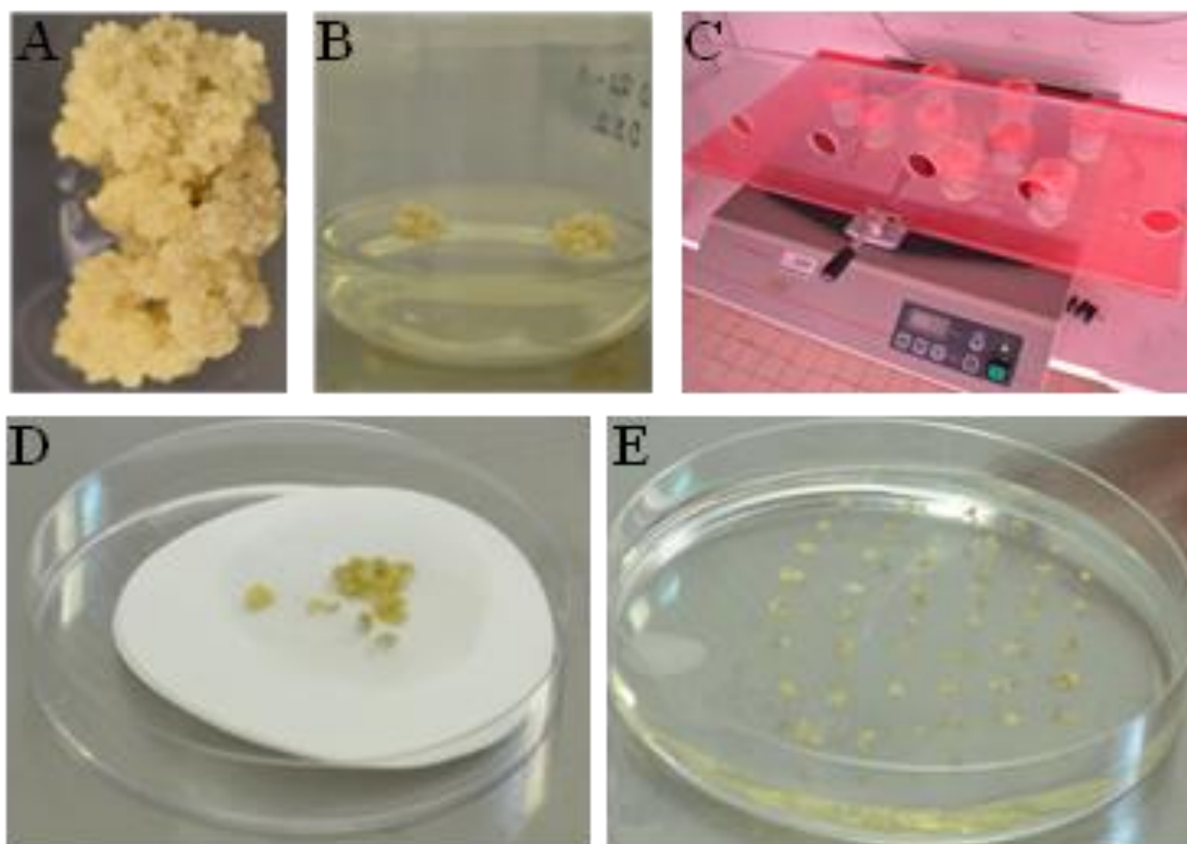


Figure 3 *Agrobacterium tumefaciens*-mediated genetic transformation procedure of *C. arabica*. (A) *C. arabica* long-term embryogenic cultures (12 months-old) were re-established from cryopreserved calli. (B) Embryogenic calli were co-cultivated with *A. tumefaciens* at 20 °C for five days in the dark. (C) Decontamination of co-cultivated calli in ECP medium containing 1.2 $g.l^{-1}$ cefotaxime for three hours. (D) Remove of the excess of bacterial solution on a Whatman paper. (E) Distribution in small cell clusters of inoculated embryogenic calli in Petri dishes containing ECP medium with 500 $mg.l^{-1}$ cefotaxime.

Regeneration of C. arabica *transgenic plants*

After co-cultivation, embryogenic calli were subcultured monthly until the complete regeneration of transgenic plants. During the first month of regeneration, embryogenic calli culture were cultivated on a ‘R’ regeneration medium (Etienne, 2005) containing 17.76 µM of 6-Benzylaminopurine (BAP) and 250 $mg.l^{-1}$cefotaxime. In this first month, two out of the three Petri dishes in each transgenic line were supplied with kanamycin, one with 200 $mg.l^{-1}$ and another with 400 $mg.l^{-1}$. After one month, all embryogenic calli were transferred to fresh media with a decreased concentration of cefotaxime (125 $mg.l^{-1}$) and increased concentration of kanamycin (400 $mg.l^{-1}$). In the following six months, the

cultures were transferred at 4-week intervals to maturation (M) medium (Etienne, 2005) without cefotaxime containing 1.35 µM of BAP and 400 $mg.l^{-1}$ kanamycin. After selection at the concentration of 400 $mg.l^{-1}$ kanamycin, surviving putatively transformed torpedo shaped-embryos were transferred to M medium without kanamycin wherein they were subcultured monthly until cotyledonary embryos developed, the whole process taking approximately 10 months. Ensuing the appearance of the first embryos, clusters of embryogenic cell harboring embryos were considered as independent transformed-events and were individually identified and referenced (Figure 4A). All further globular and cotyledonary embryos arising from the same independent transformation event were considered as "clones" (Figures 4B and 4C). Each kanamycin resistant callus regenerated multiple cotyledonary embryos (Figure 4D). Plantlets with both primordia leaf pair and root were cultivated on MS medium with active charcoal ($1g.l^{-1}$) in baby food jars under sterile conditions for 5 months with monthly subculture. During the entire regeneration process, the cultures were maintained under a 12 h photoperiod (20 $\mu mol.m^{-2}s^{-1}$ light intensity) at 26°C and 80% RH.

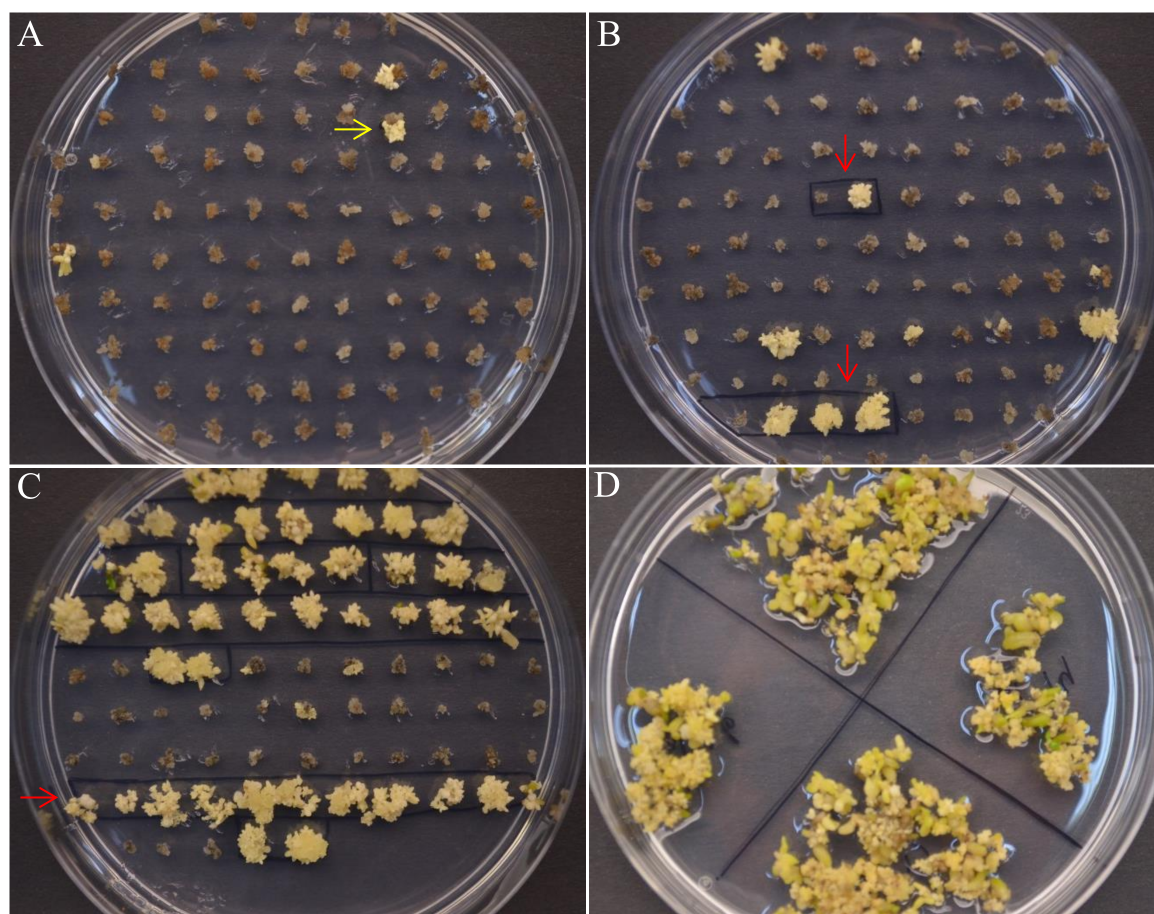


Figure 4 Regeneration of putatively transformed somatic embryos of *C. arabica* var. Caturra. (A) Development of the first somatic embryos over brownish calli after 3 months on maturation medium [M] containing 400 mg.l^{-1} of kanamycin. Putatively transformed embryogenic cultures after 5 months (B) and 6 months (C) cultivation on the same kanamycin selective medium. (D) Putatively transformed somatic embryo germinating on the selective medium. Yellow arrow indicates an independent transformation-event with the multiplication of the first putatively transformed embryogenic cells. Clusters of somatic embryos derived from same independent-event are surrounded by a black square line and indicated by red arrows.

Experiments aimed at improving the efficiency of transformants selection

To generate recombinant vectors including the *GUS* reporter gene in view of *Agrobacterium* genetic transformation, the binary vector pBI121 was used as background model. Therefore, all recombinant vectors constructed in this study carry *nptII* selective gene which confers resistance to kanamycin. Considering that the referenced protocol (Ribas et al., 2011) used hygromycin as selective marker, an adaptation of the protocol was required. Aiming at improving the selection of transformants, the effect of negative selection after co-cultivation (first cycle of subculture) was evaluated by assessing the

subsequent transformation efficiency. In the first subculture, three R media were used with distinct kanamycin concentrations (0, 200 and 400 mg.l^{-1}). After 6 months, clusters harboring somatic embryos and the total number of independent transformation events were counted in view of establishing the transformation rates. Proportion of putatively transformed calli (transformation efficiency) was calculated as the proportion (p) of embryogenic cluster with somatic embryos (p = x/n), where *x* is the number of cluster harboring somatic embryos and *n* the total number of embryogenic cluster subcultured. A $\bar{p} \pm \varepsilon_{\bar{p}}$ confidence limit for binomial distribution was calculated using the formula $p \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p(1-p)}{n}}\right)$ with a level of confidence of 97% in which $Z_{\frac{\alpha}{2}} = 1{,}96$.

PCR analysis of transgenic plants

Nine hundred putatively transformed four months-old plantlets derived from 40 independent transformation events (20 plants randomly taken per transformation event per transgenic line) were verified by qPCR on genomic DNA. Genomic DNA was extract according to Saghai-Maroof (1984) from 200 µg of grounded leaf tissue. For PCR reactions, GoTaq® DNA polymerase (Promega, Madison, WI, USA) and *GUS* gene-specific primers (Table 1) were used to amplify a 120 bp fragment of *β*-glucuronidase gene from 20 ng of genomic DNA using reagent concentrations and PCR conditions specified by manufacture.

Dehydration experiments

Genetically transformed plants carrying the different promoters were subjected to three distinct dehydration treatments performed in *in vitro* conditions to assess the inductility of the promoters by water stress.

To assess the expression of *GUS* reporter gene in tissue and organs, histochemical GUS assays were performed for a same transformation event with both cotyledonary embryos and plantlets, and for these later, several organs like root, stem, cotyledons, apical and root meristems and leaves.

Effect of development stage and tissue specificity on the *DREB1D* promoter expression

Torpedo-shaped embryos, cotyledonary embryos and *in vitro* plantlets (two months) were placed in a laminar flow cabinet under 0.49 $m.s^{-1}$ air flow for 1 hour until reaching a dehydrated state. Five independent transformation events were used per construct. Each transformation event was composed by three biological replicates for torpedo-shaped embryos and cotyledonary embryos, and by five replicates for plantlets.

Effect of water stress on the *DREB1D* promoter expression

Two drought treatments were applied to access the regulation of *CcDREB1D* under low water potential (Ψ_w) and vapor pressure deficit (*D*). To gain insight in the *CcDREB1D* promoter activity in response to drought stress, two time-course experiment was conducted with four month-old plants. For each promoter construct, 60 plants from 8 independent-transformation events were arranged randomly in eight batches of 7 plants. Each experiment was composed of four batches that were subjected independently to 0, 3, 6, 12 and 24 of stress treatment (Table 2).

Effect of low water potential

Water potential was lowered by the addition of 537$g.l^{-1}$ of polyethylene glycol (PEG) (molecular weight 6000; Sigma) to the MS medium. High-molecular-weight PEG 6000 was used to simulate drying soil conditions in which water lost occurs from both cell and cell wall, low-molecular-weight solutes can penetrate cell walls and drain water only from inside cell. Batches of 7 plants were placed over 60 mg vermiculite fully imbibed with 40 ml of PEG diluted in MS solution in sterile plastic Magenta® boxes (Figure 5A). The medium water potential was calculated from the water reactivity (a_w) of the compound vermiculite PEG MS using the equation, $W = \frac{RT \ln(a_w)}{V_w}$ (Pa), with R=8.314, T=298 K, and $V_w=18.07 \times 10^{-3}$.

Effect of vapor pressure deficit (low hygrometry)

Vapor pressure deficit (*D*) was induced by 9% relative humidity (RH) at 27°C under controlled conditions. Considering optimal conditions, a theoretical *D* was calculated applying the formula $D = \left(1 - \left(\frac{RH}{100}\right)\right) * P_{WS}$ (kPa), where relative humidity, RH=9%, and saturated vapor pressure, P_{WS} =3779 Pa at 28°C. To create a low atmosphere humidity of 9%, 400 ml of KOH supersaturate solution was poured in the lower compartment of a temporary immersion bioreactor (Matis®, CID Plastiques, France) which remained hermetically closed during the course of the experiment (Figure 5B). Three-to-four leaf pairs were placed in the upper compartment over 55mm Petri dishes having their upper part exposed to outside environment and radicles immersed in MS medium throw a small hand-made hole on the Petri dish cover (Figure 5D). To avoid any water vapor exchange between MS medium and outside bioreactor atmosphere, the hand-made hole was further closed with high-vacuum silicone grease (Dow Corning®, Sigma) and Petri dishes were sealed with plastic film (Figure 5D). Batches of 7 plants were then placed inside the bioreactor in a resulting atmosphere humidity of 9% (Figure 5C). The control groups pBI101 (negative control for promoter activation) and WT (wild type, negative control of transformation) were composed of 4 plants per batch. Plants were cultivated in a 14h photoperiod. To minimize possible effects of circadian rhythm (subset of biological rhythms with period) all experiments were started in the morning period at 10 a.m. and sampled in the afternoon period.

Table 3 Design of dehydration experiments used to assess *CcDREB1D* promoter activity under PEG and RH water stress treatments.

Promoter construct	N° of plants (Independent-events)	PEG ($\Psi_W \approx$ -2.5 MPa)					9% RH ($D \approx$ 3.43 kPa)			
		Duration (h)								
		0	3	6	12	24	3	6	12	24
pD14-hp15*pro*0.7	60 (8)	4	7	7	7	7	7	7	7	7
D14-hp15*pro*1.4	60 (8)	4	7	7	7	7	7	7	7	7
pD14-hp16*pro*1.1	60 (8)	4	7	7	7	7	7	7	7	7
pD14-hp16*pro*1.4	32 (8)	4	7	7	7	7	nd	nd	nd	nd
pD22-hp17*pro*0.7	32 (8)	4	7	7	7	7	nd	nd	nd	nd
pD22-hp17*pro*1.4	32 (8)	4	7	7	7	7	nd	nd	nd	nd
pBI121	60 (8)	4	7	7	7	7	7	7	7	7
pBI101	40 (8)	4	4	4	4	4	4	4	4	4
WT	40	4	4	4	4	4	4	4	4	4
	Total	36	57	57	57	57	36	36	36	36

nd – not done

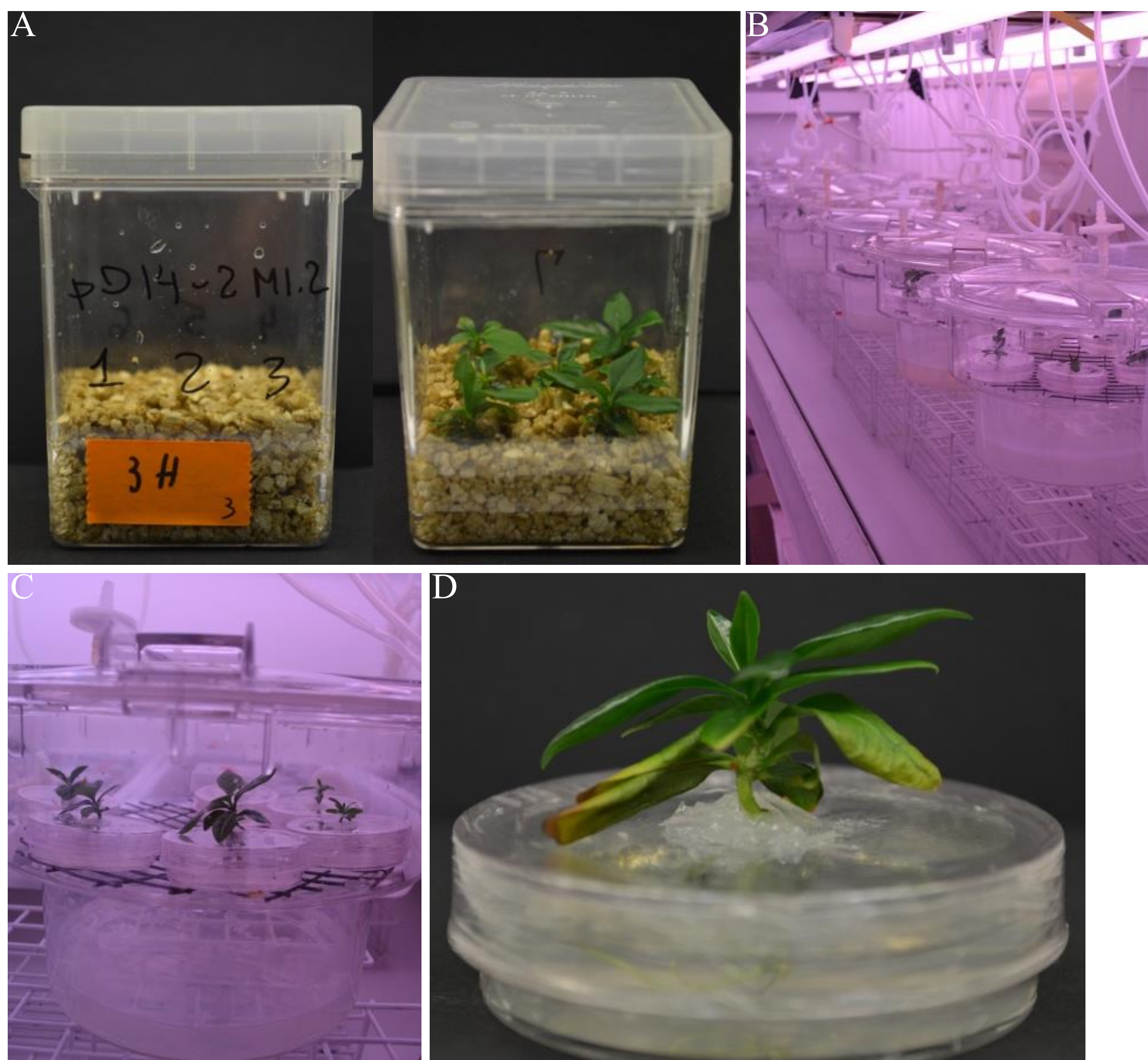


Figure 5 Transgenic plants of *C. arabica* var Caturra subject to low osmotic potential by using PEG and low RH treatments (9%). (A) One batch of 7 *C. arabica* plants cultured in vermiculite was imbibed with 45 ml of PEG 6000 MS medium solution. (B) and (C) Batches of 7 plants were incubated in a Matis® bioreactor under an induced atmosphere humidity of 9% due to the presence of a saturated solution of KOH in the lower bioreactor compartment. (D) Water stressed four month-old transgenic plant placed under 9%RH over 55mm Petri dishes throw a hand-made hole closed with high-vacuum silicone grease and sealed with plastic film. With this system, the water stress is applied only on plant shoots.

