

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
POS-GRADUAÇÃO EM BIOTECNOLOGIA

WILLIAN GERALDO DA SILVA

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS FATORES
TRANSCRICIONAIS DA FAMÍLIA AP2 / ERF ENVOLVIDOS COM OS
MECANISMOS ASSOCIADOS AO ESTRESSE HÍDRICO EM *Coffea
canephora***

PATOS DE MINAS

Fevereiro 2020

WILLIAN GERALDO DA SILVA

**IDENTIFICAÇÃO E CARACTERIZAÇÃO *IN SILICO* DOS FATORES
TRANSCRICIONAIS DA FAMÍLIA AP2/ERF ENVOLVIDOS COM OS
MECANISMOS ASSOCIADOS AO ESTRESSE HÍDRICO EM *Coffea
canephora***

Dissertação de mestrado apresentada ao
Programa de Pós-graduação em Biotecnologia
como requisito parcial para obtenção do título
de mestre.

Orientador: Prof. Dr. Matheus de Souza Gomes

PATOS DE MINAS

Fevereiro 2020

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

S586i
2020 Silva, Willian Geraldo da, 1991-
Identificação e caracterização *in silico* dos fatores transcricionais da família AP2/ERF envolvidos com os mecanismos associados ao estresse hídrico em *coffea canephora* [recurso eletrônico] / Willian Geraldo da Silva. - 2020.

Orientador: Matheus de Souza Gomes.
Dissertação (mestrado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Biotecnologia.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2021.6015>
Inclui bibliografia.
Inclui ilustrações.

1. Biotecnologia. I. Gomes, Matheus de Souza, 1981-, (Orient.). II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Biotecnologia. III. Título.

CDU: 57.08

Rejâne Maria da Silva (Bibliotecária) – CRB6/1925



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Biotecnologia
Av. Getúlio Vargas, 230, 3º andar, Sala 308 - Bairro Centro, Patos de Minas-MG,
CEP 38700-128
Telefone: (34) 3823-3714 - Ramal 39 - www.ppgbiotec.ibtec.ufu.br -
ppgbiotec@ibtec.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Biotecnologia				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	18 de fevereiro de 2020	Hora de início:	13:30	Hora de encerramento:	15:15
Matrícula do Discente:	41812BTC010				
Nome do Discente:	Willian Geraldo Silva				
Título do Trabalho:	IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS FATORES TRANSCRICIONAIS DA FAMÍLIA AP2 / ERF ENVOLVIDOS COM OS MECANISMOS ASSOCIADOS AO ESTRESSE HÍDRICO EM <i>Coffea canephora</i>				
Área de concentração:	Biotecnologia				
Linha de pesquisa:	Bioinformática e Biologia Molecular aplicada à genômica, transcriptômica e proteômica				
Projeto de Pesquisa de vinculação:	IDENTIFICAÇÃO E CARACTERIZAÇÃO DE GENES ENVOLVIDOS COM OS MECANISMOS ASSOCIADOS AO ESTRESSE HÍDRICO EM PLANTAS				

Reuniu-se na sala de videoconferência, **Palácio dos Cristais**, Campus **Patos de Minas**, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em **Biotecnologia**, assim composta: Professores Doutores: **Professora Dra. Ana Carolina S. Siquieroli - IBTEC/UFU**; **Professor Dr. Vinícius de Moraes Machado - UNIPAM**; **Professor Dr. Matheus de Souza Gomes - IBTEC/UFU** orientador do candidato.

Iniciando os trabalhos o presidente da mesa, Professor Dr. Matheus de Souza Gomes, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação

interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Matheus de Souza Gomes, Professor(a) do Magistério Superior**, em 18/02/2020, às 15:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ana Carolina Silva Siquieroli, Professor(a) do Magistério Superior**, em 18/02/2020, às 15:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **VINICIUS DE MORAIS MACHADO, Usuário Externo**, em 18/02/2020, às 15:30, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1841784** e o código CRC **CDC1643F**.