GUS staining

Coffee possesses a high phenolic and wax content in leaves, which could affect GUS staining. To improve staining, tissues and organs were dissected to enhance GUS staining solution penetration, and

leaves were cut in 5 mm large bands. Tissues and organs were immersed in GUS staining solution (100 mM sodium phosphate buffer, pH 7.2, 10 mMsodium EDTA, 0.1% Triton X-100, and 1 $mg.ml^{-1}$ 5-bromo-4-chloro-3-indolyl-D-glucuronic acid [Sigma] and 2.5 mM potassium ferrocyanidine). The staining solution was infiltrated into the tissue by subjecting samples to a vacuum for 2x10min. Samples were incubated at 37°C for 24 hours, and then rinsed with water.

Histology and immunodetection of GUS proteins in guard cells

For bright field microscopy observation, GUS stained samples were fixed in fixative (50% methanol and 10% acetic acid) at 4°C for at least 12 h. The tissues were rinsed with water and incubated for 3 days in chloral hydrate:glycerol:water solution (4:1:2, v/v/v) to remove pigments (clearing). Before observation, tissues were rinsed with 80% ethanol and assembled in microscope slides. Entire leaves were examined using a DM600 Leica microscope (Leica Microsystems GmbH, Wetzlar, Germany). Pictures were taken with a Retiga 2000R camera (G-Imaging Co., Wetzlar, Germany). For immunodetection, coffee leaves were pre-fixed with 4% paraformaldehyde in 0.01 M phosphate buffer solution (PBS) (10 mM Na-phosphate, pH 7.5, 138 mM NaCl, and 2.7 mM KCl) for 24 h. The samples were rinsed one time with 0.1 M glycine in PBS solution for 15 min. and three times in PBS (15 min. each). Cross-sections (150 μm) of leaves were obtained using a Micron HM650V vibratome. Fluorescence immunolabelling of β -glucuronidase was attained with 5% bovine serum albumin (BSA) in PBS (blocking buffer, 3 h), anti-β-glucuronidase antibody (1:200 in blocking buffer, overnight at 4 °C), PBS (3× washing, 15 min each), secondary anti-rabbit IgGs antibody conjugated to Alexa Fluor® 488 probe (4 $\mu g.ml^{-1}$ in 2% BSA in PBS, 1 h, in the dark), and PBS (3× washing). Sections were mounted in PBS and observed under a confocal microscope Axiovert 200M 510 META NLO Zeiss, equipped with a laser Chameleon Ultra II (Coherent, Glasgow, UK) (laser 488 nm, BP 500–530 nm).

4. Results

The upstream promoter region of CcDREB1D alleles and homolog genes from C. canephora *clones 14 and 22.*

Previously, we have reported a diversity study of *DREB1D* gene in *C. canephora* and *C. arabica* species indicating the promoter region as the major source of diversity for this gene with possible implications in drought adaptation (Alves et al, submitted [see Chapter 1]). Additionally, haplotype analysis revealed a possible relationship between haplotype variability of allelic *CcDREB1D* promoters of *C. canephora* clones 14 (hp15 and hp16) and 22 (hp15 and hp 17) and natural variability of drought response in *C. canephora*. Therefore, upstream promoters region of *C. canephora* clones 14 and 22 may contain important genetic information about allele-type specific transcription. In particular, three upstream domain between base pairs -38/-270, -825/-1000 and -1265/-1375 which centers the majority of nucleotide diversity present in *CcDREB1D* gene conjointly the majority of nucleotide variability present between allelic and homolog promoters of clone 14, herein conventionally called pD14-hp15 and pD14-hp16, and clone 22, named pD22-hp17 (Figure 6). To further address this finding, two regions 5' of pD14-hp15, pD14-hp16 and pD22-hp17 were subcloned to *GUS* reporter gene in a combinatorial arrangement to allow the comparison of the three domains containing the majority of nucleotide variability. A sequence of 1480 bp from upstream promoter region of *CcDREB1D* encompassing the first amino acidic position +1/-1480 of 5' promoter sequence was amplified from clone 14 and 22, and subsequently sequenced (Alves et al, submitted). Several regulatory elements implicated in abiotic-stress response were identified together with elements driving tissue-specific expression, biotic-stress response and light-inducible expression. Altogether 21 abiotic-stress responsive elements were identified of those, three ABA-responsive element (ABRE: -513/-521, -724/-731, -1441/-1445), one dehydration-responsive element (DRE: -437/-441), one inducer of CBF expression region 2 (ICEr2: -54/-59), nine MYB-related binding sites (MYBRS: -65/-70, -160/-165, -744/-749, -864/-869, -1151/-1156, -1252/-1257, -1397/-1402, -1404/-1409 and -1417/-1422), and seven MYC-related binding sites (MYCRS: -65/-70, -389/-394, -516/-521, -509/-514, -

556/-561, -744/-749, -1011/-1016). Further a search for *cis*-regulatory regions on *CcDREB1D* promoter of pD14-hp15, pD14-16 and pD22-hp17 focusing on the upstream divergent domain regions was carried out to identify potential polymorphisms that could be related to allelic-specific regulation (Figure 7). Any of the found polymorphisms promote directly the loss of *cis*-regulatory elements known to positively regulate drought response such as ABRE and DRE motifs (Figure 7). However, several polymorphisms do occur near *cis*-regulatory element such as MYC and MYB, which can act, respectively, as activator and repressors of transcription during drought response. Particularly, in one of the three upstream polymorphic domain, -825/-1000, was identified a cluster of MYB- and MYC-like regulatory elements (Figure 7). In this region, a structural arrangement of three unidentified motif (GAAWTT: -896/-913) in tandem was observed only in the pD14-hp16 allele due to nucleotide substitutions within the sequence. We identified one regulatory element usually found in biotic-stress responsive genes, W-BOX (-401/-405, -676/-680). Additionally, nine DOF (DNA-binding one zinc finger) binding sites (A/TAAAC) that are required in guard-cells specific expression were also identified, seven on sense (5'-3': -121/-125, -406/-410, -657/-661, -699/-703, -754/-758, -1095/-1099, -1138/-1142, -1170/-1174, -1187/-1191 and -1353/-1357) and two on anti-sense (3'-5': -76/-80, -1100/-1104) strand. Interestingly, we report a cluster of five DOF-binding sites within a distance of 106 bp occurring after the *cis*-regulatory repressors domain, -825 to -1000, and also the presence of a SNP which promote the loss of one DOF *cis*-element (-1187/-1191) (Figure 7). The *CcDREB1D* promoter also contained regulatory elements essential for light-inducible expression as GT-1 binding sites (-1348/1353, -1298/-1303, -1306/-1311, -1173/-1178, -1090/-1095, -976/-982, -890/-895 and -658/-663), all mainly located in the most upstream 500 bp. Interestingly, a particular arrangement of light-inducible element and DOF-binding site was observed between -1089/-1358, two modules comprising one light-inducible element and one DOF-binding site in tandem and one comprising one light-inducible element and two DOF-binding site in tandem. Finally, the promoter sequence also contained three calmodulin binding transcription activator (CAMTA) target site, known as CG-1 element (CG-1: -317/-322, -1424/-1429 and -1432/-1437) and a repressor of *DREB1/CBF* activation conserved motif 6 (CM6: -791/-800). Four sites for the assembly of transcription machinery were

found within the first 760 base pairs of the upstream promoter region, three initiator element (INR: -606/-613, -693/-700 and -735/-742), found in TATA-less of plant promoters (Nakamura et al., 2002), and a TATA-BOX (-330/-333) corresponding to the RNA polymerase II DNA binding site.

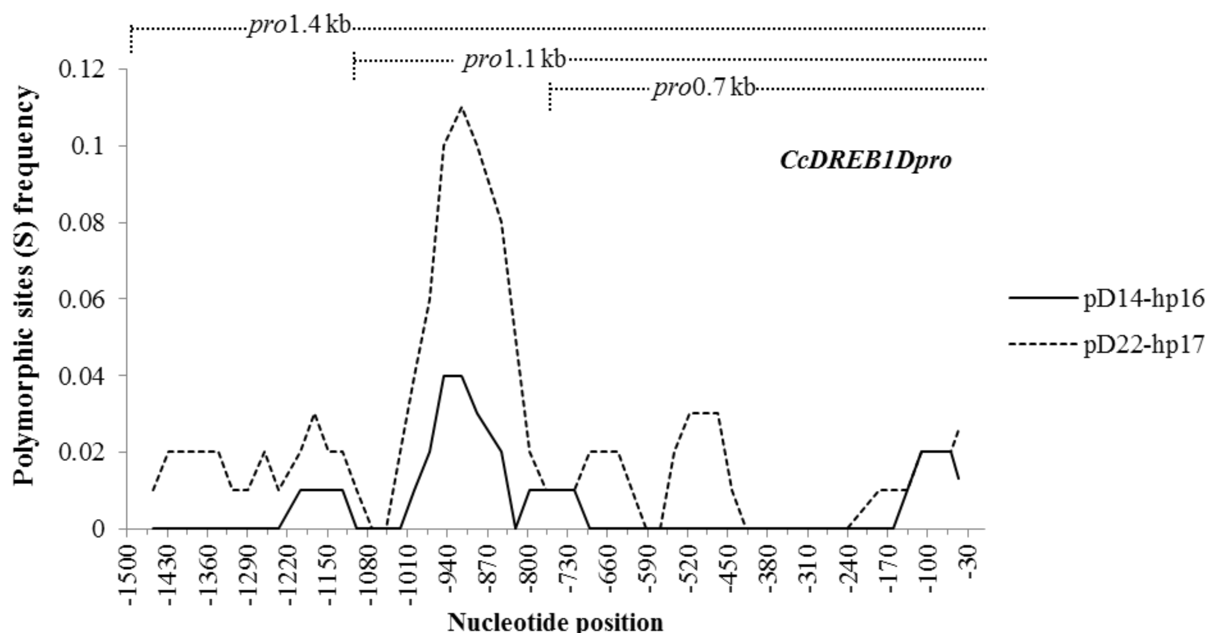


Figure 6 Polymorphic site frequency of the *C. canephora* clone 14 and 22 homolog *CcDREB1D* promoters. The nucleotide variability detected in the upstream promoter region (-1 to -1456) are presented based on a sliding window graphic. The upper bars indicate the extension and the polymorphic upstream domains assessed by full-length or truncated versions of *CcDREB1Dpro*:*GUS* constructions. Three upstream domains between base pairs -38 to -270, -825 to -1000 and -1265 to 1375 contained the majority of nucleotide diversity present between allelic and homolog promoters of clones 14 and 22. The frequency of polymorphic sites (S) is given with observed frequency of polymorphic sites among pD14-hp15, pD14-hp16 and pD22-hp22, using the common allele pD14-hp15 as reference.

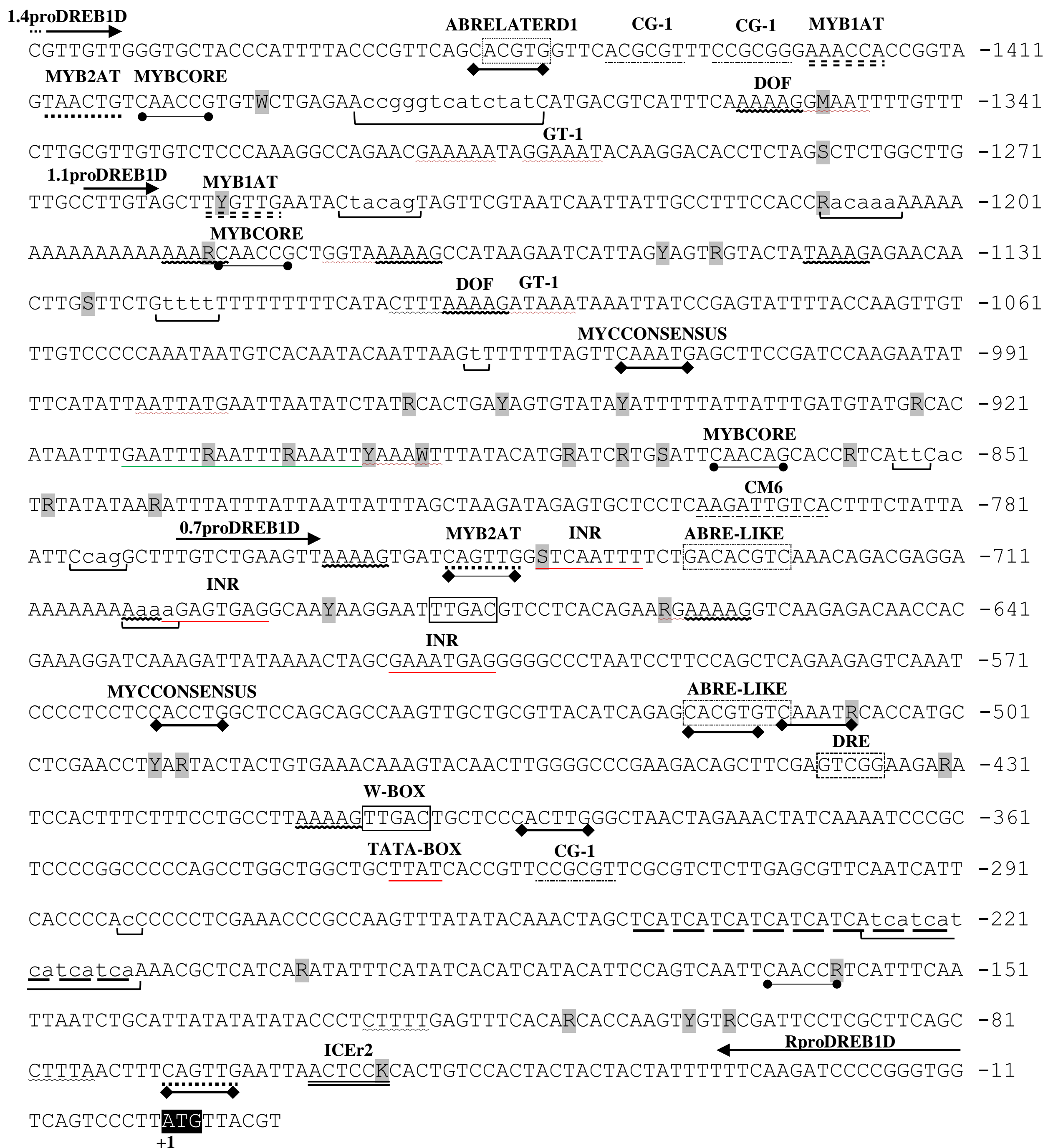


Figure 7 Promoter consensus sequence of *CcDREB1D* and the observed *cis* regulatory elements acting in osmotic- and cold-stress responses (for details see Table 3). The promoter sequence represents a consensus sequence of all genotypes and the IUPAC code indicates the polymorphic sites (gray boxes). Black shading indicates the translation start site and the first base of this start codon corresponds to +1. Insertion sequences are indicated in lower case, deletions are represented by (␣) and (▁) indicates microsatellite core repeats. The drought-response related motifs ABRE-LIKE motif (dot-dashed box), ABRELATERD1 (dotted box), DRE (dashed box), MYBCORE (filled circled), MYB1AT motifs (double-dashed line), MYB2AT motif (dotted line), MYCCONSENSUS motifs (filled diamond), and a cold-response related ICEr2 motif (double underlined) are indicated. The W-BOX (solid line box) is usually found in biotic-stress responsive genes. The DOF-binding site required for guard-cell specific expression are identified in 5'-3' (bold wavy underlined) and 3'-5' direction (wavy underlined). The CG-1 element (dot-dot-dash underline) is a bind site of a calmodulin binding transcription activator protein (CAMTA). The GT-1 (red wavy underline) is a light

inducible element normally target GT-1 proteins. The CM6 motif was identified as repressor of CBF activation (dot-dash underline) in *Arabidopsis*. *Cis*-acting elements for assembly of the transcriptional complex (red underlined) were found in *CcDREB1D* upstream promoter region, TATA-BOX and INR. A structural formation of three unidentified motif (GAAWTT) in tandem is underlined in green, SNP modify the structure that is present only in the pD14-hp16 allele. The nucleotides are numbered (on the right) on each lane. Horizontal arrows indicate primers (Table 1) used to amplify the full-length and truncated versions of coffee *CcDREB1D* promoter sequences. The +1 position, indicating the first translated nucleotide of the *GUS* reporter gene, also corresponds to the start [0] of base numbering (indicated in the right part of the figure).

Table 4 *Cis*-regulatory elements involved in osmotic- and cold-stress responsive gene expression identified using PlantPAN analysis and experimentally supported results from previously published reports.

	cis element	Sequence	TFs that bind to *cis* elements	Gene	Stress condition	Reference
ABRE	ABRELATERD1	ACGTG	bZIP	*erd1*	Water deficit, ABA	PLACE
	ABRE-like	BACGTGKM	bZIP	*rd29*	Water deficit, ABA	AGRIS
DRE	DRE	CCGNC	ERF/AP2	*rd29A*	Water deficit, cold	1
	LTRECOREATCOR15	CCGAC	ERF/AP2	*cor15A*	Water deficit, cold	PLACE
MYB	MYBCORE	CNGTTR	bHLH	*rd22*	Water deficit, ABA	PLACE
	MYB1AT	WAACCA	bHLH (MYB15)	*rd22*	Water deficit, ABA	PLACE
	MYB2AT	YAACTG	bHLH (MYB2)	*rd22*	Water deficit, ABA	PLACE
MYC	MYCCONSENSUSAT	CANNTG	bHLH	*rd22*	Water deficit	PLACE
ICE	ICEr2	ACTCCG	bHLH	*CBF2/ DREB1C*	Cold	2
CG-1	CG-1	VCGCGB	bHLH	*PAL1*	CaM/Ca^{+2} -inducible	3
GT-1	GT-1	GRWAAW	bHLH	*PR1-A*	Light-inducible	PLACE
DOF	DOF	WAAAC	ZF	*KST1*	Water deficit, ABA	4
CM6	CM6	AAGATTGTCA	unkown	*unknown*	Water deficit, ABA	5
W-BOX	W-BOX	TTGAC	ZF	*NPR1*	Wounding	PLACE

1- (Sakuma et al., 2002)
2- (Yamaguchi-Shinozaki and Shinozaki, 2005)
3- (da Costa and Silva, 1994)
4- (Plesch et al., 2001)
5- (Doherty et al., 2009)

Transgenic coffee plant regeneration and molecular analysis

Using promoter-reporter constructs, the *CcDREB1D* promoter activity from allelic forms of *C. canephora* clone 14 and 22 was characterized by homologous transformation of *C. arabica* plants. In the present study were generated *CcDREB1Dpro* recombinants vectors harboring the *nptII* gene as selective marker which confers resistance to kanamycin. Kanamycin selection of coffee transgenic embryos has always been controversial (Etienne et al., 2008). Successful selection of transformed embryos was reported at 400 and 100 $mg.l^{-1}$ of kanamycin (Canche-Moo et al., 2006; Albuquerque et al., 2009), however regeneration of non-transformed embryos was also observed at 400 $mg.l^{-1}$ (Barton et al., 1991; Spiral et al., 1993). Therefore, to limit the risk of escape and increase the reliability and efficiency of transgenic plant selection, the effect of time-course application and concentration of kanamycin was assessed for their efficiency in selecting somatic transformed embryos. In this aim, embryogenic calli submitted to transformation conditions from the transgenic lines pD22-hp17*pro*1.4 and pBI101 were transferred to R medium with 0, 200 or 400 $mg.l^{-1}$ of kanamycin ensuing co-cultivation. Following the first month of subculture, all coming selection was performed with 400 $mg.l^{-1}$ kanamycin. When kanamycin was applied immediately after co-cultivation, 23 independent transformation events, characterized by embryo development on necrotic calli, were regenerated from 292 co-cultivated calli (7.8 % transformation efficiency [TE]) of pD22-hp17*pro*1.4 on 200 $mg.l^{-1}$ selective medium. On the same medium, 64 putative transformed embryos were observed starting from 986 embryogenic calli of the pBI101 transgenic line (6.5% TE). Under the same conditions but on 400 $mg.l^{-1}$ selective medium, 11 and 46 putative transformed embryos were obtained from 110 (10% TE) and 158 (29% TE) co-cultivated calli of pD22-hp17*pro*1.4 and pBI101 lines, respectively. By applying kanamycin only in the second month of subculture, we counted 9 putative regenerated embryos when initiating with 786 embryogenic calli (1.1% TE) for pD22-hp17*pro*1.4, and 51 embryos from 1100 co-cultivated calli of pBI101.

In total, 1278 and 268 calli were initially subcultured in 200 and 400 $mg.l^{-1}$ of kanamycin, respectively. After six month of selection, 87 putatively transformation events were obtained for the

onset concentration of 200 $mg.l^{-1}$ (6.8% TE) and 57 (21.3% TE) for the initial concentration of 400 $mg.l^{-1}$. Under different concentration of kanamycin, regeneration of putatively transformed embryos proved to be more efficient in 400 $mg.l^{-1}$ achieving a proportion (ρ) of 0.21 putatively transformed callus per callus subjected to transformations conditions compared to a ρ of 0.07 when 200 $mg.l^{-1}$ were initially used for selection. If no antibiotic was used in the first month of selection, an even lower ρ was observed (p = 0.03), with 60 putatively transformed calli obtained from 1886 cultivated calli subjected to transformation conditions (Figure 8). The initial addition of 400 $mg.l^{-1}$ of kanamycin was proved to be the most efficient treatment (Figure 8E) with transformation efficiencies in a range of 10-30%. However these values are much lower than those reported by Ribas et al. (2011) that used the same transformation procedure but with a hygromycin selection (70-95% transformation efficiency). However, an accumulative effect of kanamycin during the selection process is described. Although commonly observed during coffee plants regeneration through somatic embryogenesis, the secondary or repetitive somatic embryogenesis process was unexpectedly long and spectacular in presence of kanamycin with oxidation of germinated embryos and regeneration of numerous globular secondary embryos at the root pole which lasted up to 3 months the regeneration of transgenic plants (Figure 8).

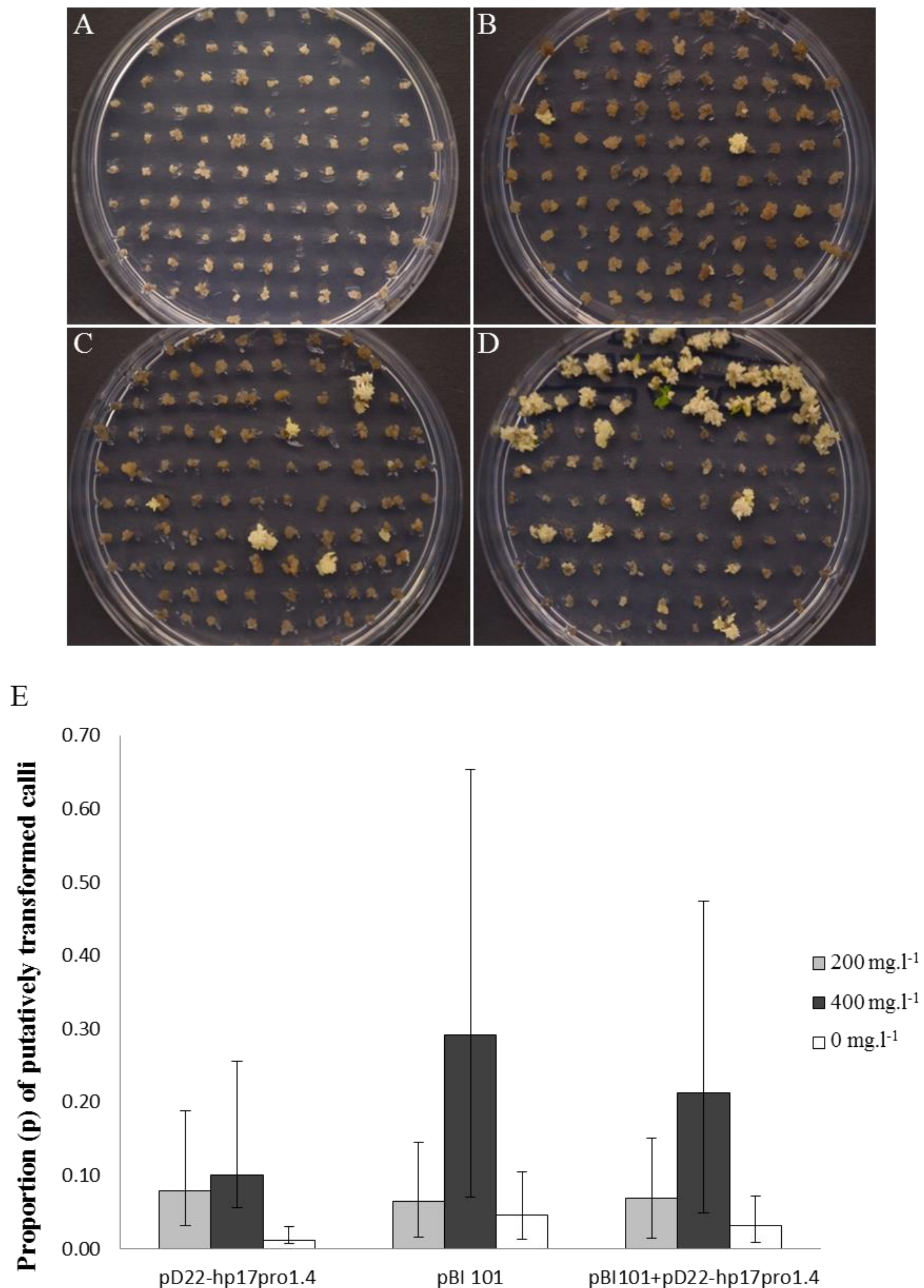


Figure 8 Effect of the onset concentration of kanamycin on transformation efficiency (proportion of putatively transformed calli/ number of calli submit to transformation conditions). (A) Necrotic non-transformed embryogenic calli (negative control) subcultured on R medium supplied with kanamycin. Embryo regeneration on putatively transformed calli when onset

kanamycin concentration of 200 $mg.l^{-1}$ (C) or 400 $mg.l^{-1}$ (D) was applied immediately after co-cultivation, or when applied only after the second month of subculture (B). The onset kanamycin concentrations were applied in R medium in the first cycle of subculture immediately after co-cultivation, and for the following subcultures selection was performed with 400 $mg.l^{-1}$ kanamycin. The pictures were taken after five month selection on 400 $mg.l^{-1}$ kanamycin. (E) Transformation efficiency was estimated 6 months after the end of co-cultivation. The transformation efficiency was estimated by the proportion (ρ) of embryogenic clusters with somatic embryos ($p = \frac{x}{n}$), where x is the number of cluster harboring somatic embryos and n the total number of embryogenic cluster calli submitted to co-cultivation conditions with $A.\ tumefaciens$. A $\bar{p} \pm \varepsilon_{\bar{p}}$ confidence limit for binomial distribution was calculated using the formula $p \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{p(1-p)}{n}}\right)$ with a level of confidence of 97% in which $Z_{\frac{\alpha}{2}} = 1{,}96$.

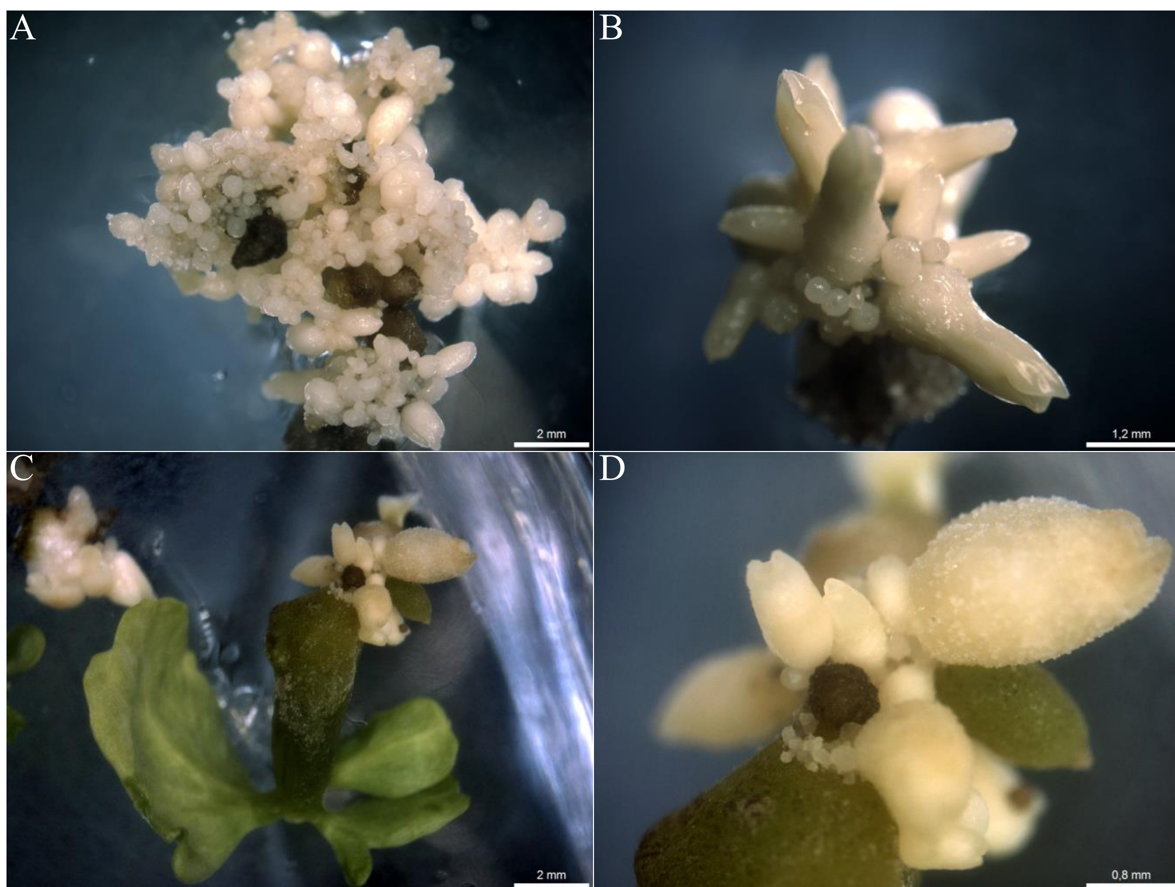


Figure 9 Effect of accumulative concentration of kanamycin on secondary embryogenesis during the transgenic plant regeneration process. (A) Globular and torpedo-shaped embryos of *C. arabica* var. Caturra developing on putatively transformed calli. (B) Putatively transformed cotyledonary-shaped embryos. (C) and (D) Beginning of the secondary embryogenesis at the root pole of genetically transformed germinated embryos.

Several thousand putative transgenic plants were regenerated. Finally, 100 plants from 5 independent transformation events for each transgenic promoter-reporter construct and 100 wild type plants (in total 900 plants), were chosen for molecular analyses an further bioassays. PCR analysis was performed to detect the *GUS* reporter gene in the DNA of the 800 putative transformed plants and 100 non-transformed plants. The amplification of 120 bp insert with GUS specific primers confirmed the transgenic nature of regenerated plants (Figure 10). Consistently, with the previously reported results the average of kanamycin selection efficiency was 88% (Table 4). Stable transgenic plants were further subjected to drought treatment experiments.

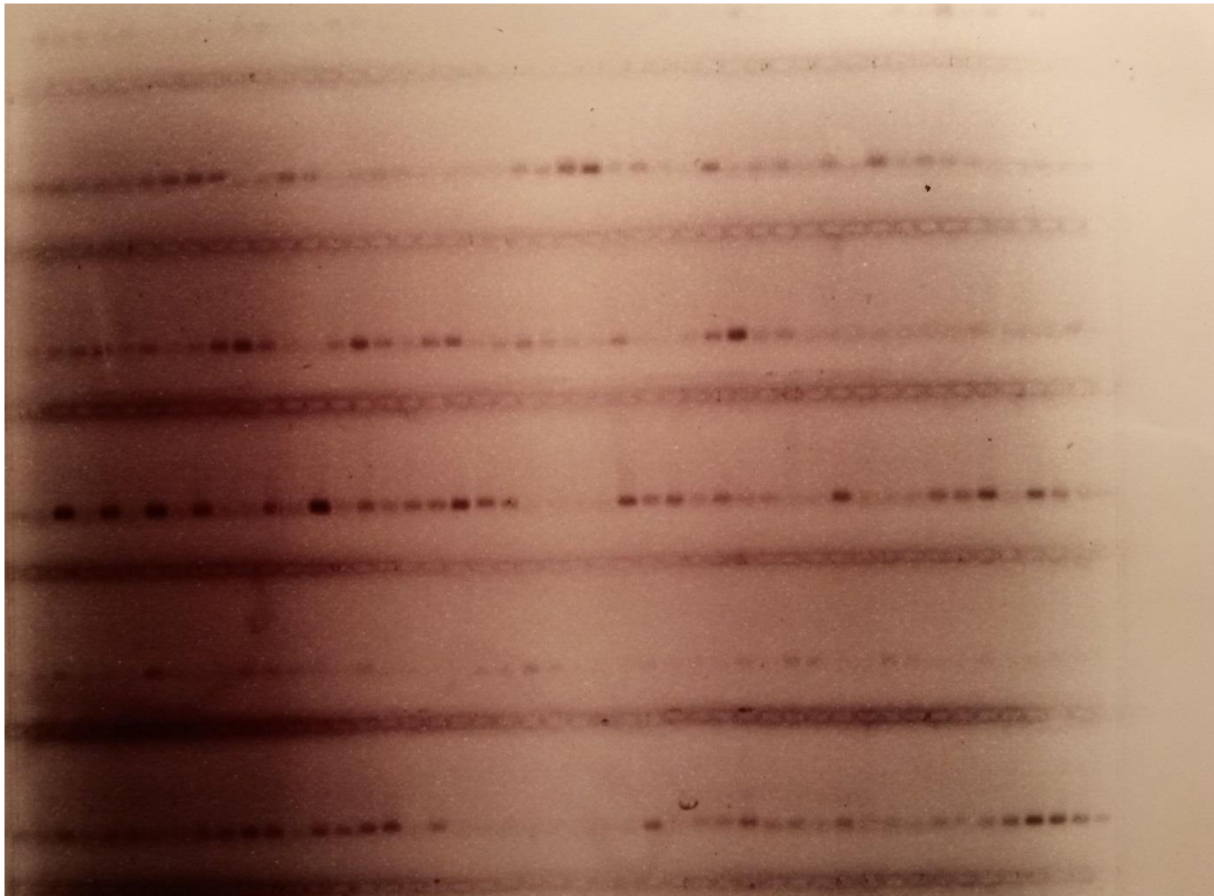


Figure 10 Amplification fragments of the *GUS* reporter gene from regenerated plants of *C. arabica*.

Table 5 Selection efficiency of transformed *C. arabica* var. Caturra plants assessed for each promoter-reporter construct.

Recombinant vector	N° of Plants Regenerated (Events)	N° of Transformed Plants confirmed by PCR	N° of Positive Independent-Events	Selection Efficiency (%)
pD14-hp15pro0.7	100 (5)	100	5	100.00
pD14-hp15pro1.4	100 (5)	81	4	81.00
pD14-hp16pro1.1	120 (6)	67	4	55.83
pD14-hp16pro1.4	120 (6)	119	6	99.10
pD22-hp17pro0.7	100 (5)	93	5	93.00
pD22-hp17pro1.4	100 (5)	95	5	95.00
pBI101	100 (6)	84	5	84.00
pBI121	100 (5)	97	5	97.00
WT	100 (na)	0	0	0.00

na – not applicable.
* The first column indicates the number of putatively transformed plants on kanamycin selective medium and in brackets the number of independent-events regenerated for each promoter-reporter construct. The final number of transformed plants was confirmed by PCR analyses. The number of positive independent-events was obtained by subtracting the total number of independent-events tested to the number of non-transformed event for each transgenic line.

Temporal and spatial regulation of CcDREB1D promoter in transgenic plants of C. arabica *var. Caturra exposed to drought stress*

The activity of the different *CcDREB1D* promoters was investigated in stable transgenic lines of coffee (*C. arabica* var. Caturra), expressing the reporter gene *GUS* under the control of the full-length (1.466 kb), the medium (1.113 kb) or the proximal (752 bp) regulatory regions of *CcDREB1D*. To evaluate the genetic effect of *CcDREB1D* alleles in the regulation of gene expression, we selected the alleles of *C. canephora* drought-tolerant clone 14 and drought-sensitive clone 22. Both clones are heterozygous at the *CcDREB1D* loci. Moreover, they share a common allele (hp15) but are distinguished by the second allele, hp16 present exclusively in clone 14 and hp17 present only in clone 22. Thus, the three alleles were selected for reporter gene analysis by homologous transformation in *C. arabica* with the truncated version thereof. Six transgenic lines (pD14-hp15*pro*1.4, pD14-hp15*pro*0.7, pD14-hp16*pro*1.4, pD14-hp16*pro*1.1, pD22-hp17*pro*1.4, pD22-hp17*pro*0.7), along with one positive (pBI121) and one negative control (pBI101) for GUS enzymatic activity, and a non-transformed control (WT) were generated.

Non stressed embryos and plants: First, to characterize the GUS histochemical pattern of tissues transformed by the *CcDREB1Dpro:GUS* constructions, non-stressed transformed torpedo-shaped and cotyledonary embryos from all transgenic lines and controls were stained in GUS solution. Very intense GUS activity was observed in non-stressed torpedo-shaped and cotyledonary embryos of pBI121 (positive control GUS activity) while no GUS activity was seen in pBI101 and WT lines (negative controls of GUS activity). Non-stressed torpedo-shaped embryos of the transgenic lines pD14-hp15*pro*1.4, pD14-hp15*pro*0.7, pD14-hp16*pro*1.4, pD14-hp16*pro*1.1, pD22-hp17*pro*1.4, pD22-hp17*pro*0.7 showed a weak GUS enzymatic activity. Alike, histochemical detection of GUS activity in plantlets revealed weak and disperse staining in radicle, hypocotyl, epicotyl and cotyledons of *CcDREB1Dpro* constructions when no stress was applied. Taking together, these observations suggest a basal activity of *CcDREB1Dpro* cassettes at the embryogenic state.

Under water stress: Staining of torpedo-shaped and cotyledonary embryos subjected to drought stress in a laminar flow cabinet revealed heterogeneously GUS activities. GUS stainings in torpedo-shaped embryos showed most frequently blue coloration at the root pole (Figure 11). The pD14-hp15*pro*0.7 and pD22-hp17*pro*0.7 torpedo-shaped embryos (Figures 11B and 11D) harboring the proximal promoter allelic versions showed more intense GUS signals compared with corresponding longer versions pD14-hp15*pro*1.4 and pD22-hp17*pro*1.4 (Figures 11F and 11H). Unlike, GUS activity increased from proximal to full-length promoter in the pD14-hp16 (Figures 11C and 11G). The histochemical analysis also revealed larger and more intense areas showing GUS activity for pD14-hp16*pro*1.4 suggesting a strongest activity of this promoter among all full-length promoter at the embryo stage (Figures 11F, 11G and 11H).

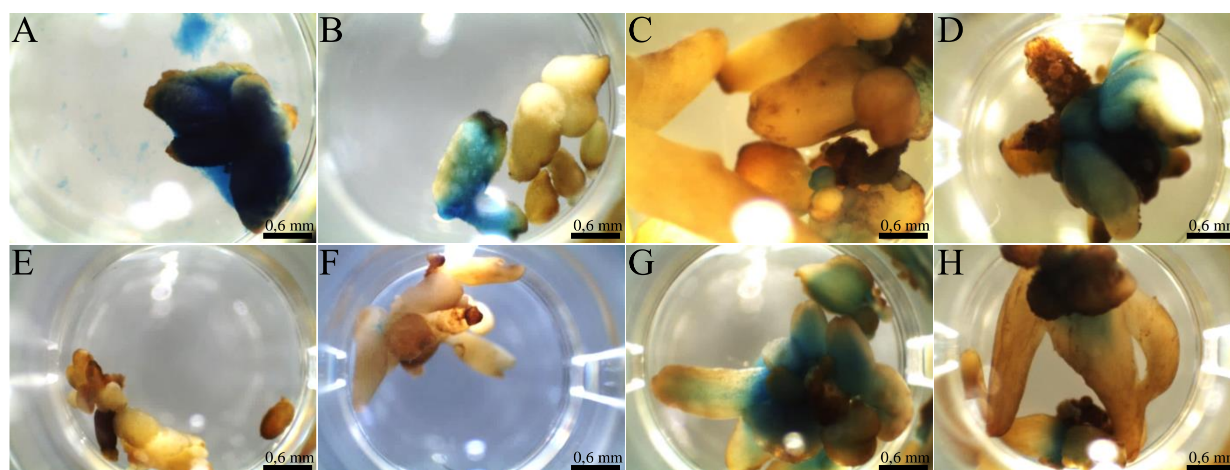


Figure 11 Histochemical localization of GUS activity in *C. arabica* stably transformed embryos carrying *CcDREB1Dpro:GUS* cassettes. (A and E) Embryos regenerated from the positive control pBI121 (*CaMV35Spro*:*GUS*) and the negative control pB101 (*GUS*-promoterless) transgenic lines, respectively. (B, C and D) Embryos regenerated from pD14-hp15*pro*0.7, pD14-hp16*pro*1.1 and pD22-hp17*pro*0.7 transgenic lines carrying cassettes truncated *CcDREB1D* promoters cassettes. (F, G and H) Embryos regenerated from pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 transgenic lines carrying full-length *CcDREB1D* promoters cassettes.

As for GUS analyses on drought-stressed plantlets, very intense tissue-nonspecific GUS activities were again observed for the positive control pBI121 while no GUS enzymatic signals were detected for the negative controls WT and pBI101. In stressed plantlets, intense and diffuse GUS activities were also observed throughout plant organs i.e. cotyledons, hypocotyls and radicle without any evidence of tissue specificity for pD14-hp15*pro*1.4, pD14-hp15*pro*0.7, pD14-hp16*pro*1.4, pD14-hp16*pro*1.1, pD22-hp17*pro*1.4, pD22-hp17*pro*0.7 transgenic lines.

Further, to gain more insight into the cell and tissue specificity of *GUS* gene expression during plant development, GUS activity was evaluated in water stressed small plants from five independent-events per transgenic line and wild type. GUS histochemical activities was assessed in roots, stems, meristems with leaf primordia, cotyledons and leaves of transgenic plantlets after the same drought stress (Figures 12, 13 and 14). Similarly to that found in embryos, no GUS activity was seen in all tissues tested from both WT and pBI101 plants (Figures 12C and D). However, constitutive and very intense GUS activities tissue-nonspecific were observed in roots, stem, meristem with leaf primordia,

cotyledons and leaves of all plantlets harboring *CaMV35S:GUS* construction (Figures 12E and F). Under *CcDREB1D* promoter regulation, GUS activity varied in intensity, tissue- and cell-specificity accordingly to promoter constructs (Figures 13 and 14). Tissue and cell-specificity of GUS activity was observed by the staining of very specific structures.