Referência: Processo nº 23117.006336/2020-26

SEI nº 1841784

WILLIAN GERALDO DA SILVA

**IDENTIFICAÇÃO E CARACTERIZAÇÃO *IN SILICO* DOS FATORES
TRANSCRICIONAIS DA FAMÍLIA AP2/ERF ENVOLVIDOS COM OS
MECANISMOS ASSOCIADOS AO ESTRESSE HÍDRICO EM *Coffea*
*canephora***

Dissertação de mestrado apresentada ao
Programa de Pós-graduação em Biotecnologia
como requisito parcial para obtenção do título
de mestre.

Aprovado em ___/___/___

BANCA EXAMINADORA

Prof. Dr. Matheus de Souza Gomes

Prof^ª. Dr^ª. Ana Carolina Silva Siquieroli

Prof. Dr. Vinícius de Moraes Machado

PATOS DE MINAS

Fevereiro 2019

AGRADECIMENTOS

A minha mãe e irmão, pelo amor, apoio e constante incentivo durante o desenvolvimento deste trabalho.

Ao orientador Dr. Matheus de Souza Gomes, por tornar possível a realização deste projeto, pelos valiosos ensinamentos, pela preocupação, paciência e confiança. Minhas dúvidas e problemas sempre foram atendidos. Só tenho a lhe agradecer pela oportunidade e pelo exemplo de pessoa e profissional!

A futura mestre a Ana Carolina Martins pela ajuda na condução do trabalho, pelos conhecimentos transmitidos e pela enorme paciência.

A Carlos Bruno de Araújo pelo auxílio na elaboração de partes fundamentais deste trabalho.

RESUMO

O café é uma valiosa exportações agrícolas mundial, sendo a segunda commodity mais comercializada. Nesse cenário, o Brasil se destaca, sendo considerado o maior produtor e exportador de café no mundo. Essa cultura representa fonte de empregos e agrega volume considerável para a receita do país, sendo indiscutível a sua importância para a conjuntura socioeconômica local e mundial. Entretanto, estresses abióticos, como a seca, são fatores que afetam significativamente a sua produção, os quais requerem a expressão de genes responsivos ao estresse, que são regulados por uma rede de fatores de transcrição. A família AP2/ERF é uma extensa família de fatores de transcrição de plantas que compartilham um domínio de ligação ao DNA bem conservado. Dessa maneira, acredita-se que a identificação dos fatores que controlam os mecanismos de defesa e adaptação são importantes para a sustentabilidade dessa cultura. Assim, o objetivo deste trabalho foi identificar e caracterizar, por meio de análises *in silico*, os fatores de transcrição da família AP2/ERF em sequências genômicas depositadas em bancos de dados públicos de *Coffea canephora*. Foram realizadas análises de espécies de vegetais para busca de proteínas referência relacionadas aos fatores de transcrição da família AP2/ERF e suas ortólogas. A busca de domínios conservados foi realizada utilizando PFAM. As sequências obtidas foram classificadas em árvore filogenética feita com o MEGA X 5.1. Foram identificadas 39 prováveis proteínas da família AP2/ERF no proteoma predito de *Coffea canephora*. Todas as proteínas identificadas no proteoma predito de *Coffea canephora* apresentaram o domínio conservado Apetala 2 (AP2). As análises demonstraram que as proteínas conservaram aminoácidos essenciais para a função da proteína tais como YRG e RAYD, que estão localizados dentro de cada domínio AP2. Este estudo foi capaz de identificar e caracterizar os prováveis genes da família AP2/ERF no genoma de *Coffea canephora*. Esta descoberta abre novas perspectivas de estudo para a verificação e a busca do envolvimento da participação destes genes na resistência do café à seca.

Palavras-chave: Análises *in silico*; *Coffea canephora*; Estresse hídrico.