The pD14-hp15*pro*0.7 line showed intense GUS signal in vascular tissues of roots (Figure 13 A1 and B1). In stems, a less intense activity of pD14-hp15*pro*0.7 was detected in vascular tissues as well as in endodermic and epidermic cells (Figures 13 B2). Longitudinal-sections of pD14-hp15*pro*0.7 meristem evidenced less intense GUS activity in leaf primordia, upper auxiliary buds and ground meristem but a very intense activity in apical meristem (Figures 13 B3). Staining of pD14-hp15*pro*0.7 cotyledons and leaves also revealed a very intense GUS signal, comparable to *CaMV35S*:*GUS* control lines, in epidermis and spongy and palisade parenchymas cells even as in guard-cells (Figures 13 B4 and B5), however, no GUS activity was detected in vascular tissue of leaves. Under pD22-hp17*pro*0.7 regulation, GUS activity showed the same spatial, tissue and cell-specific stringency as in pD14-hp15*pro*0.7 line with the exception of roots wherein no GUS activity was seen (Figures 13 and F).

Substantial differences in GUS activity started to appear when comparing the 1.1 bp promoter, pD14-hp16*pro*1.1, with 0.7 bp promoters, pD14-hp15*pro*0.7 and pD22-hp17*pro*0.7. Histochemical staining in pD14-hp16*pro*1.1 line revealed GUS signal exclusively in leaves (Figures 13C and D). Intense GUS tissue-specific activities were detected in guard-cells but also in leaves epidermis and in spongy and palisade parenchyma (Figure 13 D5). Again, no GUS activity was detected in vascular tissue of leaves. When observing GUS signal in full-length promoter lines pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 an even fine-tune regulation of GUS activity was evidenced (Figures 14A to F). Intense GUS activities were observed strictly in leaves and cotyledons being preferentially localized in guard-cell and stomata. Occasionally, diffuse and weak staining was found in parenchyma and epidermic cells and once more no GUS staining was observed in vascular tissues. Interestingly, variability in the number and spatial localization of GUS stained regions was observed in leaves and cotyledons from the same plant for *CcDREB1Dpro* lines. We could not find any obvious relation between leaf developmental stages and the number or spatial localization of stained regions. In the

same way, GUS activity in guard cells was intense but with staining of punctual cells groups that were heterogeneously distributed along leaf abaxial region.

Figure 12 Histochemical localization of GUS activity in *C. arabica* var. Caturra wild type and transgenic plants carrying *CcDREB1Dpro:GUS* cassettes. Patterns of GUS activity observed in organs and tissues of small coffee plants subjected to drought stress under an air flow of 0.49 m. s^{-1} for 1 hour. Binocular loupe images present GUS staining in entire organs (A, C and E) and bright field microscopy images of dissected organs in longitudinal- or cross-sections (B, D and F). WT, non-transformed control (A and B). pBI101 line (*GUS* promoterless:), negative control of transcriptional activity (C and D). pBI121 line (*CaMV35S*:*GUS*), positive control of *GUS* gene expression (E and F). A same plant was dissected in the following organs (from the left to the right on the same line): (1) roots; (2) stems; (3) apical meristem; (4) cotyledons and (5) leaves. Five independent-events (each one composed by three biological repetitions) were assessed for GUS activity in each of the promoter-reporter cassette. Images show the most representative pattern of GUS activity observed in each promoter-reporter cassette. For microscopy images black bars represent 30 µm.

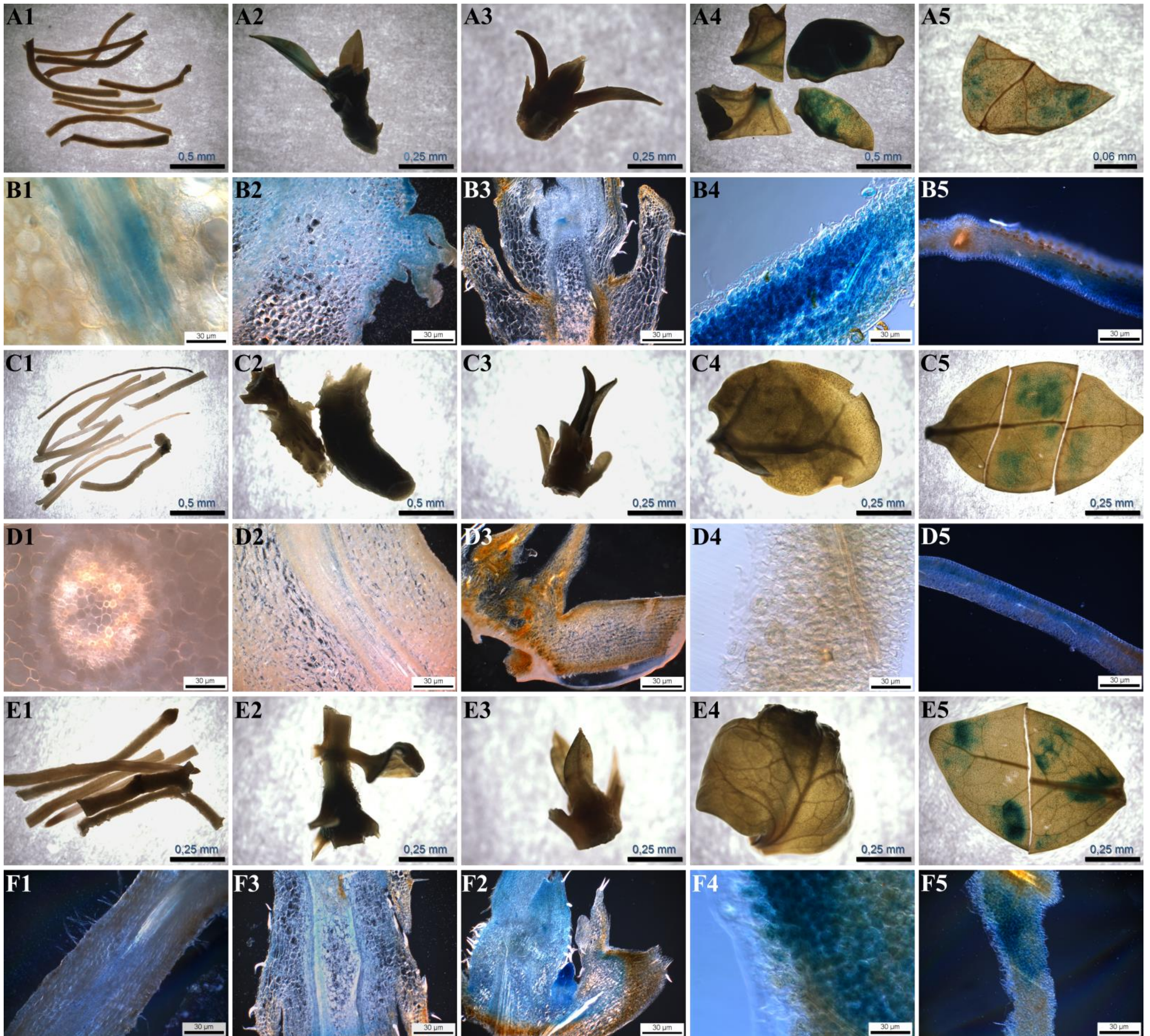


Figure 13 Histochemical localization of GUS activity in *C. arabica* var. Caturra transgenic plants harboring *CcDREB1Dpro*:*GUS* cassettes. Patterns of GUS activity detected in organs and tissues of small coffee plants subjected to drought stress under an air flow of 0.49 m.s^{-1} for 1 hour. Binocular loupe images present GUS staining in entire organs (A, C and E) and bright field microscopy images of dissected organs in longitudinal- or cross-sections (B, D and F). pD14-hp15*pro*0.7 line (A and B). pD14-hp16*pro*1.1 line (C and D). pD22-hp17*pro*0.7 line (E and F). A same plant was dissected in the following organs (from the left to the right): (1) roots; (2) stems; (3) apical meristems; (4) cotyledons and (5) leaves. Five independent-events (each one composed by three biological repetitions) were assessed for GUS activity in each of the promoter-reporter cassette. Images show the most representative pattern of GUS signal observed in each promoter-reporter construct. For microscopy images black bars represent 30 µm.

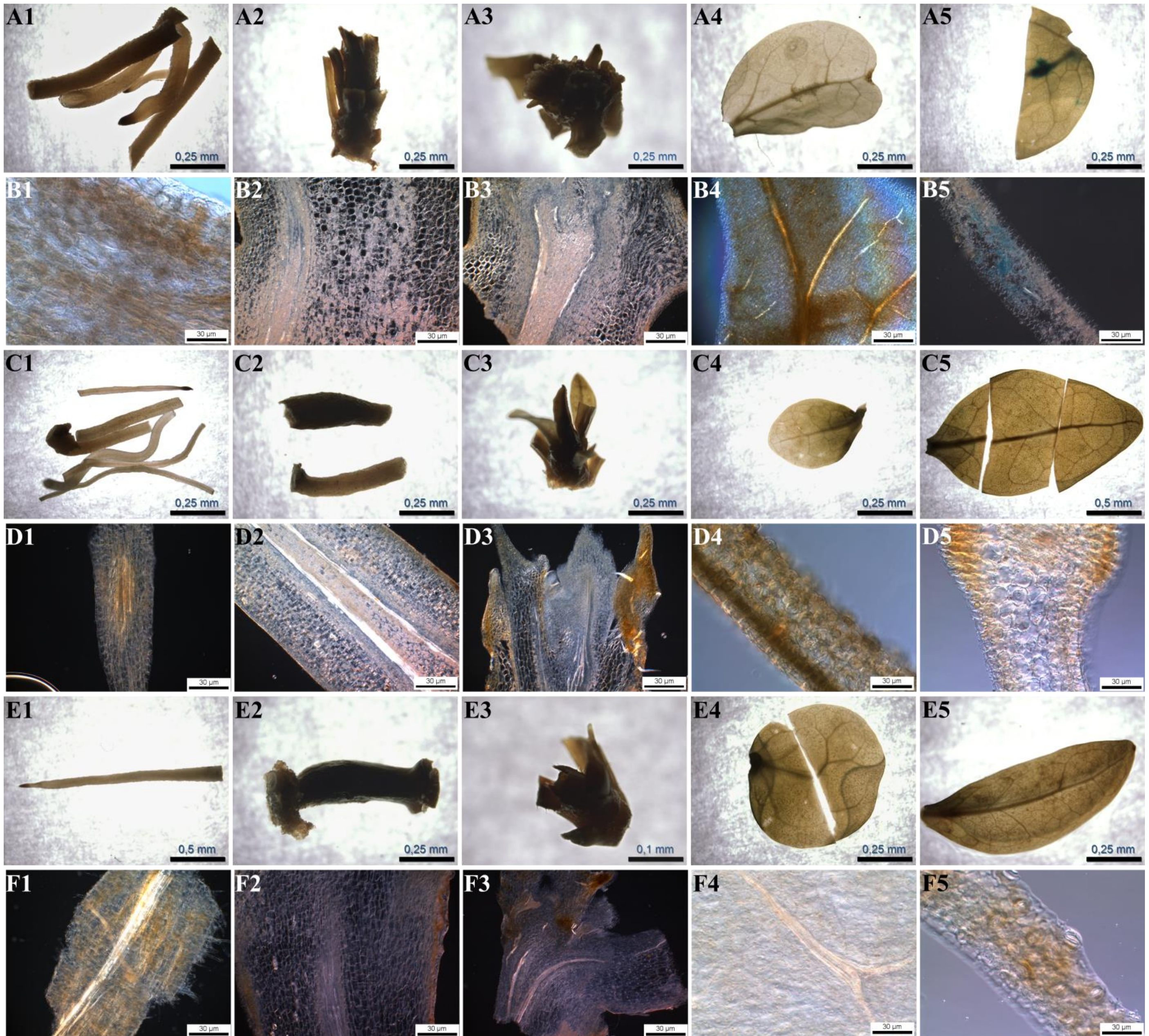


Figure 14 Histochemical localization of GUS activity in *C. arabica* var. Caturra transgenic plants harboring *CcDREB1Dpro*:*GUS* cassettes. Patterns of GUS activity detected in organs and tissues of small coffee plants subjected to drought stress under an air flow of 0.49 m.s^{-1} for 1 hour. Binocular loupe images present GUS staining in entire organs (A, C and E) and bright field microscopy images of dissected organs in longitudinal- or cross-sections (B, D and F).pD14-hp15*pro*1.4 line (A and B). pD14-hp16*pro*1.4 line (C and D). pD22-hp17*pro*1.4 line (E and F). A same plant was dissected in the following organs: (1) roots; (2) stems; (3) apical meristem; (4) cotyledons; and (5) leaves. Five independent-events (each one composed by three biological repetitions) were assessed for GUS activity in each of the promoter cassette. Images show the most representative pattern of GUS signal observed in each promoter-reporter cassette. For microscopy images black bars represent 30 µm.

GUS activity driven by CcDREB1D coffee promoters are detected in guard cells, epidermis and parenchymas of coffee leaves

All examined *CcDREB1Dpro*:*GUS* constructions revealed activities in guard-cells, but intensity and staining patterns varied among them. GUS activity from promoters pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD14-hp16*pro*1.1 were to a various intensity degree and mainly present in ventral side of guard cells (Figures 15E, G and H), while staining of pD14-hp15*pro*0.7, pD22-hp17*pro*0.7 and pD22-hp17*pro*1.4 were also detected in periclinal and dorsal sides (Figure 15 D, F, I).

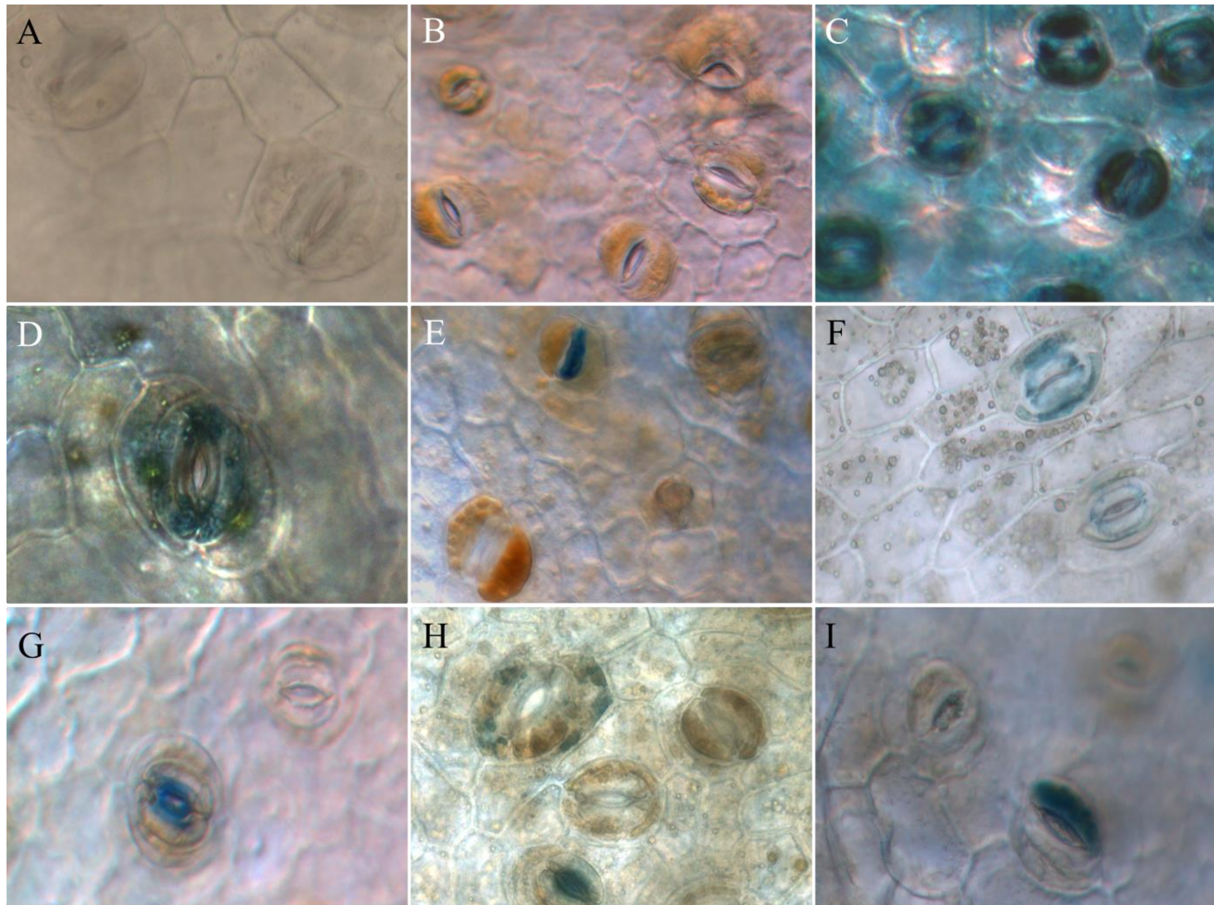


Figure 15 Histochemical detection of GUS activity in guard cells of leaves from *C. arabica* var. Caturra wild type and transgenic plants harboring *CcDREB1Dpro*:*GUS* cassettes. Guard cells visualized by bright field microscopy on abaxial region of coffee leaves. WT, non-transformed control (A). pBI101 line (GUS negative), negative control of transcriptional activity (B). pBI121 line (*CaMV35S*:*GUS*), positive control of transcriptional activity (C). *CcDREB1D*:*GUS* transgenic lines harboring pD14-hp15*pro*0.7 (D), pD14-hp16*pro*1.1 (E), pD22-hp17*pro*0.7 (F), pD14-hp15*pro*1.4 (G), pD14-hp16*pro*1.4 (H), pD22-hp17*pro*1.4 (I) cassettes. Leaves were sampled following 0, 3, 6, 12 or 24h of exposure to drought stress with osmotic pressure of -2.0 MPa (PEG) or 9% Relative Hygrometry (RH) .

Protein immunolocalization

Protein immunolocalization was used to confirm that blue staining was specifically associated to GUS proteins and to precise the location of GUS proteins at a cellular level. The immunolocalization of *β*-glucuronidase protein (GUS) was evaluated using an anti-*β*-glucuronidase polyclonal antibody and a secondary antibody conjugated to Alexa Fluor 488 fluorescent dye. Longitudinal and cross-sections of several leaves from *CcDREB1Dpro*:*GUS* constructions matching the staining pattern previously described for guard, epidermic and parenchyma cells were used for immunolocalization. Leaves sections were observed by confocal microscopy and *in vivo β*-glucuronidase localization

identified by spectral analysis combined with the Linear Unmixing technique with a multiphoton microscope at λ_{exc}=543 nm (Figure 16). Firstly, a spectral picture was obtained from stained sections in the Lambda mode (554–683 nm range) (Figure 16A). Further, an emission spectrum was collected on the region of interest (ROI) wherein GUS signal was detected by bright field microscopy. Three main peaks at 488, 543 and 633 nm were obtained in this spectral signature on ROI. Secondly, spectral acquisitions using the same optical parameters on not stained leaf sections of pBI101 line (control of transcription activity) allowed the collection of a reference spectrum for autofluorescence signal. The Linear Unmixing technique was applied to leaf sections using the reference spectra of pBI101 (Figures 16 and 17). This technique was used with an advanced iterative option and a residual channel. A red autofluorescence homogeneous signal was observed in leaf epidermis, mesophyll and guard cells being present even on the thickened inner wall (Figure 16B). A green heterogeneous signal located in some of parenchyma and guard cells with high intensity on the thickened inner wall was attributed to Alexa Fluor 488 (λ_{em}=519 nm) and thus to a β-glucuronidase epitope (Figure 16C). Finally, an overlay of the three spectrums was generated (Figure 16D).

Consistently, localization of signals highly condensed in thickened inner wall of guard cell confirmed the pattern of GUS protein location on the ventral pore edge side of stomata (Figure 17D and 17B). A fluorescence signal on the center of periclinal part in guard cells was located in cross-sections of leaf and matched the periclinal and dorsal GUS staining pattern previously described (Figure 17C). Importantly, green signals were not detected in all guard cells, alike shown in histochemical analysis, leading evidence of guard cells independent-activity (Figures 16C, 17A and B). GUS localization was also confirmed on subsidiary cells, spongy and palisade parenchymas cells, and epidermis cells (Figures 16D and E). Spectra acquisition of primary anticorps did not show signal in Alexa Fluor 488 emission spectrum, therefore confirming the exclusive emission of green signal by secondary anticorps coupled (Figure 16F). These findings confirmed a developmental expression pattern of *CcDREB1D* homologous promoter which confers tissues-specific activity for *CcDREB1D* promoter in coffee cotyledons and leaves ensuing plant development. Homologous pD14-hp15, pD14-hp16 and pD22-hp17 full-length promoters showed activity restricted to leaf epidermic, parenchyma and guard cells.

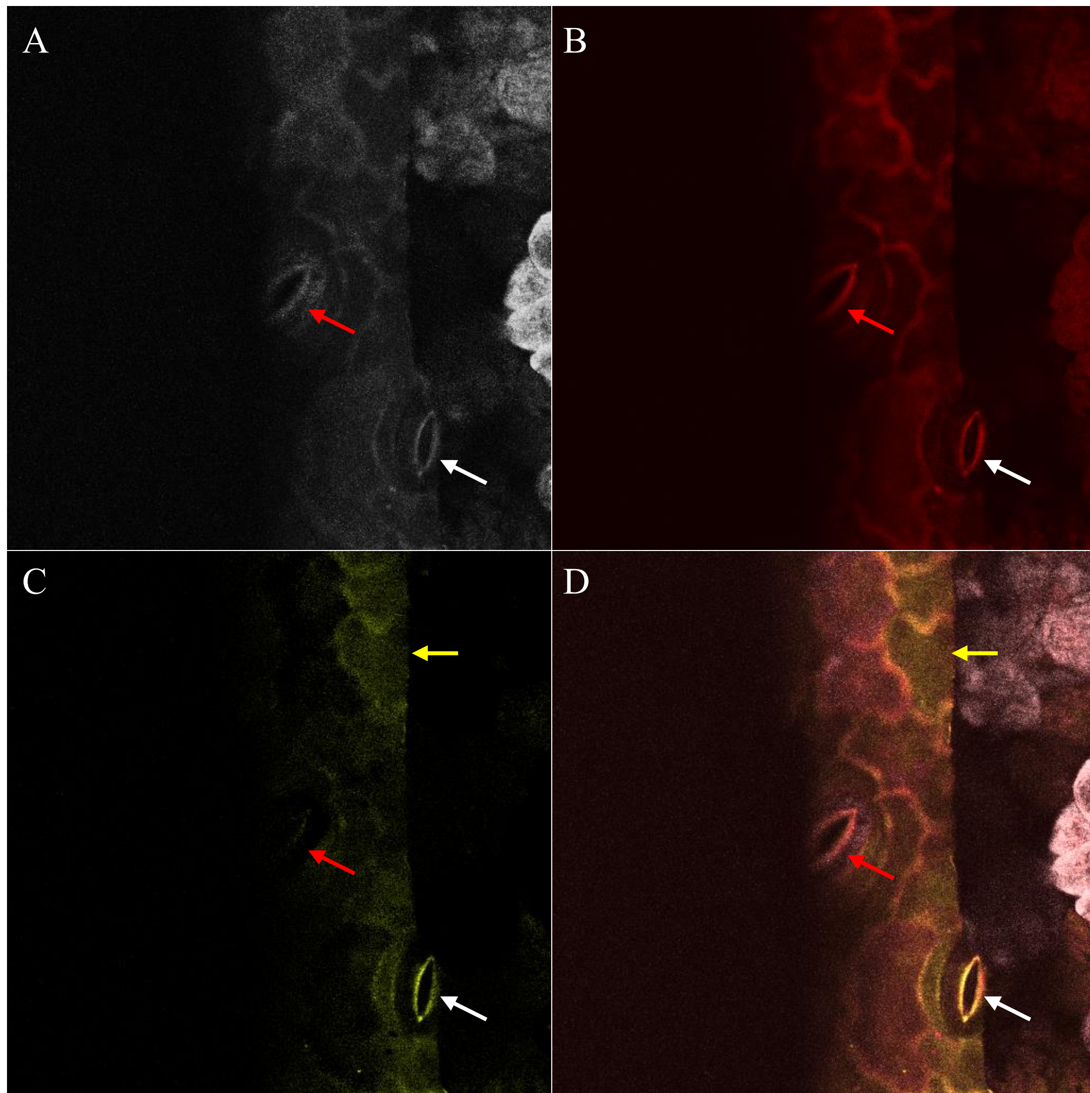


Figure 16 Confocal laser scanning microscopy (CLSM) micrographs of guard cells from longitudinal-section of leaf with immunolocalization of *β*-glucuronidase. Section from GUS stained coffee leaf was treated with anti-*β*-glucuronidase rabbit polyclonal antibody and then secondary anti-rabbit IgGs mouse antibody coupled to an Alexa Fluor® 488 probe. Micrograph deconvolution of spectra signals was achieved using Linear Unmixing technique. (A) Residual channel, spectral picture obtained from leaf section in the Lambda mode (554–683 nm range). (B) Red channel, guard cell and epidermic cell autofluorescence. An intense red autofluorescence detected in the ventral pore edge side of stomata (thickened inner wall) outside of Alexa Fluor 488 emission spectra. (C) Green channel, visible thickened inner wall of guard cell (white arrow) and epidermic cells (yellow arrow) exhibited green fluorescent for *β*-glucuronidase in the absence of red autoflurosence. (D) Overlay of the three channels. Guard cells tested positive and negative for *β*-glucuronidase green fluorescent are indicated by white and red arrows, respectively.

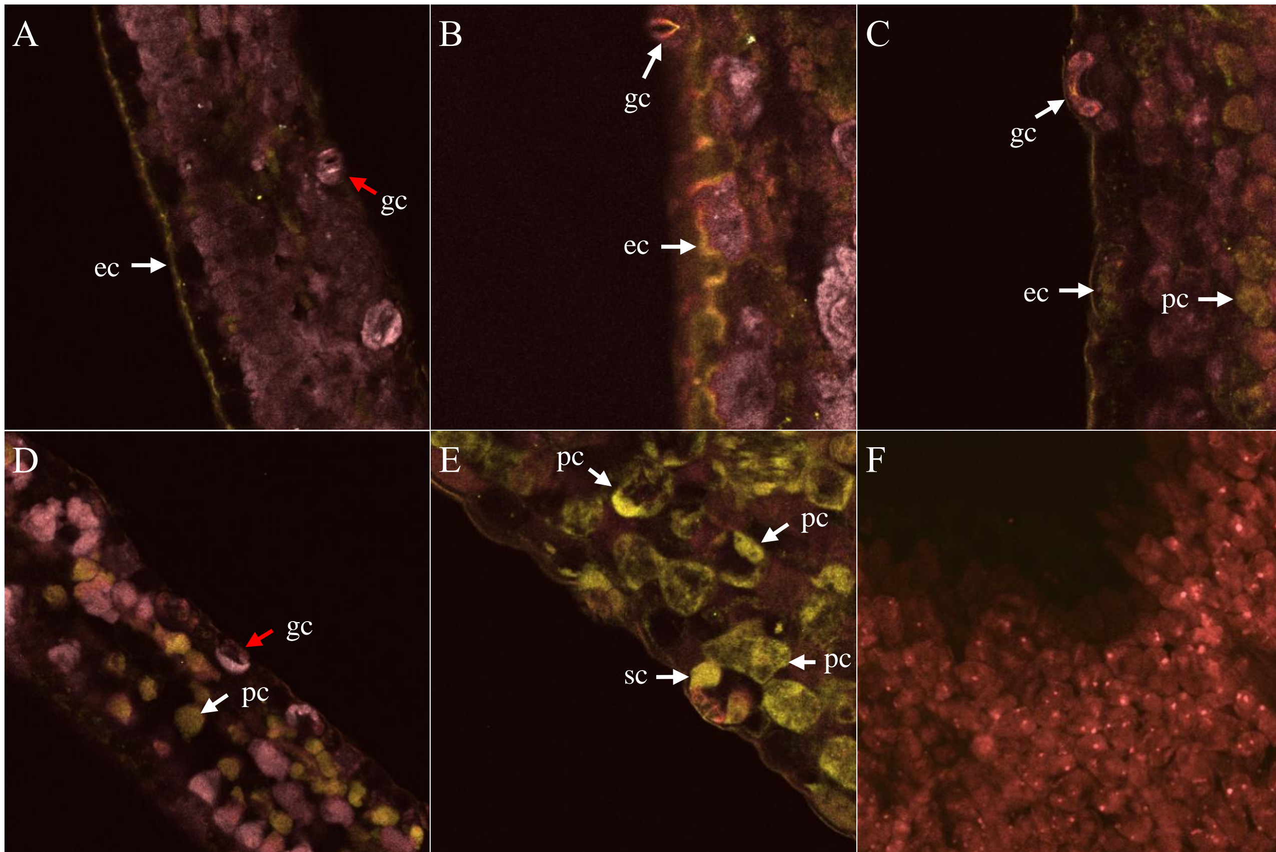


Figure 17 Confocal laser scanning microscopy (CLSM) micrographs of guard cells from cross-sections of leaf with immunolocalization of *β*-glucuronidase. Sections from GUS stained coffee leaves were treated with anti-*β*-glucuronidase rabbit antibody and then secondary anti-rabbit IgGs mouse antibody coupled to an Alexa Fluor® 488 probe. Micrograph deconvolution of spectra signals was achieved using Linear Unmixing technique. (A) Detection of *β*-glucuronidase fluorescence on epidermis cells and absence of fluorescent signal in guard cells. (B) Highly condensed signals in thickened inner wall of guard cell confirming the pattern of GUS protein location on the ventral pore edge side of stomata. (C) Fluorescent signal on the center of periclinal part in guard cells. (D and E) Localization of intense *β*-glucuronidase signal on subsidiary cells, spongy and palisade parenchymas cells. (F) Background signal from primary anticorps showing no spectra emission in Alexa Fluor® 488 fluorescence spectrum. Guard cells tested positive and negative for *β*-glucuronidase green fluorescent are indicated by white and red arrows, respectively. ec, epidermic cell; gc, guard cell; pc, parenchyma cell; sc, subsidiary cell.

The activity of the CcDREB1D promoter is up-regulated by relative humidity in coffee leaves of transgenic coffee plants

As we demonstrate herein, temporal and spatial analyses of *CcDREB1D* promoter activity revealed leaf-specific activity in pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 lines. We also report that full-length *CcDREB1Dpro*:*GUS* activity was observed strictly in epidermic, parenchyma and guard cells. Additionally, genetic variability present on the allelic promoter regions of pD14-hp15 and pD14-hp16, and homolog promoter region pD22-hp17 led to some differences in tissue modulation of *CcDREB1D* transcriptional activity, leading us to wonder if

allele-specific regulation of the *CcDREB1D* gene could be influenced by different drought stress stimulus. Therefore, *CcDREB1D* promoter activity was evaluated in time-course response to drought stress treatments by assessing GUS staining activity in guard cells. If considered that the frequency of stained guard cells is correlated with *CcDREB1D* promoter regulation of GUS transcriptional activity in *CcDREB1Dpro:GUS* cassettes, by calculating the frequency of GUS stained guard cells on abaxial region of leaves it is likely to estimate *CcDREB1D* promoter response to drought treatments. For each *CcDREB1Dpro:GUS* lines, frequency of GUS activity, obtained by the ratio of observed stained guard cells per total number of observed guard cells, was assessed in 24x36 mm area randomly distributed in six pre-delimited zones of coffee leaves (Figure 18). This hypothesis was verified by studying the time-course response of *CcDREB1Dpro:GUS* activity in four-months old coffee transgenic plants subjected to drought treatments. Sixty transgenic plants from eight independent-event (7 plants/event) arranged randomly in batch of 7 plants were independently exposed to 0, 3, 6, 12 and 24 hours of drought treatment. We first evaluated the frequency of GUS stained guard cells for *CcDREB1D* promoter constructs under osmotic stress in a MS medium supplied with polyethylene glycol (PEG) in required amount to reproduce a theoretical water potential Ψ_W equivalent to -2.0 MPa. When dissolved in the medium, PEG is uptaken by the root system and generates a positive osmotic gradient to water efflux from cells and by hydrostatic pressure it reaches leaf tissues and cells. Under PEG treatment, the frequency of GUS activity of the full-length promoter cassettes pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 increased following the first hours of osmotic stress reaching 10% and 20% maximum average after 6h and 12h, respectively, whereas GUS activity for the common allele pD14-hp15*pro*1.4 maintained stable under this stress stimulus (Figure 19A). Oppositely, under the same treatment the frequency of GUS activity from the truncated promoter cassette pD14-hp15*pro*0.7 of the common allele increased mainly in the earlier hours of stress achieving 15% of guard cells, whereas no apparent variation in GUS activity was detected during stress exposure in plants bearing either pD14-hp16*pro*1.1 or pD22-hp17*pro*0.7 cassettes (Figure 19A).

We next tested whether other environmental drought stimulus that is known to modulate stomatal activity could differentially induce *CcDREB1D* promoter activity. Plants transformed with pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 were subjected to a low 9% RH (Figure 19A). As observed for PEG treatment, RH also induced changes in frequency of GUS activity, however duration of drought stress accounted differently in frequencies variation. Interestingly, cassettes of promoter alleles from the drought^{T} clones, pD14-hp15*pro*1.4 and pD14-

hp16*pro*1.4 revealed completely opposite patterns to variation in GUS activity during stress. While pD14-hp16*pro*1.4 frequency of GUS activity increased gradually towards drought period, reaching 30% of guard cells after 24h of drought stress, pD14-hp15*pro*1.4 GUS activity ncreased rapidly (20-30% of guard cells) and decreased over the time course until a very low 5% frequency (Figure 19A). The pD22-hp17*pro*1.4 cassette, from drought susceptible clone 22, showed increasing frequency of GUS stained guard cells ensuing the first 12 hours of stress arriving to an average of 10% of guard cells, alike cassette pD14-hp16*pro*1.4 from the tolerant clone, but conversely to pD14-hp16*pro*1.4, pD22-hp17*pro*1.4 expression down-regulated shortly thereafter (Figure 19A).

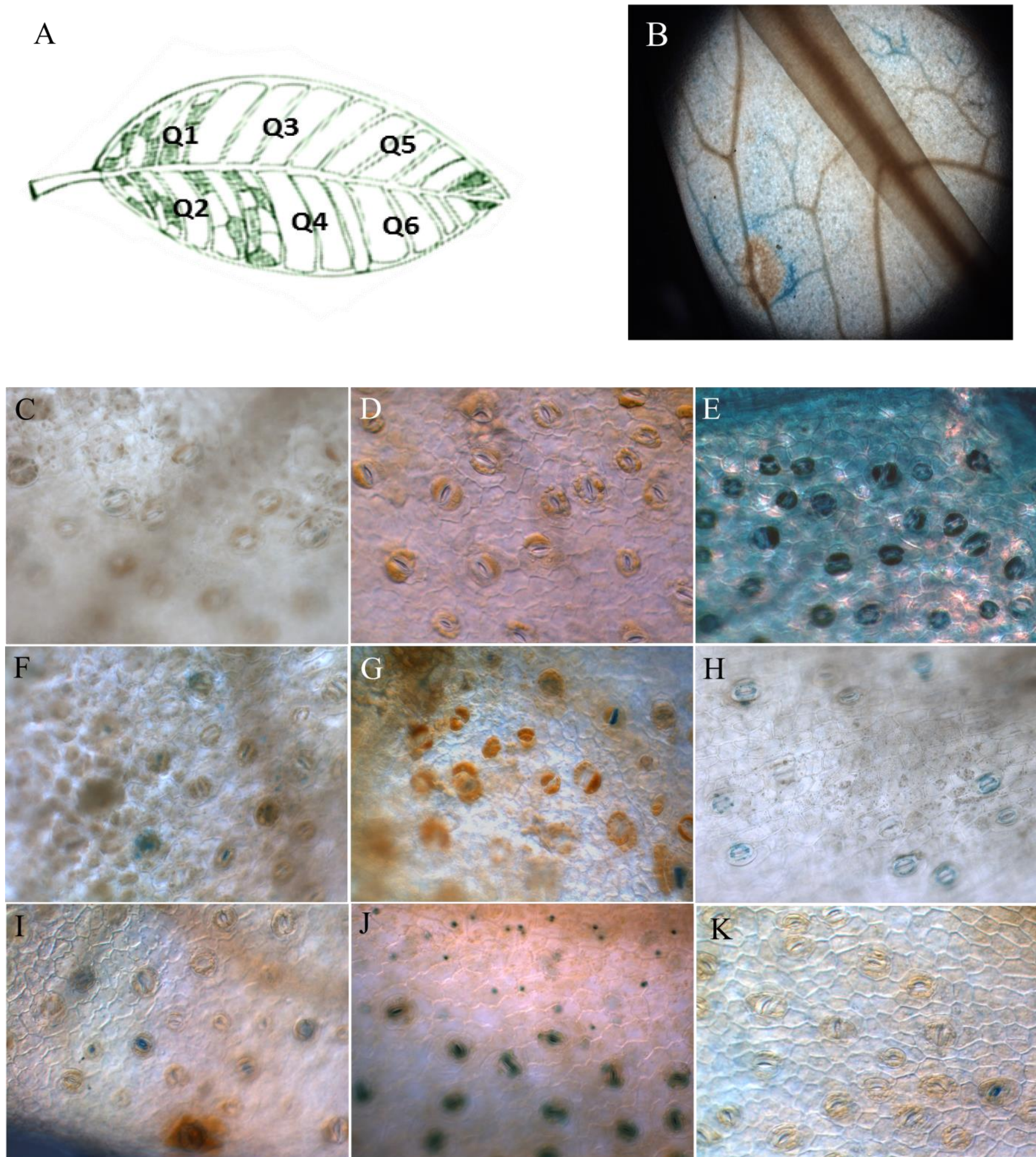
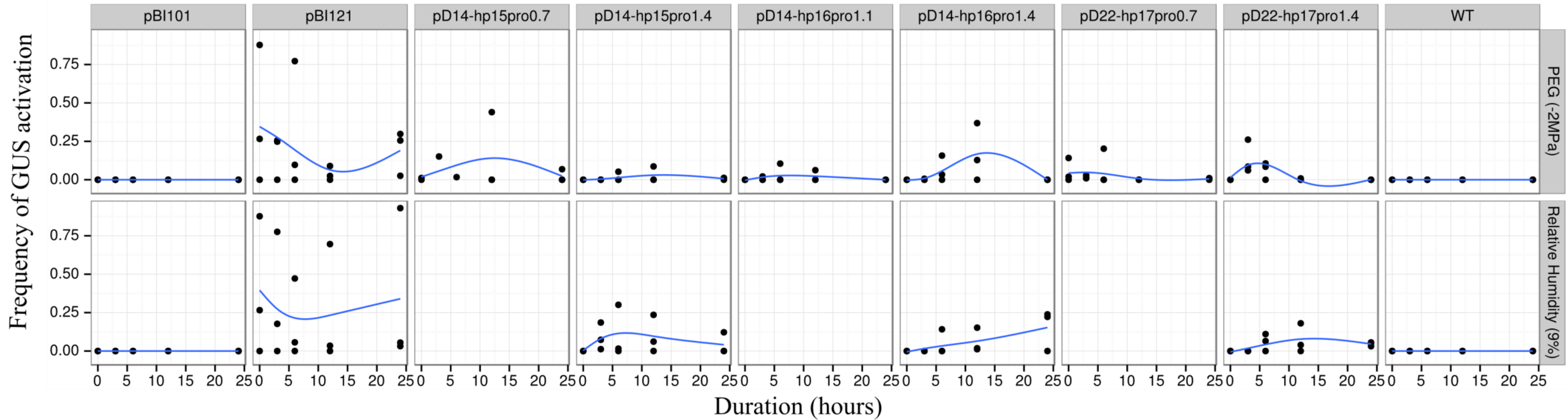


Figure 18 Histochemical detection of GUS activity in guard cells from *C. arabica* var. Caturra wild type and transgenic plants. Guard cells visualized by bright field microscopy on abaxial region of coffee leaves. Six 24x36 mm area randomly distributed in pre-delimited zones [Q] of coffee leaves were observed to estimate the frequency of GUS stained guard cells. Schematic representation of the six pre-delimited zones [Q1 to Q6] observed on abaxial region of coffee leaves (A). Bright field microscopy image of one of the six observed zones (B). WT, non-transformed control (C). pBI101 line (GUS negative), negative control of transcriptional activity (D). pBI121 line (*CaMV35S*:*GUS*), positive control of transcriptional activity (E). *CcDREB1D*:*GUS* transgenic lines harboring pD14-hp15*pro*0.7 (F), pD14-hp16*pro*1.1 (G), pD22-hp17*pro*0.7 (H), pD14-hp15*pro*1.4 (I), pD14-hp16*pro*1.4 (J), pD22-hp17*pro*1.4 (K) cassettes. Leaves were sampled following 0, 3, 6, 12 or 24h of exposure to drought stress with osmotic pressure of -2.0 MPa (PEG) or 9% Relative Hygrometry (RH) .