ABSTRACT

Coffee is one of the most valuable agricultural products in the world and a second most traded commodity. In this scenario, Brazil stands out, being considered the largest coffee producer and exporter in the world. This culture represents a source of jobs and adds considerable volume to the country's revenue, and its importance for the local and global socioeconomic situation is indisputable. However, abiotic stresses, such as drought, are factors that significantly affect their production, which require the expression of stress-responsive genes, which are regulated by a network of transcription factors. The AP2 / ERF family is an extended family of plant transcription factors that share a well-conserved DNA binding domain. Thus, it is believed that the identification of the factors that control the defense and adaptation mechanisms are important for the sustainability of this culture. Thus, the aim of this work was to identify and characterize, through *in silico* analyzes, the transcription factors of the AP2 / ERF family in genomic sequences deposited in public databases of *Coffea canephora*. Analyzes were performed using public databases of plant species to search for reference proteins related to the transcription factors of the AP2 / ERF family and their orthologs. The search for conserved domains was performed using PFAM. The obtained sequences were classified in a phylogenetic tree made with MEGA X 5.1. 39 putative AP2 / ERF family proteins were identified in the predicted proteome of *Coffea canephora*. All proteins identified in the predicted proteome of *Coffea canephora* presented the conserved domain Apetala 2 (AP2). The analyzes showed that proteins conserved amino acids essential for protein function such as YRG and RAYD, which are located within each AP2 domain. The analyzes carried out in this work were able to identify and characterize the probable genes of the AP2 / ERF family in the genome of *Coffea canephora*. These findings open an avenue for studies for the verification and search for the involvement of the participation of these genes in the resistance of coffee to drought.

Palavras-chave: Análises *in silico*; *Coffea canephora*; Estresse hídrico.

LISTA DE ABREVIATURAS E SIGLAS

ABA: Abscisic Acid

AP2: APETALA2

AP2/ERF: APETALA2/Ethylene Response Factors

BLAST: Basic Local Alignment Search Tool

DNA: Deoxyribonucleic acid

DREB: Dehydration Responsive Element Binding Protein

ERF: Ethylene Response Factors

ESTs: Expressed Sequence Tags

MEGA: Molecular Evolutionary Genetics Analysis

NCBI: National Center for Biotechnology Information

RAP2: Related to APETALA2

RAP2.7: Related to APETALA2 7

TFs: Transcription Factors

SUMÁRIO

1	INTRODUÇÃO	7
2	REVISÃO DE LITERATURA	9
2.1	O CAFÉ	9
2.2	Estresse hídrico.....	10
	Fonte: (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2007).	
2.3	Fatores de transcrição (TFs)	11
2.3.1	Família AP2/ERF	11
2.4	BIOINFORMÁTICA	12
2.4.1	Sequenciamento de nova geração	13
2.4.2	Sequenciamento genômico <i>coffea</i>	14
2.2.4	Sequenciamento do genoma estrutural de <i>Coffea canephora</i>	15
	Referências	16
3	INTRODUCTION	23
4	MATERIAL AND METHODS	25
5	RESULTS	26
6	DISCUSSION	27
	REFERÊNCIAS	31

1 INTRODUÇÃO

A crescente demanda pela utilização dos recursos hídricos para a irrigação agrícola, para o abastecimento público e industrial, e para o aproveitamento hidrelétrico (SANTOS; STEVAUX, 2015), aliados a condições climáticas desfavoráveis, tem suscitado a ocorrência de conflitos pelo direito de uso das águas.

Perante esta utilização em demasia, a ideia de grande parte da população em relação à água é de que tal recurso é infinitamente abundante, além de sua renovação acontecer de modo natural. No entanto, a mesma ocupa 71 % da superfície do planeta, em que 97 % deste total constituem as águas salgadas, 2,07 % são as águas doces em geleiras e calotas polares (água em estado sólido) e apenas 0,63 % restam de água doce em vias de ser consumido, não totalmente aproveitada devido a questões de inviabilidade técnica, econômica e financeira, bem como de sustentabilidade ambiental (UNESCO, 2015).

Nesse contexto de distribuição hídrica, tem-se o Brasil como o país de maior abundância em água, concentrando 18 % do potencial hídrico superficial do planeta (ANA, 2010). Porém, existem consideráveis diferenças quanto à situação do recurso entre as distintas regiões do país, nas quais as problemáticas se devem, sobretudo, ao baixo rendimento de utilização, gerenciamento, contaminação e degradação ambiental.

Em relação ao setor agrícola no Brasil, a prática de irrigação das culturas plantadas é responsável por 70 % do uso de água doce (ANA, 2017). Assim, tal recurso torna-se o elemento indispensável ao desenvolvimento agrícola, de maneira que se não houver o controle e a administração adequados, além de confiáveis, não existe a possibilidade de promover uma agricultura sustentável.

Nesse setor, destaca-se o café, cultura da qual o Brasil tem sido considerado o maior produtor, assim como exportador em âmbito mundial (OIC, 2018). Desse modo, se tornou no país uma das “*commodities*” agrícolas de maior importância, sendo o quinto item agrícola mais exportado (CONAB, 2017).

Contudo, grande parte da produção cafeeira é limitada por estresses abióticos, como a seca, considerada o principal estresse que afeta a produção no Brasil (PINHEIRO et al., 2004), de forma que a mesma é responsável por uma série de respostas nas plantas, incluindo mudanças na expressão gênica, acúmulo de metabólitos e síntese de proteínas específicas.

Inúmeras vias de sinalização regulam as respostas de estresse nas plantas. Dentre estas, tem-se que as modificações para aumentar a tolerância à seca estão relacionadas à ativação e regulação de genes que protegem e mantêm as funções e componentes da estrutura celular

(WANG et al., 2009). Os produtos dos genes quando induzidos por estresse podem ser classificados em dois grupos (SHINOZAKI et al., 2003).

O primeiro grupo inclui proteínas que funcionam protegendo as células da desidratação, como as enzimas envolvidas na biossíntese de vários osmoprotetores, tais como prolina e betaínas, chaperonas e enzimas detoxificadoras. Já o segundo grupo inclui os fatores de transcrição (FTs), que desempenham um papel fundamental na resposta da planta ao estresse, por meio da ativação de genes específicos (SEKI et al., 2003; SHINOZAKI et al., 2003).

Uma das maneiras de entender como os fatores transcricionais respondem ao estresse abiótico em plantas é com a utilização de ferramentas de bioinformática. Essas ferramentas são usadas para extrair informações significativas da grande massa de dados biológicos auxiliando na resposta às perguntas biológicas (LUSCOMBE, 2001).

O sequenciamento completo do genoma de *Coffea canephora* foi realizado por um grupo de pesquisadores participantes de um consórcio internacional composto por 11 países (DENOEUDE et al., 2014). Os resultados mostraram a presença de 25.574 genes no genoma. Para analisar esse grande número de dados é necessário a utilização de ferramentas computacionais, as quais propiciam o maior entendimento de como os fatores transcricionais considerados resistentes respondem ao estresse abiótico em *C. canephora*.

Dessa forma, o objetivo deste trabalho foi identificar e caracterizar, por meio de análises *in silico*, os fatores de transcrição da família AP2/ERF envolvidos com os mecanismos associados ao estresse hídrico em sequências genômicas depositadas em bancos de dados públicos do café.

2 REVISÃO DE LITERATURA

2.1 O CAFÉ

O cafeeiro, que na antiga denominação nomeava-se cafezeiro, é uma planta nativa do continente africano, especificamente da Etiópia, que teve sua inserção por volta do ano de 1727 no Brasil, ao norte do país, através das Guianas (SAES; NAKAZONE, 2004). Já no ano seguinte, foi introduzido no Maranhão, se expandindo para os Estados vizinhos, em pequenas plantações, atingindo a Bahia em 1773. Algumas sementes de café foram então transportadas do Maranhão para o Rio de Janeiro, local em que se fomentou a ampliação da cultura na Serra do Mar (ALONSO-SALCES et al., 2009).