A



B

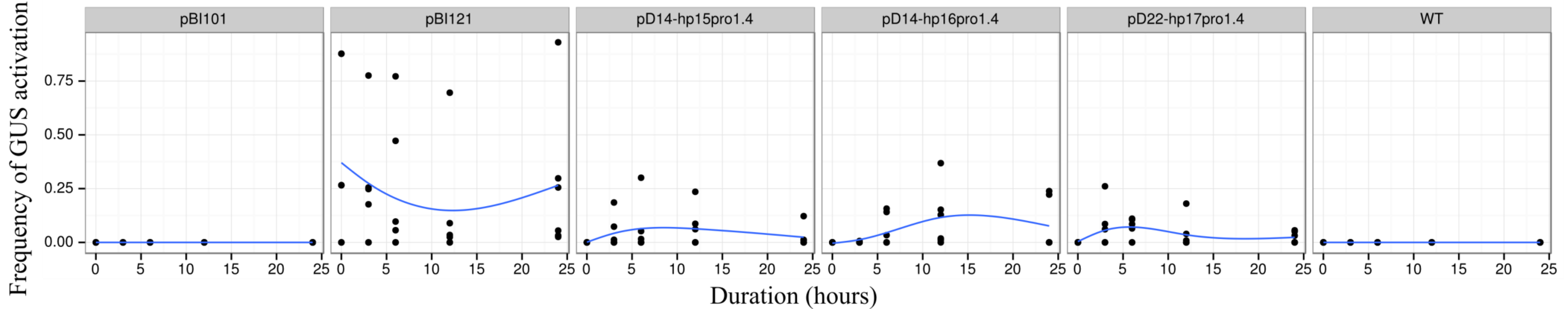


Figure 19 Drought stress-induced regulation of GUS activity in mature leaf stomata driven by *CcDREB1D* homologous promoters. (A) Proportion of GUS stained stomata in response to osmotic pressure of -2.0 MPa (PEG) or 9% Relative Hygrometry (RH) during a 24h drought stress. (B) pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 promoter activity matching the assembled of frequency data from the two drought stimuli. Frequency of GUS activity, obtained by the ratio of observed stained guard cells per total number of observed guard cells, is the average of six observations in 24x36 mm area on abaxial zones of coffee leaves. GUS activity was assessed in leaves of coffee plants following 0, 3, 6, 12 and 24 h of exposure to drought stress.

PEG and RH treatments are known to change water status of the leaf cells generating an osmotic stress that can be sensed by guard cells which use this stimulus to control leaf transpiration. Considering that *CcDREB1D* gene expression varies under water constrain conditions, irrigated and non-irrigated, in coffee plants (Marraccini et al., 2012), and expecting this gene expression variation to be likely a result of promoter activity regulation, statistical analyses were applied to verify if variation in GUS activity was systematically affected by drought stress duration. If that is the case, we expect to see a statistically significant variation in frequency of stained guard cells during time-course evolution of stress stimulus. When analyzed together the GUS stained guard cells frequency data from pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4, it was obvious that the observed variations in the frequencies under RH treatment were significantly correlated with the duration of the stress. This confirmed that exposure of coffee leaves to a low RH stress directly interferes on the activity of the promoters pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4. Such relation could not be confirmed for PEG treatment using this analysis (Figure 20). The same statistical analysis was further applied to the totality of GUS stained stomata frequency data from PEG and RH bioassays, to assess the effect of the pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 cassettes on the observed variations in GUS stained guard cells frequencies over drought stress period. We found that pD14-hp16*pro*1.4 cassette was the only to present a statistically significant variation in GUS activity which positively increased during drought stress for both treatments (Figure 20A). This confirmed that pD14-hp16*pro*1.4 transcriptional activity is up-regulated during the evolution of both water stress conditions (Figure 20B).

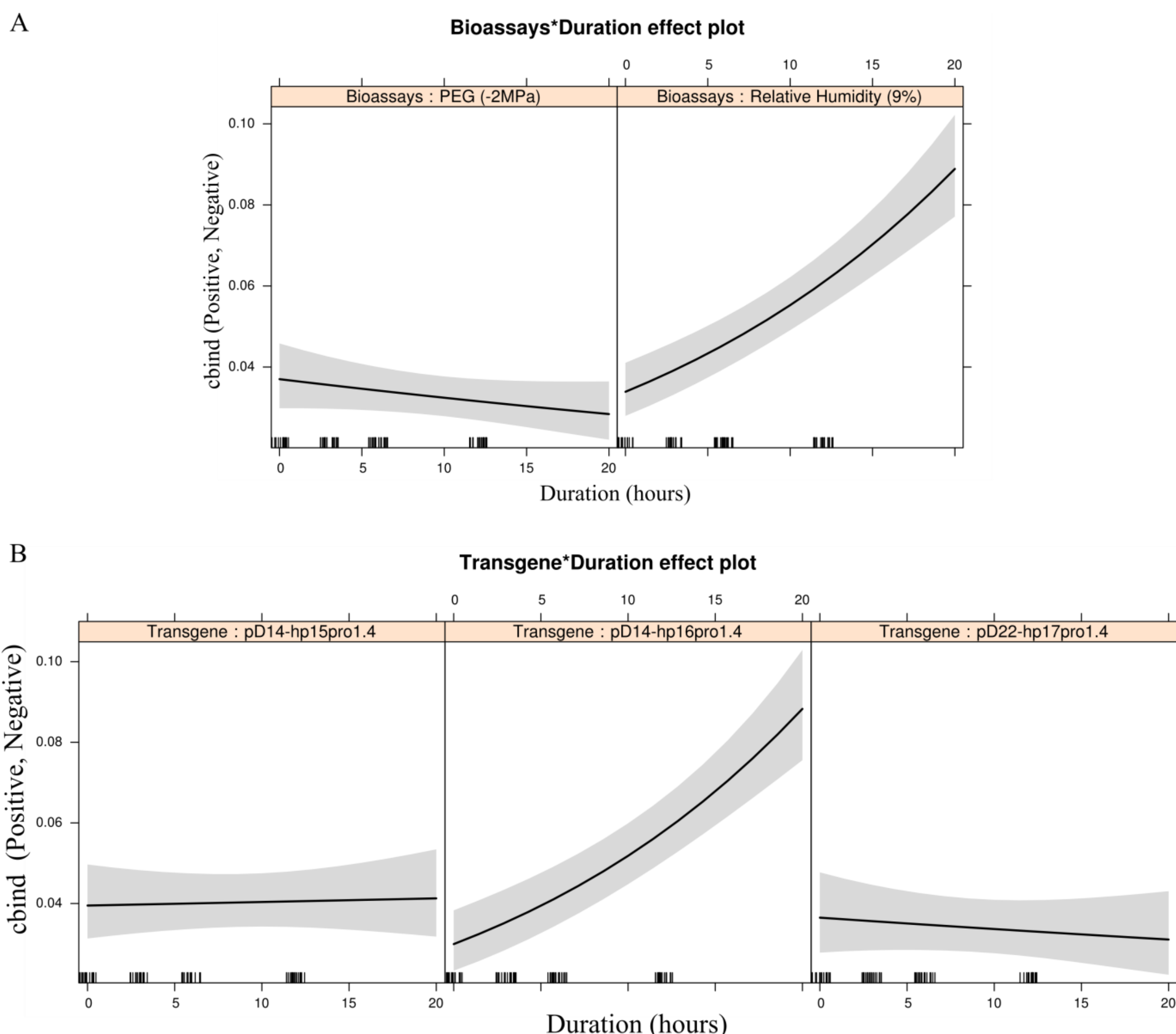


Figure 20 Effect of stress time-course in *CcDREB1D* promoter regulation. (A) Statistical analysis of variance between drought stress duration (hours) and totality of GUS activity frequencies in guard cells from pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 promoters for each treatment separately. (B) Statistical analysis of variance for the totality of GUS activity frequencies in guard cells from PEG and RH treatments combined showing that pD14-hp16*pro*1.4 is the only full promoter that presented a significant positive correlation between time-course and expression activity for each of drought stress conditions.

5. Discussion

One of the main objectives of drought adaptation research is to determine how plants sense water stress and convert this information into molecular mechanism to promote physiological responses that increase drought resistance. One approach to address this issue is to identify the actors in the perception and network response of drought stress. *Cis*-acting elements and *trans*-acting factors are major actors that work in the first wave of network response. *DREB* transcription factors genes are

strongly and rapidly up-regulated in response to drought stress in numerous plant species being ideal candidates to elucidate the mechanisms underlying the transmission of drought stress signals (Agarwal et al., 2006b). In coffee, an investigation of the molecular mechanisms that account for drought acclimation in green-house plants evidenced *CcDREB1D* as a candidate gene, among others, to drought tolerance (Marraccini et al., 2012; Vieira et al., 2013). Upon drought, coffee leaves showed a significant increase in *CcDREB1D* mRNA level compared to non-stress conditions, having an up-regulation even more evident in drought^T clone 14 than drought^S clone 22. Analysis of nucleotide diversity in *CcDREB1D* promoter and coding region of several accessions from *C. canephora* and *C. arabica* revealed a great diversity in promoter region of this gene and possible implication of this genetic variability in drought adaptation were postulated (Alves et al, accepted [see Chapter 1]). The present study concentrated the efforts to identify DNA regulatory elements present upstream coffee *DREB1D/CBF4* homolog genes and characterize the promoter activity in response to water stress. The results indicated a complex interaction of both positive and negative DNA regulatory elements in upstream region of *CcDREB1D* resulting in spatiotemporal- and tissue- specific expression during plant development. In addition, these results evidenced activity of *CcDREB1D* promoter in guard-cell and existence of nucleotide variability in this region governing differentially gene expression towards water stress stimulus. These results suggest the existence of a mechanism integrating guard-cell perception of water stress and *CcDREB1D* gene expression regulation.

Multiple cis-elements governing CcDREB1D expression

Gene expression is mostly controlled by dynamic interactions between *cis*-regulatory modules localized in the promoter sequence and their cognate transcription factors. To elucidate the regulatory elements involved in the different patterns of *CcDREB1D* regulation, homolog promoter sequences from clones 14 and 22 were isolated and studied. The results presented herein indicate that *CcDREB1D* promoter harbor several abiotic-stress responsive *cis*-elements apart from tissue-specific expression and light-responsive elements. The presence of ABA-dependent (ABRE) and ABA-

independent (DRE) elements indicate the possible activation of *CcDREB1D* by both networks responses as it has been predicted and confirmed for its homolog in *Arabidopsis* (Haake et al., 2002; Yamaguchi-Shinozaki and Shinozaki, 2005). Moreover, the co-occurrence within 100 bp of distance in parallel strand and oppositely direction of ABRE (-514/-521) and DRE (-437/-441) elements is the most active structural organization observed for these *cis*-acting elements in drought-responsive genes (Mishra et al., 2014). The presence of DRE regulatory element indicate *CcDREB1D* involvement in ABA-independent network of drought response, but considering that DRE share a common core with LTRE (low-temperature responsive element) it could be expected a functional implication of this gene in cold response as observed for its homologs and other subgroup members in several plants (Gilmour et al., 1998; Fowler and Thomashow, 2002; Dubouzet et al., 2003; Maruyama et al., 2004; Qin et al., 2004; Zhang et al., 2004; Ito et al., 2006; Chen et al., 2009; Mao and Chen, 2012). In this case, *CcDREB1D* should present other *cis*-elements known to compose cold-stress perception and transduction of cold-response in *DREB1/CBF* genes. This was exactly the case for the *CcDREB1D* promoters, that contain one ICEr2 box, target of bHLH transcription activator protein ICE2 (inducer of *CBF* expression 2), and seven MYC-related binding site (CNNNTG). For the latter, two of which contained the preferable target sequence (CAAATG) of another positive regulator ICE1 (inducer of *CBF* expression 1), a MYC-type bHLH protein. ICE1 and ICE2 play a key role in the perception of cold stress signals and activation of *DREB1/CBF* expression (Chinnusamy et al., 2003; Fursova et al., 2009). These proteins are constitutively expressed and respond to stress stimulus throw a ABA-independent pathway that implicate Ca^{+2} signaling and activation of kinase proteins leading to ICE phosphorylation and activation (Chinnusamy et al., 2007). Similarly, the CG-1 element (CCGCGT) found in -322/-317, is an up-regulator of *DREB1/CBF* expression that interacts with CAMTA transcription activator (Doherty et al., 2009). CAMTA proteins have four functional domains, an isoleucine/glutamine [IQ] domain for calmodulin (Ca^{+2}/CaM) binding, an ankyrin repeats for protein interactions, a threonin/isoleucine/glycine [TIG] domain with nonspecific DNA binding activity, and a CG-1 proteic domain which has specific DNA binding activity and recognize the CG-1 *cis*-element (VCGCGB) (Bouché et al., 2002). Together, IQ and CG-1 domains allow CAMTA protein to be one

of the first actors in *DREB1* stress-response ensuing Ca^{+2} signaling, due to its conjointly interaction with Ca^{+2} and CG-1 regulatory element. *CcDREB1D* promoter contains several MYB-binding sites that can act as enchancer or represor. Particulary, two MYB elements (DAACCA) have high binding affinity to MYB15, which can target *DREB1D/CBF* promoter and regulate negatively its expression (Agarwal et al., 2006a). The remaining seven MYB *cis*-elements are potential positive regulators of *DREB1/CBF* expression, specially three MYB2AT (YAACTG) binding sites required for AtMYB2 induction of salt and dehydration responsive genes (Nakagoshi et al., 1990; Urao et al., 1993; Abe et al., 1997; Abe et al., 2003). Not only abiotic stress responsive elements were identified in *CcDREB1D* promoter, *in silico* analyses revealed the occurance of DOF-binding sites, key determinants in mediating guard cell expression (Cominelli et al., 2011). Interstingly, two DOF sites clusters (-657/-758 and -1095/-1191 in Figure 7), composed respectively by three and five elements in sense (5'-3') or antisense (3'-5') position, were identified in *CcDREB1D* promoter. DOF (DNA-binding one zinc finger) are transcription factors family exclusively found in higher plants that mediate responses to signals and regulate many fundamental processes, including guard cell-specific gene expression (Plesch et al., 2001; Galbiati et al., 2008; Yang et al., 2008; Gardner et al., 2009).

Modular organization of DOF and light-induced elements account for preferential CcDREB1D guard cell expression

Organ and tissue specificity of the *CcDREB1D* promoter from three allelic forms (pD14-hp15, pD14-hp16 and pD22-hp17) of *C. canephora* were investigated in *C. arabica* transgenic plants. Analysis of several independent coffee lines carrying the *GUS* gene under the control of proximal (-1/-762), medium (-762/-1103) or full length (-1103/-1466) *CcDREB1D* promoters revealed a strong role of negative and tissue-specific regulatory elements in fine tuning regulation of reporter activity. The analysis of *CcDREB1Dpro:GUS* constructs in transgenic coffee plants clearly demonstrated that *CcDREB1D* full length promoter directs the expression of the gene in a temporal and stress-specific manner. We did not observe any activation of the *CcDREB1D* promoter in normal non stress conditions. Deletion analyzes of *CcDREB1D* upstream region from all allelic forms (pD14-hp15,

pD14-hp16 and pD22-hp17) showed that the most distal *cis*-elements (in the region -762/-1466) play a major role in driving leaf specific expression compared to others regulatory motifs for the most proximal region (-1/-762). The unspecific and strong GUS activity revealed by *CcDREB1D* proximal truncated promoters (pD14-hp15*pro*0.7and pD22-hp17*pro*0.7) contrasted enormously with the fine-tuned leaf specific regulation of full length promoters (pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4). Histochemical patterns and immunolocalization of GUS revealed tissue-specific expression of full length promoters in mesophyll, epidermic and guard-cells of leaves. In addition, stringency of *CcDREB1D* promoter activity to guard cells increased proportionally with inclusion of DOF binding sites from proximal promoters (pD14-hp15*pro*0.7and pD22-hp17*pro*0.7) to medium (pD14-hp16*pro*1.1) and full length promoters (pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4). These results can be explained by the increasing number of negative regulators and tissue-specific binding sites among the most distal *cis*-elements of *CcDREB1D* promoter. Of the six MYB occurring within -762/-1466, five are negative regulators and specially two are binding sites of MYB15. MYB15 were shown to repress *DREB1/CBF* genes (Agarwal et al., 2006a) but also to regulate guard-cell stomatal closure under water-deficit conditions inducing drought tolerance (Ding et al., 2009). The presence of repressor could reasonably explain the decrease in GUS activity observed in pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4 lines compared to pD14-hp15*pro*0.7and pD22-hp17*pro*0.7. Stronger evidence can be placed for the increasing leaf-specific activity from proximal (pD14-hp15*pro*0.7 and pD22-hp17*pro*0.7) to medium (pD14-hp16*pro*1.1) and full length promoters (pD14-hp15*pro*1.4, pD14-hp16*pro*1.4 and pD22-hp17*pro*1.4). This additively effect is in accordance with DOF cluster hypothesis postulated by Galbiati et al. (2005). Several studies have confirmed this hypothesis. Regions enriched in DOF *cis*-elements have been reported in several promoters controling guard-cell specific expression (Plesch et al., 2001; Cominelli et al., 2005; Galbiati et al., 2008; Yang et al., 2008; Gardner et al., 2009). A strong funtional cluster organization of DOF-binding sites in the promoter of *AtMYB60* gene specifically expressed in guard cells was revealed (Cominelli et al., 2011). The partially additive role of DOF-elements was clearly demonstrated by the combined mutagenesis of these sites in *AtMYB60:GUS* transgenic lines which

resulted in a drastically reduction of GUS activity. Although, in the same study, single or multiple artificial DOF-binding site in tandem were not able to drive guard cell expression. There seems to be a consensus that other *cis*-reulatory elements function conjoitnly with DOF-binding site to fine tune transcriptional regulation of expression in guard cell (Galbiati et al., 2008; Gardner et al., 2009; Cominelli et al., 2011). *In silico* analyses in *CcDREB1D* promoter highlighted three 10 bp motifs composed by one ligth-induced (GRWAAW) and one DOF-binding elements. Two of them are present in the medium promoter region (-762/-1103) and one in the most distal region (--1103/-1466), in addition, the most distal region also harboured two consecutive ligth-induced elements. Alike Cominelli et al. (2011) reported, the progressive inclusion of DOF binding site herein increased the preferential expression of GUS in guard cell. Interestingly, the series of *CcDREB1D* promoter deletions promoted also an increase of light-induced elements almost proportionally to DOF elements. By reviewing *AtMYB60* promoter the number of light-induced elements in sense (5'-3') and antisense (3'-5') direction increase proportionally to promoter activity, in special a cluster organization within -911/-1043 (Cominelli et al., 2011). Therefore, light-induced motifs could be candidate as additional *cis*-regulatory element that could act mutually with DOF elements in regulation of guard-cell specific expressed genes. Our result endorse the hypothesis of general evolutionary conservation of DOF cluster organization proposed by Cominelli *et al* (2011), adding the DOF cluster structure observed in coffee *CcDREB1D* promoter to the previously reported DOF cluster structures in grape *VvMYB60*, potato *KST1* and cotton *GbSLSP* promoters (Plesch et al., 2001; Galbiati et al., 2011; Han et al., 2013).

CcDREB1D promoter responds differentially to environmental conditions

In this study, we also investigated the regulation of the homologs *CcDREB1D* promoter (pD14-hp15, pD14-hp16 and pD22-hp17) activity in response to water deficit mimicked by either vapor pressure deficit (RH: relative humidity of 9%) or low water potential (PEG: theoretical Ψ_W equivalent to -2.0 MPa). The effect of drought treatments on promoter activity was evaluated by accessing the frequency of GUS activity in guard cells. A reduction in Ψ_W increased the frequency of promoter activity of full

length allele from drought^{T} (allele^{T}) pD14-hp16 and drought^{S} (allele^{S}) pD22-hp17. However, the expression of the common allele (allele^{C}: pD14-hp15) of *CcDREB1Dpro* was induced mainly in the truncated version (pD14-hp15*pro*0.7) of this promoter compared to expression driven by pD14-hp15*pro*1.4 that appeared nule or very low. Diversely, increased vapor pressure deficit (*D*) led to upregulation of the full length promoters allele^{C} pD14-hp15, as well as allele^{T} pD14-hp16 and allele^{S} pD22-hp17. Drought stress treatment affected differently the spatio-temporal activity of allele^{T} and allele^{S} of *CcDREB1D* promoters. According to stress duration, our results showed that allele^{T} pD14-hp16 and allele^{S} pD22-hp17 presented a relatively early promoter activity in response to reduced Ψ_W compared to increased *D*. This could be due to the involvement of ABRE regulatory elements in *CcDREB1D* promoter activation. Regarding the results obtained under PEG treatment, it is well known that this chemical compound mediates low Ψ_W by reducing osmotic potential of nutrient solution, therefore mimicking low soil water potential and leading to a systemic water deficit in the whole plant (Blum, 2014). Changes in soil moisture can indirectly affect stomata control through a hydraulic feedback involving hormone signaling such as ABA. The production of ABA in roots (major organ producer of this hormone) following the first signal of drought stress, would rapidly amplify drought stress signal and achieve guard cell triggering *CcDREB1D* promoter activity. Noteworthy, an evidence of such rapid feedback response has been constantly reported for *C. canephora* clones with rapid recovery of Ψ_W accompanied by increase in stomatal conductance (G_S) after re-watering (Pinheiro et al., 2005; Marraccini et al., 2012; Vieira et al., 2013).

If ABA is really contributing to amplify drought stress signal and induce an earlier response of *CcDREB1D* promoter through root sensing of water stress, an inverse drought stimuli where water stress is "sensed" by shoot rather than roots is expected to delay the response of *CcDREB1D* promoter to drought stress. To simulate such condition, a water deficit was generated by increasing vapor pressure deficit (*D*) and maintaining high soil water potential. Low relative humidity increase the difference between the saturation vapor of the leaves and that in the atmosphere creating a leaf-to-air vapor pressure deficit (leaf-to-air *D*) (Kerstiens, 1996; Carr, 2001; DaMatta, 2004; DaMatta and Ramalho, 2006; Farooq et al., 2009). In turn, increased leaf-to-air *D* generates a high evaporative

demand elevating transpiration rate (*E*). To keep photosynthetic rate with increasing vapor pressure deficit, plants guard cells sensing of stomatal transpiration activates a feedback response to maintain *E* as evaporative demand increase. As expected, the gradual increases in activity of allele^{T} pD14-hp16 and allele^{S} pD22-hp17 *CcDREB1D* promoters towards drought stress progression under high *D* support such mechanism. In addition, allele^{T} pD14-hp16 upregulation of GUS activity frequency in guard cell lasted the entire period of drought stress (24 h). This was not observed for the allele^{S} pD22-hp17 for what the increase of GUS activity was maintained only during the first 12 h of drought stress. Considering the role of leaf in plant water-use control, leaves can lose water through the cuticle, leading to loss of turgor from epidermis and mesophyll cells, and guard cell stomata pores (Kerstiens, 1996). Evaporation of water from the intercellular spaces, mesophyll, or from stomata affect *E* efficiency and subsequently plants water-use (Kerstiens, 1996). Therefore, drought-tolerant plants generally have a greater water-use and larger and deeper root systems that allow deep soil moisture extraction (Blum, 2009). The same physiological and morphological remarks were made when comparing drought^{T} clones of *C. canephora* with drought^{S} ones (Pinheiro et al., 2005). Several studies showed that effective use of water (EUW) can lead to yield improvement under drought rather than water use efficiency (WUE) (Pinheiro et al., 2005; Blum, 2009). The activity of *CcDREB1D* promoter in epidermic, mesophyll and guard cells observed during our study point out the importance of this gene in the network of molecular responses involved in coffe regarding drought acclimation.

Linking guard cell feedback response to atmosphere changes and CcDREB1D expression

It is well established that transport of ions and water through channel proteins across the plasma and vacuolar membranes changes turgor and guard cell volume, thereby regulating stomatal aperture (Raschke et al., 1988; Schroeder et al., 2001; Pandey et al., 2007). Changes in guard cell turgor can induced release of calcium from both extracellular and vacuolar stores increasing cytosolic calcium $[Ca^{2+}]cyt$ (Kim et al., 2010). A connection between the changes of calcium concentration and *DREB/CBF* gene expression has been demonstrated in *Arabidopsis*, the rapid influx of calcium into

the cytosol being required for normal induction of the *CBF* target genes (Knight et al., 1996; Tähtiharju et al., 1997). Environmental stimuli (e.g. cold, circadian clock), leading to oscillations of cytosolic calcium, were shown to induce expression of *DREB1/CBF* genes (Dodd et al., 2005 and 2006). Circadian-gated expression of *CBF* genes also occurs in tomato, suggesting that this is a highly conserved form of regulation and of fundamental importance (Pennycooke et al., 2008). As previously mentioned, the regulation of *CBF1* and *CBF2* gene expression involves calmodulin-binding CAMTA transcription factors (Doherty et al., 2009). For example, CAMTA3 has a role in stress-induced expression of *DREB1/CBF* genes and genes rapidly induced in response to $[Ca^{2+}]$cyt oscillation are enriched in CG-1 binding element. The two proteins ICE1 and ICE2 also implicated in activation *DREB1D/CBF* gene expression, are constitutively expressed but postranslationaly activated by calcium pathway (Thomashow, 2010). Based on these observations and on our results, we can suggest a regulation of *CcDREB1D* gene expression by guard cell feedback in response to atmosphere changes, for example mediated by humidity that increases leaf-to-air *D* and subsequently *E*. In that case, as *D* increases, water loss from guard cells increases and guard cell turgor declines. Therefore, in order to maintain stomatal conductance (g_S), local gradients of water potential at the mesophyll/epidermis/guard cell are scaled rather than whole leaf water potential (Kerstiens, 1996). Water content changes occurring in epidermis, mesophyll and guard cell can elevate $[Ca^{2+}]$cyt concentration, which leads to the formation of a calcium-calmodulin-CAMTA3 complex and can also activate posttranslationaly ICE1 proteins through SIZ1-mediated sumoylation (Allen et al., 2001; Yang et al., 2003; Li et al., 2004; Mahajan and Tuteja, 2005; Doherty et al., 2009; Laanemets et al., 2013). Once actived, CAMTA or ICE proteins could stimulate the transcription of *CcDREB1D* gene after binding to CG-1 and MYC DNA boxes previously identified in these promoters (Figure 7). Even if further experiments are required to validate the existence of such regulatory mechanism, the activation of drought stress-responsive genes expressed in guard cells through calcium signaling might explain how guard cell maintained turgor and stomatal conductance (g_S) when leaf water potential decreases and leaf-to-air vapor deficit increases in the morning.

6. References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell **15**: 63–78

Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K (1997) Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell **9**: 1859–1868

Agarwal M, Hao Y, Kapoor A, Dong C-H, Fujii H, Zheng X, Zhu J-K (2006a) A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. J Biol Chem **281**: 37636–37645

Agarwal PK, Agarwal P, Reddy MK, Sopory SK (2006b) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep **25**: 1263–1274

Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, Singh NK (2012) DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. J Genet **91**: 385–395

Albuquerque EVS, Cunha WG, Barbosa AEAD, Costa PM, Teixeira JB, Vianna GR, Cabral GB, Fernandez D, Grossi-de-Sa MF (2009) Transgenic coffee fruits from *Coffea arabica* genetically modified by bombardment. Vitr Cell Dev Biol - Plant **45**: 532–539

Ashraf M (2010) Inducing drought tolerance in plants: Recent advances. Biotechnol Adv **28**: 169–183

Barton C, Adams T, Zarowitz M (1991) Stable transformation of foreign DNA into *Coffea arabica* plants. Proc Intl Conf Coffee Sci **14**: 460–464

Bihani P, Char B, Bhargava S (2010) Transgenic expression of sorghum *DREB2* in rice improves tolerance and yield under water limitation. J Agric Sci **149**: 95–101

Blum A (2009) Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. F Crop Res **112**: 119–123

Blum A (2014) Genomics for drought resistance – getting down to earth. Funct Plant Biol **41**: 1191-1198

Bouché N, Scharlat A, Snedden W, Bouchez D, Fromm H (2002) A novel family of calmodulin-binding transcription activators in multicellular organisms. J Biol Chem **277**: 21851–21861

Canche-Moo RLR, Ku-Gonzalez A, Burgeff C, Loyola-Vargas VM, Rodríguez-Zapata LC, Castaño E (2006) Genetic transformation of *Coffea canephora* by vacuum infiltration. Plant Cell Tissue Organ Cult **84**: 373–377

Chang W-C, Lee T-Y, Huang H-D, Huang H-Y, Pan R-L (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial *cis*-regulatory elements with distance constraint in plant gene groups. BMC Genomics **9**: 561p

Chen H, Je J, Song C, Hwang JE, Lim CO (2012) A proximal promoter region of *Arabidopsis DREB2C* confers tissue-specific expression under heat stress. J Integr Plant Biol **54**: 640–651

Chen M, Xu Z, Xia L, Li L, Cheng X, Dong J, Wang Q, Ma Y (2009) Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, *GmDREB3*, in soybean (*Glycine max* L.). J Exp Bot **60**: 121–135

Chinnusamy V, Ohta M, Kanrar S, Lee B, Hong X, Agarwal M, Zhu J (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes Dev **17**: 1043–1054

Chinnusamy V, Zhu J, Zhu J-K (2007) Cold stress regulation of gene expression in plants. Trends Plant Sci **12**: 444–451

Cominelli E, Galbiati M, Albertini A, Fornara F, Conti L, Coupland G, Tonelli C (2011) DOF-binding sites additively contribute to guard cell-specificity of AtMYB60 promoter. BMC Plant Biol **11**: 162p

Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL, Tonelli C (2005) A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. Curr Biol **15**: 1196–2000

Da Costa, Silva O (1994) CG-1, a parsley light-induced DNA-binding protein. Plant Mol Biol **25**: 921–924

Cotta MG, Barros LMG, de Almeida JD, de Lamotte F, Barbosa E a, Vieira NG, Alves GSC, Vinecky F, Andrade AC, Marraccini P (2014) Lipid transfer proteins in coffee: isolation of *Coffea* orthologs, *Coffea arabica* homeologs, expression during coffee fruit development and promoter analysis in transgenic tobacco plants. Plant Mol Biol **85**: 11–31

Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2013) Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol **60**: 483–501

Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, Poncet V, Anthony F, Dufour M, Leroy T (2008) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome **51**: 50–63

DaMatta F, Maestri M, Barros R, Regazzi A (1993) Water relations of coffee leaves (*Coffea arabica* and *C. canephora*) in response to drought. J Hortic Sci Biotechnol **68**: 741–746

DaMatta FM (2004) Exploring drought tolerance in coffee: a physiological approach with some insights for plant breeding. Brazilian J Plant Physiol **16**: 1–6

DaMatta FM, Chaves ARMM, Pinheiro HA, Ducatti C, Loureiro ME (2003) Drought tolerance of two field-grown clones of *Coffea canephora*. Plant Sci **164**: 111–117

DaMatta FM, Ramalho JDC (2006) Impacts of drought and temperature stress on coffee physiology and production: a review. Brazilian J Plant Physiol **18**: 55–81

Ding Z, Li S, An X, Liu X, Qin H, Wang D (2009) Transgenic expression of *MYB15* confers enhanced sensitivity to abscisic acid and improved drought tolerance in *Arabidopsis thaliana*. J Genet Genomics **36**: 17–29

Dodd AN, Jakobsen MK, Baker AJ, Telzerow A, Hou S-W, Laplaze L, Barrot L, Poethig RS, Haseloff J, Webb AAR (2006) Time of day modulates low-temperature Ca signals in *Arabidopsis*. Plant J **48**: 962–73

Dodd AN, Love J, Webb AAR (2005) The plant clock shows its metal: circadian regulation of cytosolic free Ca(2+). Trends Plant Sci **10**: 15–21

Doherty CJ, Van Buskirk HA, Myers SJ, Thomashow MF (2009) Roles for *Arabidopsis* CAMTA transcription factors in cold-regulated gene expression and freezing tolerance. Plant Cell **21**: 972–984

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J **33**: 751–763

Etienne H (2005) Somatic embryogenesis protocol: Coffee (*Coffea arabica* L . and *C . canephora* P.). *In* SM Jain, PK Gupta, eds, Protoc. Somat. Embryog. Woody Plants. Springer, Netherlands, pp 167–179

Ferrão R, Fonseca A, Silveira J, Ferrão M, Bragança S (2000) EMCAPA 8141 - Robustão Capixaba, variedade clonal de café conilon tolerante à seca, desenvolvida para o estado do Espírito Santo. Rev Ceres **47**: 555–560

Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell **14**: 1675–1690

Fursova O V, Pogorelko G V, Tarasov VA (2009) Identification of *ICE2*, a gene involved in cold acclimation which determines freezing tolerance in *Arabidopsis thaliana*. Gene **429**: 98–103

Galbiati M, Matus J, Francia P, Rusconi F, Cañón P, Medina C, Conti L, Cominelli E, Tonelli C, Arce-Johnson P (2011) The grapevine guard cell-related VvMYB60 transcription factor is involved in the regulation of stomatal activity and is differentially expressed in response to ABA and osmotic stress. BMC Plant Biol **11**: 142p

Galbiati M, Simoni L, Pavesi G, Cominelli E, Francia P, Vavasseur A, Nelson T, Bevan M, Tonelli C (2008) Gene trap lines identify *Arabidopsis* genes expressed in stomatal guard cells. Plant J **53**: 750–762

Gardner MJ, Baker AJ, Assie JM, Poethig RS, Haseloff JP, Webb AAR (2009) GAL4 GFP enhancer trap lines for analysis of stomatal guard cell development and gene expression. J Exp Bot **60**: 213–226

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. Plant J **16**: 433–442

Gutha LR, Reddy AR (2008a) Rice DREB1B promoter shows distinct stress-specific responses, and the overexpression of cDNA in tobacco confers improved abiotic and biotic stress tolerance. Plant Mol Biol **68**: 533–555

Gutha LR, Reddy AR (2008b) Rice DREB1B promoter shows distinct stress-specific responses, and the overexpression of cDNA in tobacco confers improved abiotic and biotic stress tolerance. Plant Mol Biol **68**: 533–555

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol **130**: 639–648

Han L, Han YN, Xiao XG (2013) Truncated Cotton Subtilase Promoter Directs Guard Cell-Specific Expression of Foreign Genes in Tobacco and *Arabidopsis*. PLoS One **8**: 1–10

Hardy A (2010) Candidate stress response genes for developing commercial drought tolerant crops ABA-induced transcription factors. **6**: 54–58

Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2006) Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. Plant Cell Physiol **47**: 141–153

Josis P, Ndayishimiye V, Rénard C (1982) Etude des relations hydriques chez *Coffea arabica* L. II. Évaluation de la resistance à la secheresse de divers cultivars à Ghisa (Burundi). Café Cacao Thé **27**: 275–282

Kasuga M, Miura S, Shinozaki K, Yamaguchi-Shinozaki K (2004) A combination of the *Arabidopsis* DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. Plant Cell Physiol **45**: 346–350

Kerstiens G (1996) Cuticular water permeability and its physiological significance. J Exp Bot **47**: 1813–1832

Kim J-S, Mizoi J, Yoshida T, Fujita Y, Nakajima J, Ohori T, Todaka D, Nakashima K, Hirayama T, Shinozaki K, et al. (2011) An ABRE promoter sequence is involved in osmotic stress-responsive expression of the *DREB2A* gene, which encodes a transcription factor regulating drought-inducible genes in *Arabidopsis*. Plant Cell Physiol **52**: 2136–2146

Kim T-H, Böhmer M, Hu H, Nishimura N, Schroeder JI (2010) Guard cell signal transduction network: advances in understanding abscisic acid, CO2, and Ca2+ signaling. Annu Rev Plant Biol **61**: 561–591

Knight H, Trewavas AJ, Knight MR (1996) Cold calcium signaling in *Arabidopsis* involves two cellular pools and a change in calcium signature after acclimation. Plant Cell **8**: 489–503

Krishnaswamy S, Verma S, Rahman MH, Kav NN V (2011) Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in *Arabidopsis*. Plant Mol Biol **75**: 107–127

Lambot C, Crouzillat D, Fonseca A, Leloup V, Haiduc A, Broun P, Pétiard V, Ferrão M, Ferrão R (2008) Evaluation of conilons for genetic diversity, cup quality and biochemical composition. Proc. 22th Int. Sci. Colloq. Coffee. Campinas, SP, Brazil, p A128

Lashermes P, Andrade AC, Etienne H (2008) Genomics of coffee one of the world's largest traded commodities. *In* PH Moore, R Ming, eds, Genomics Trop. Crop plants. Springer New York, pp 203–226

Lata C, Prasad M (2011) Role of DREBs in regulation of abiotic stress responses in plants. J Exp Bot **62**: 4731–48

Lim CJ, Hwang JE, Chen H, Hong JK, Yang K, Choi MS, Lee KO, Chung WS, Lee SY, Lim CO (2007) Over-expression of the *Arabidopsis* DRE/CRT-binding transcription factor DREB2C enhances thermotolerance. Biochem Biophys Res Commun **362**: 431–436

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: an overview. Arch Biochem Biophys **444**: 139–158

Mao D, Chen C (2012) Colinearity and similar expression pattern of rice DREB1s reveal their functional conservation in the cold-responsive pathway. PLoS One **7**: e47275

Marraccini P, Deshayes A, Pétiard V, Rogers WJ (1999) Molecular cloning of the complete 11S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. Plant Physiol Biochem **37**: 273–282

Marraccini P, Vinecky F, Alves GSC, Ramos HJO, Elbelt S, Vieira NG, Carneiro FA, Sujii PS, Alekcevetch JC, Silva VA, et al. (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. J Exp Bot **63**: 4191–4212

Maruyama K, Sakuma Y, Kasuga M, Ito Y, Seki M, Goda H, Shimada Y, Yoshida S, Shinozaki K, Yamaguchi-Shinozaki K (2004) Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. Plant J **38**: 982–993

Mishra S, Shukla A, Upadhyay S, Sanchita, Sharma P, Singh S, Phukan UJ, Meena A, Khan F, Tripathi V, et al (2014) Identification, occurrence, and validation of DRE and ABRE *Cis*-regulatory motifs in the promoter regions of genes of *Arabidopsis thaliana*. J Integr Plant Biol **56**: 388–399

Montagnon C, Cubry P, Leroy T (2012) Amélioration génétique du caféier *Coffea canephora* Pierre: connaissances acquises, stratégies et perspectives. Cah Agric **21**: 143–153

Montagnon C, Leroy T (1993) Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de Côte d'Ivoire appartenant à différents groupes génétiques. Café Cacao Thé **37**: 179–190

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant **15**: 473–497

Nakagoshi H, Nagase T, Kanei-Ishii C, Ueno Y, Ishii S (1990) Binding of the *c-myb* proto-oncogene product to the simian virus 40 enhancer stimulates transcription. J Biol Chem **265**: 3479–3483

Nakamura M, Tsunoda T, Obokata J (2002) Photosynthesis nuclear genes generally lack TATA-boxes: A tobacco photosystem I gene responds to light through an initiator. Plant J **29**: 1–10

Nakashima K, Shinwari ZK, Sakuma Y, Seki M, Miura S, Shinozaki K, Yamaguchi-Shinozaki K (2000) Organization and expression of two *Arabidopsis DREB2* genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. Plant Mol Biol **42**: 657–665

Novillo F, Alonso JM, Ecker JR, Salinas J (2004) CBF2/DREB1C is a negative regulator of CBF1/DREB1B and CBF3/DREB1A expression and plays a central role in stress tolerance in *Arabidopsis*. Proc Natl Acad Sci U S A **101**: 3985–3990

Nunes MA (1976) Water relations in coffee. Significance of plant water deficits to growth and yield: A review. J Coffee Res **6**: 4–21

Pandey S, Zhang W, Assmann SM (2007) Roles of ion channels and transporters in guard cell signal transduction. FEBS Lett **581**: 2325–2336

Pennycooke JC, Cheng H, Roberts SM, Yang Q, Rhee SY, Stockinger EJ (2008) The low temperature-responsive, *Solanum CBF1* genes maintain high identity in their upstream regions in a genomic environment undergoing gene duplications, deletions, and rearrangements. Plant Mol Biol **67**: 483–497

Petitot A-S, Barsalobres-Cavallari C, Ramiro D, Albuquerque Freire E, Etienne H, Fernandez D (2013) Promoter analysis of the WRKY transcription factors CaWRKY1a and CaWRKY1b homoeologous genes in coffee (*Coffea arabica*). Plant Cell Rep **32**: 1263–1276

Pinheiro HA, Damatta FM, Chaves ARM, Loureiro ME, Ducatti C (2005) Drought tolerance is associated with rooting depth and stomatal control of water use in clones of *Coffea canephora*. Ann Bot **96**: 101–108

Pinheiro HA, DaMatta FM, Chaves ARM, Fontes EPB, Loureiro ME (2004) Drought tolerance in relation to protection against oxidative stress in clones of *Coffea canephora* subjected to long-term drought. Plant Sci **167**: 1307–1314

Plesch G, Ehrhardt T, Mueller-Roeber B (2001) Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. Plant J **28**: 455–464

Praxedes SC, DaMatta FM, Loureiro ME, G. Ferrão MA, Cordeiro AT (2006) Effects of long-term soil drought on photosynthesis and carbohydrate metabolism in mature robusta coffee (*Coffea canephora* Pierre var. kouillou) leaves. Environ Exp Bot **56**: 263–273

Qin F, Sakuma Y, Li J, Liu Q, Li Y-Q, Shinozaki K, Yamaguchi-Shinozaki K (2004) Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. Plant Cell Physiol **45**: 1042–1052

Raschke K, Hedrich R, Reckmann U, Schroeder JI (1988) Exploring biophysical and biochemical components of the osmotic motor that drives stomatal movement. Bot Acta **101**: 283–294

Ribas AF, Dechamp E, Champion A, Bertrand B, Combes M, Verdeil J-L, Lapeyre F, Lashermes P, Etienne H (2011) *Agrobacterium*-mediated genetic transformation of *Coffea arabica* (L.) is greatly enhanced by using established embryogenic callus cultures. BMC Plant Biol **11**: 92p

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290**: 998–1009

Schroeder JI, Allen GJ, Hugouvieux V, Kwak JM, Waner D (2001) Guard cell signal transduction. Annu Rev Plant Physiol Plant Mol Biol **52**: 627–658

Shahmuradov IA, Solovyev V V, Gammerman AJ (2005) Plant promoter prediction with confidence estimation. Nucleic Acids Res **33**: 1069–1076

Shanker A, Venkateswarlu B (2011) Abiotic stress response in plants-physiological, biochemical and genetic perspectives. 358p

Spiral J, Thierry C, Paillard M, Pétiard V (1993) Obtention de plantules de *Coffea canephora* Pierre (Robusta) transformées par *Agrobacterium rhizogenes*. C R Acad Sci Paris **3**: 1–6

Tähtiharju S, Sangwan V, Monroy AF, Dhindsa RS, Borg M (1997) The induction of kin genes in cold-acclimating *Arabidopsis thaliana*. Evidence of a role for calcium. Planta **203**: 442–447

Takasaki H, Maruyama K, Kidokoro S, Ito Y, Fujita Y, Shinozaki K, Yamaguchi-Shinozaki K, Nakashima K (2010) The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. Mol Genet Genomics **284**: 173–183

Tong Z, Hong B, Yang Y, Li Q, Ma N, Ma C, Gao J (2009) Overexpression of two chrysanthemum *DgDREB1* group genes causing delayed flowering or dwarfism in *Arabidopsis*. Plant Mol Biol **71**: 115–129

Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive *cis*-element in the early responsive to dehydration stress 1 promoter. Plant Cell **16**: 2481–2498

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci **11**: 405–412

Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K (1993) An *Arabidopsis myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5**: 1529–1539

Van Hilten HJ (2011) The coffee exporter's guide, 3rd ed. 247p

Vieira NG, Carneiro FA, Sujii PS, Alekcevetch JC, Freire LP, Vinecky F, Elbelt S, Silva VA, DaMatta FM, Ferrão MAG, et al. (2013) Different molecular mechanisms account for drought tolerance in *Coffea canephora* var. Conilon. Trop Plant Biol **6**: 181–190

Vinecky F, Silva FR da, Andrade AC (2012) Análise *in silico* das bibliotecas de cDNA SH2 e SH3 para a identificação de genes responsivos à seca em cafeeiro. Coffee Sci **7**: 1–19

Wang CT, Yang Q, Wang CT (2011) Isolation and functional characterization of *ZmDBP2* encoding a dehydration-responsive element-binding protein in *Zea mays*. Plant Mol Biol Report **29**: 60–68

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci **10**: 88–94

Yang W, Liu X-D, Chi X-J, Wu C-A, Li Y-Z, Song L-L, Liu X-M, Wang Y-F, Wang F-W, Zhang C, et al (2011) Dwarf apple MbDREB1 enhances plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-independent pathways. Planta **233**: 219–229

Yang Y, Costa A, Leonhardt N, Siegel RS, Schroeder JI (2008) Isolation of a strong *Arabidopsis* guard cell promoter and its potential as a research tool. Plant Methods **4**: 6p

Zhang X, Fowler SG, Cheng H, Lou Y, Rhee SY, Stockinger EJ, Thomashow MF (2004) Freezing-sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing-tolerant *Arabidopsis*. Plant J **39**: 905–919

7. Conclusions and Perspectives

Collectively, the results presented in this thesis provided important insights towards the characterization of the coffee homolog gene *DREB1D*, that represents one of the candidate genes involved in drought tolerance in *Coffea sp*. putting forward several molecular data suggesting its implication on this important agronomical trait. Using the recent advances in coffee DNA genomic sequencing, it was possible to initiate an in-depth search for *DREBs* homologous genes in the *Coffea canephora* reference genome. To, our knowledge, and after the study of genes coding for *N*-methyltransferases enzymes involved in alkaloid and flavonoid pathway of caffeine synthesis that accompanied the completion of coffee genome sequencing (Denoeud et al., 2014), it is also worth noting that our work is (1) the second one realized in coffee dealing to perform an in depth analysis of a complete gene family, and (2) the first one that analyzed the structure of *DREB* genes in perennial crops.

This led us to identify 31 *DREB* subfamily members divided in four subgroups (I to IV), subgroup I was composed by five members, subgroup II by six, subgroup III by twelve and subgroup IV by eight. *Cc02g03430, formerly CcDREB1D,* showed higher phylogenetic relation to *AtDREB1D/CBF4* than to other DREB1/CBF members (*AtDREB1A/CBF3*, *AtDREB1B/CBF1*, *AtDREB1C/CBF2*), however, no clear orthology relationship between *Arabidopsis* and coffee sequences could be seen. Based on these results, it is now possible to initiate further analyses to better elucidate the specific role of these *DREB* genes particularly in relation to adaptation of coffee plants to abiotic stress. The analysis of nucleotide diversity of *DREB1D* loci in several accessions of *Coffea sp*, demonstrated (1) a high conservation of the protein coding sequence and (2) a great genetic variability of *DREB1D* promoter regions. The search for signatures of natural selection to drought adaptation evidenced a possible association between this promoter nucleotide diversity in adaptative evolutionary process of the *Coffea* species, mainly *C. canephora* and *C. arabica*. Despite prominent indication of promoter genetic diversity towards this association, it is important to mention that the implication of coding region cannot completely be ruled out, as recently reported in foxtail millet (*Setaria italica* (L.), where association of a synonymous SNP (a A/G transition) with dehydration tolerance was observed in the coding sequence

of the *SiDREB2* gene (Lata and Prasad, 2011 and 2014; Lata et al., 2011). It is noteworthy that additional work through wet lab experiments is required to confirm the existence of significant association between the observed genetic diversity of *DREB1D* and drought tolerance. . Such validation could be done by analyzing *DREB1D* gene expression, for example by quantitative polymerase chain reaction (qPCR), in leaves all *Coffea* accessions studied during this work. However, to perform such kind of experiment, it should be previously necessary (1) to share all the plants (assigning MTA and following quarantine procedures to export these materials) analyzed during the genetic study, (2) to get cuttings for all *C. canephora* accession (as this species is allogamous, mother plants cannot be propagated by seeds and (3) to grow together all these plants in controlled (greenhouse) conditions under irrigated and non-irrigated treatments. Moreover, the facts that several of these accessions originated from recent prospection (Cubry et al. 2008 and 2013) and physically dispersed in several research institutes around the world, did not allow to conduct such experiment in the framework of this thesis.

Another possibility to evaluate this association should be to analyze the genetic diversity of *DREB1D* loci in a population with the same genetic background and grown in identical (field) conditions. As part of a genome-wide selection program (GWS), our group initiated in 2012 the phenotyping for responses to drought of 403 individuals plants of *C. canephora* Conilon planted in field at experimental station of the Embrapa Cerrado research center (Planaltina - DF, Brazil) in full-sun condition. Because the dry climate of Brazilian Cerrado is characterized by well-defined dry (from May to mid-October) and a wet season concentrating more than 80 % of annual precipitations (Silva et al., 2008), it is therefore possible to evaluate drought acclimation in plants subjected to an extensive stress by measure leaf water potential (Ψ_{pd}) at the end of the dry season during. Such evaluations were performed during two consecutive years((2012 and 2013), together with estimating several other phenotypic traits like plant vigor, responses to leaf rust infection and fruit production, for example (Carneiro et al., 2014).

Altogether, these data led to the identification of drought tolerant and sensitive plants. In addition, assessment of the genetic diversity of these plants done by SSR markers, revealed a strong genetic background of SG1 subgroup in progenitors and individuals of this *C. canephora* population

(Alekcevetch, 2013), like already reported for conilon (Lambot et al., 2008). Interestingly, a SG2 genetic background also occurs within population, howsoever in much lesser extend (Alekcevetch, 2013). Such situation explains why it was possible to identify after 60 days of irrigation suspension plants of *C. canephora* highly drought^T (Ψ_{pd} > -0.5 MPa) corresponding to SG1 phenotype, and others highly drought^S (Ψ_{pd} < -3.0 MPa) corresponding to SG2 phenotype (Figure 21).

Figure 21 Characterization of *C. canephora* clones grown in field condition under water stress. On the left, drought^T clone (L5P47) presented a leaf water potential pre-dawn (Ψ_{PD}) of -0,46 MPa after 60 days of water withholding. On the right, clone (L12P100) under same conditions presented Ψ_{PD} = -2,47 MPa.

Analyses of *CcDREB1D* gene sequences and expression study of this gene in drought^T and drought^S plants of this *C. canephora* population could certainly be interesting to validate our model. Moreover, the search for association could be extended to other members of *DREB* subfamily identified by phylogenetic analyses. For this, a deeper comprehension of *DREB* gene subfamily in *Coffea* would be required. Herein, we identified and annotated new *DREB* genes in coffee genome which showed a relative small number of duplication in this subfamily. This revealed similar subgroups structuration between *Arabidopsis* and coffee subfamilies, suggesting recent and common evolutionary origins of *DREB* genes. The high level of protein sequence similarity out-side signature motifs identified may

serve to assist further functional characterization of coffee genes. Experimental validation is essential to provide reliable functional information on *DREB* genes. To achieve this objective, primers were designed for the genes composing *DREB* subfamily and should be tested to analyze their expression profiles in responses to drought stress by qPCR. RNA extracted from leaves of transgenic coffee plants generated during this study, could be used to assesses the expression of these genes in response to stress treatments such as PEG, RH, exogenous ABA and cold. This information will be essential to evidence the function of each homolog gene and start drawing the regulatory network of drought response in coffee.

To our knowledge, the functional characterization of *CcDREB1D* promoters, performed by stable genetic transformation of *C. arabica*, is the first work to analyze simultaneously homologous *DREB* promoters in a perennial crop and highlight the expression of a DREB gene in guard cells. The results presented here clearly detected GUS activity in stomatal guard cells of *CcDREB1Dpro:GUS* constructs, mainly for the full length pD14-hp16 suggesting the existence of an direct pathway to activate DREB1D gene specifically in this tissue. Such results contribute to diversify the range of functional activity reported to *DREB* genes.

Herein, a new methodology was developed to study promoter activity by estimating GUS activity frequency in guard cells. This approach should permit to know the feasibility of GUS quantification, especially when studying promoters with tissue-specificity and spatiotemporal expression. However, the comprehension of *CcDREB1D* promoter activity, primarily deduced form observed GUS activities should be validated by analyzing expression of the *GUS* reporter gene by qPCR (currently on going in the laboratory of Dr. H. Etienne), or by quantifying GUS activities by fluorimetric assays.

In addition to primers pairs designed during this work to analyze expression of *DREB* genes, the release of the *Coffea canephora* genomic sequence in the public domain also permitted to design several primer pairs for other stress-responsive genes known to function upstream (e.g. *ICE1*, *ICE2*, *CAMTA* among others) or downstream (e.g. *RD29*, *NAC*, and others) the *CcDREB1D* in the regulatory network of drought response (Figure 8 [Introduction]). In order to precise the functions of these genes in the coffee regulatory network, these primers could be used soon to check the expression profiles of these genes using RNA extracted from transgenic coffee plants subjected to PEG and RH treatments.

In this sense, phylogenetic analyses were conducted to identify coffee homolog genes outside *DREB* subfamily (such as *ICE1*, *ICE2*, *MYB15* and others) that integrate *DREB1/CBF* regulon supposed to control *DREB* gene expression, through its role of negative regulator (such as *MYB60*) of stomatal opening. New experiments dealing to characterize of the expression of p*CcDREB1D:GUS* constructs under cold acclimation, application of exogenous ABA, heat stress and circadian rhythm, should provide valuable information about *cis*-regulatory elements required to mediate stress-response of *DREB1D* promoters. Such approaches are currently on going (Luana Ferreira Torres - UFLA/UMR RPB, personal communication). Another experiment of great interest will be to repeat the drought stress experiments realized during this work with young transgenic plants in older ones and to analyze *DREB1D*promoter functioning with systematic measurements of physiological parameters such as leaf water potential (Ψ_W) and vapor pressure deficit (D), in order to provide better insight of the proposed mechanisms of guard cell feedback response and transcriptional control of *CcDREB1D* gene expression.

Without any doubts, several others approaches remain open and could constitute interesting starting points for future studies. However, based on the results obtained collectively during this work, we are confident that our strategy already provided several and consistent evidences that we are in the right direction in deciphering the molecular mechanism implicated in drought tolerance in coffee.

8. Reference

Alekcevetch JC (2013) Estudo da diversidade genética, por meio de marcadores moleculares de uma população de *Coffea canephora* var. Conilon. Master thesis, University of Lavras. p92

Cubry P, Musoli P, Legnaté H, Pot D, de Bellis F, et al. (2008) Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. Genome **51**: 50–63.

Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2013) Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol **60**: 483–501.

Carneiro FA, Rêgo ECS, Costa TS, Lima EA, Aquino SO, Rocha OC, Rodrigues GC, Carvalho MAF, Bartholo GF, Guerra AF, et al. (2014) Phenotyping and genotyping a *Coffea canephora* population, cultivated at high altitude, aiming at a GWS program for coffee. 25th International Conference on Coffee Science. Arménia, Colombia.

Lata C, Bhutty S, Bahadur RP, Majee M, Prasad M (2011) Association of an SNP in a novel DREB2-like gene SiDREB2 with stress tolerance in foxtail millet [*Setaria italica* (L.)]. J Exp Bot **62:**3387–3401.

Lata C, Prasad M (2012) Validation of an allele-specific marker associated with dehydration stress tolerance in a core set of foxtail millet accessions**.** Plant Breed **132**:496–499**.**

Lata C, Prasad M (2014) Association of an allele-specific marker with dehydration stress tolerance in foxtail millet suggests *SiDREB2* to be an important QTL. J. Plant Biochem Biotechnol **23:**119–122.

Lambot C, Crouzillat D, Fonseca AFA, Leloup V, Haiduc A, Broun P, Pétiard V, Ferrão MAG, Ferrão RG (2008) Evaluation of conilons for genetic diversity, cup quality and biochemical composition. In: Proceedings of the 22th International Scientific Colloquium on Coffee. Campinas, SP, Brazil,

Silva FAM, Assad ED, Evangelista BA (2008) Caracterização climática do bioma Cerrado. In *Cerrado: Ecologia e Flora* (Eds S.M. Sano, S.P. Almeida & J.F. Ribeiro), 71-88pp. Brasilia, Brazil: Embrapa Cerrados.

Titre: Caractérisation d'un gène candidat pour la résistance à la sécheresse chez *Coffea* : le gène *CcDREB1D* dans des génotypes contrastés de *Coffea canephora* et d'espèces apparentées.

Autour: G.S.C. Alves

Thèses présente en cotutelle à l'Ecole Nationale Supérieure d'Agronomie et l'Université de Lavras

Discipline: Biologie Intégrative

Résume

Après le pétrole, le café est la seconde matière première cotée sur les marchés financiers. Comme pour d'autres grandes cultures, la sécheresse affecte à la fois le développement de la plante, la floraison, la productivité, le développement de fruits et leur qualité. Dans ce contexte, la création de variétés de caféiers tolérantes à la sécheresse est l'une des priorités des instituts de recherche. Chez le caféier, plusieurs gènes candidats pour la tolérance à la sécheresse ont été identifiés. C'est le cas de *CcDREB1D* dont l'expression augmente à la sécheresse dans les feuilles des clones de *Coffea canephora* résistants. Les gènes *DREBs* jouent un rôle clé dans les réponses aux stress abiotiques des végétaux. Sur la base de ces informations, une analyse approfondie de *CcDREB1D* a été initiée (i) par l'étude de la diversité génétique de ce gène, (ii) en identifiant les membres des sous-familles *DREB* de caféier et en évaluant des relations phylogénétiques entre *CcDREB1D* et ses homologues chez d'autres espèces et (iii) en menant une caractérisation fonctionnelle des promoteurs *CcDREB1D* par transformation génétique chez *C. arabica*. La variabilité des promoteurs et des régions codantes *DREB1D* a été évaluée chez 38 accessions de *Coffea*, la plupart caractérisées par différents phénotypes vis-à-vis de la sécheresse. Nos résultats montrent qu'il existe des associations entre la tolérance à la sécheresse et les variations génétiques des promoteurs *DREB1D*. D'autres analyses ont montré que ces promoteurs ont évolué par réarrangements d'éléments *cis*-régulateurs qui pourraient influencer l'expression des gènes *DREB1D*. La récente publication de la séquence du génome de *C. canephora*, nous a permis d'identifier la famille de gènes *DREB* de cette espèce composée d'au moins 31 gènes de type *DREB*. Par comparaison aux autres espèces, aucune amplification spécifique de la famille des gènes *DREB* n'est observée. Un arbre phylogénétique construit avec les gènes homologues d'autres espèces a permis d'identifier les gènes de caféier des sous-groupes *DREB* I, II, III et IV, en référence à leur classification préétablie chez *Arabidopsis*. De plus, *CcDREB1D* est l'orthologue de *C. canephora* le plus proche du groupe *AtDREB1/CBF* mais, en raison d'une duplication récente menant à une sous-fonctionnalisation du groupe *DREB1/CBF* d'*Arabidopsis*, aucune inférence de fonction par homologie n'est possible. Ces résultats renforcent la nécessité de mener une caractérisation fonctionnelle des gènes *CcDREB1D* de caféier.

L'étude de la variabilité génétique a révélé une forte diversité nucléique dans les promoteurs des gènes *DREB1D* de *C. canephora*. C'est par exemple le cas pour les promoteurs des clones 14 (tolérant à la sécheresse) et 22 (sensible à la sécheresse) de *C. canephora* connus pour présenter des profils d'expression différentiels du gène *CcDREB1D* vis-à-vis de la sécheresse. Afin de réaliser l'analyse fonctionnelle de ces promoteurs, les séquences entières ou tronquées des formes alléliques et homologues des promoteurs *CcDREB1D* ont été fusionnées avec le gène rapporteur *GUS*. Les plantes transgéniques de *C. arabica* contenant ces constructions ont été régénérées. La régulation spatio-temporelle, tissulaire, cellulaire et allèle-spécifique des promoteurs *CcDREB1D* a été analysée par détection de l'activité enzymatique GUS. En conditions normales (sans stress), aucune activité GUS n'a été observée démontrant le caractère stress dépendant de l'activité de ces promoteurs. En condition de sécheresse, la version pleine longueur du promoteur *CcDREB1D* permet une activité GUS spécifiquement dans les feuilles et seulement dans l'épiderme, le parenchyme et les cellules de garde. En comparant les différentes constructions, les 700 paires de base les plus en amont des allèles et homologues des promoteurs *CcDREB1D* se révèlent essentielles pour l'expression feuille-spécifique. Par ailleurs, l'expression allèle-spécifique observée en réponse aux différents traitements (PEG et humidité relative) imitant la sécheresse, montre que les variations nucléiques des formes alléliques *CcDREB1D* affectent la régulation spatio-temporelle du gène rapporteur *GUS*.

Title: Characterization of a candidate gene for drought tolerance in *Coffea*: the *CcDREB1D* gene, in contrasting genotypes of *Coffea Canephora* and related species
Author: G.S.C. Alves
Thesis presented in cotutelle to Ecole Nationale Supérieure d'Agronomie and University of Lavras
Discipline: Integrative Biology

Abstract

Coffee is the world's second most valuable traded commodity after crude oil. Like for other crops, drought is the key factor affecting plant development, flowering, productivity, fruits development and their quality. In such a context, the generation of drought-tolerant coffee varieties has now turned into one of the priorities of many research institutes. In coffee, several candidate genes for drought tolerance have been identified. This was the case of *CcDREB1D* that showed increased gene expression upon drought stress mainly in leaves of *Coffea canephora* drought-tolerant clones. In several plant species, the *DREBs* genes play a key role in plant responses to abiotic stress. Based on these information, an in depth characterization of *CcDREB1D* was initiated by (i) studying the genetic diversity present in *DREB1D* loci, (ii) identifying *DREB* coffee subfamily members and evaluating phylogenetic relations between *CcDREB1D* and homologs from other plant species, and (iii) performing a functional characterization of *CcDREB1D* promoters by homologous genetic transformation of *C. arabica*. Sequence variability in *DREB1D* promoter and coding regions was evaluated using 38 *Coffea* accessions most of them characterized by different phenotypes (tolerance *vs.* susceptibility) regarding to drought. Our findings show several evidences of association between drought tolerance and the genetic variations on *DREB1D* promoter region. Further analyses indicated that these promoters are evolving by the rearrangement of *cis*-regulatory elements, and could influence *DREB1D* expression. The recent release of the *C. canephora* genome sequence allowed us to identify the *DREB* gene family in this species composed of at least 31 canonical *DREB* genes. Compared to other plants, no specific amplification of the *DREB* gene family was observed. A reconciled phylogenetic tree constructed with homologs from other species allowed us to identify the coffee members of the *DREB* subgroups I, II, III and IV, in reference to their classification previously established in *Arabidopsis*. In addition, *CcDREB1D* is the nearest *C. canephora* orthologue of *AtDREB1/CBF* group, but any functional inference could not be done due to a recent duplication that lead to *DREB1/CBF* group subfunctionalization in *Arabidopsis*. These results reinforced the necessity of *CcDREB1D* functional characterization in coffee.

The study of genetic variability revealed a major source of nucleic diversity in the promoter region of *DREB1D* genes from *C. canephora*, with potential implication in local adaptation to drought tolerance. This is for example the case for the promoters of drought-tolerant clone 14 and drought-sensitive clone 22 of *C. canephora* known to present different *CcDREB1D* expression profiles regarding drought. In order to perform the functional analysis of these promoters, full length and truncated version of allelic and homolog forms of *CcDREB1D* promoters were fused to the *GUS* reporter gene. Transgenic plants of *C. arabica* harboring these constructions were regenerated. The spatiotemporal, tissue, cell and allelic-specific regulation of *CcDREB1D* promoters was analyzed by detecting GUS enzymatic activity. During drought stress, the full-length *CcDREB1D* promoter drives leaf-specific GUS activity only in epidermic, parenchyma and guard-cells. By comparing the different constructs, the most upstream 700 bp of allelic and homolog *CcDREB1D* promoters were shown to be necessary for the leaf-specific expression. In addition, allelic-specific expression observed in response to different (PEG and relative humidity) treatments mimicking drought stress showed that nucleic variations present between *CcDREB1D* promoter allelic forms affect the spatiotemporal regulation of the *GUS* reporter gene.