O café é uma das *commodities* agrícolas mais importantes no mundo. Responsável por quase metade do total das exportações de produtos tropicais, configura-se como o segundo maior no mercado mundial de produtos naturais, sendo o petróleo o primeiro (OIC, 2018).

Nesse sentido, o Brasil detém a liderança internacional em produção e exportação desse produto, sendo o maior produtor global de grãos de café por mais de 100 anos. Tem-se que as plantações de café recobrem cerca de 27.000 km² do país, em sua maioria localizadas em Minas Gerais, São Paulo, e Paraná, estados onde o clima e a temperatura são ideais para produção dessa cultura (CONAB, 2017).

Dentre as inúmeras espécies deste gênero, apenas duas têm importância econômica: *Coffea arabica* L., conhecida como “café arábica” e *Coffea canephora* Pierre, conhecida como “café robusta”. Devido à sua superioridade na qualidade de bebida, o *C. arabica* responde por mais de 70% do café produzido no mundo (CONAB, 2015).

O cafeeiro é uma planta perene, dicotiledônea de folhas persistentes e flores hermafroditas, porte arbustivo ou arbóreo e caule lenhoso. Quanto ao cafeeiro arábico, este tem altura média de 3 a 5 metros, podendo alcançar 10 m. Seu tronco tem de 8 a 10 cm de diâmetro e suas raízes podem chegar até 1,5 m de profundidade. As flores possuem fragrância de jasmim (CLARKE; MACRAE, 1989). Evolutivamente, o início da diversificação deste gênero ocorreu entre 5 e 25 milhões de anos atrás (CUBRY et al., 2008).

Espécie mais cultivada devido a sua qualidade superior, quando comparada com outras espécies ao ser cultivada em ambientes de planalto, é considerada por obter uma melhor qualidade da bebida. Tem-se que este fato parece depender da qualidade e quantidade de compostos, tais como açúcares armazenados no endosperma do fruto durante a sua maturação (GEROMEL et al., 2006; LEROY et al., 2006).

C. canephora, esta é uma espécie alógama, diplóide ($2n = 2x$) = 22 cromossomos, que apresenta autoincompatibilidade do tipo gametofítica, controlada por um único gene com vários alelos. Este mecanismo impede a ocorrência de autofecundação e de cruzamentos entre indivíduos aparentados, em consequência de interações entre proteínas presentes no pólen e no estigma, apresentando variabilidade genética significativa (SOUZA et.al., 2015).

Ainda, o *C. canephora* é mais bem adaptado às planícies equatoriais quentes e úmidas, sendo considerado mais resistente a doenças e pragas, apresentando um maior teor de cafeína do que *C. arábica*. Entretanto, a bebida preparada com seus grãos é caracterizada como neutra, de sabor leve, e eventualmente amarga (LEROY et al., 2005).

2.2 Estresse hídrico

As plantas podem ser expostas a múltiplos estresses abióticos (alteração ambientais), tais como déficit hídrico, excesso de salinidade, temperaturas extremas e uso descontrolado de defensivos agrícolas. Esses fatores limitam austeramente o crescimento e a reprodução, e conseqüentemente a produtividade agrícola, de maneira que, geralmente, as plantas respondem as mudanças ambientais de um modo complexo e integrado, por meio de mudanças morfológicas, fisiológicas, bioquímicas e moleculares (LATA et al., 2011).

Inúmeras vias de sinalização regulam as respostas de estresse nas plantas. Nesse sentido, as modificações para aumentar a tolerância à seca estão relacionadas à ativação e regulação de genes que protegem e mantêm as funções e componentes da estrutura celular (WANG et al., 2009).

Os produtos dos genes induzidos por estresse podem ser classificados em dois grupos: aqueles que protegem diretamente contra estresses e aqueles que regulam a expressão de genes e a transdução de sinais em resposta ao estresse (HASEGAWA et al., 2000; SHINOZAKI et al., 2003).

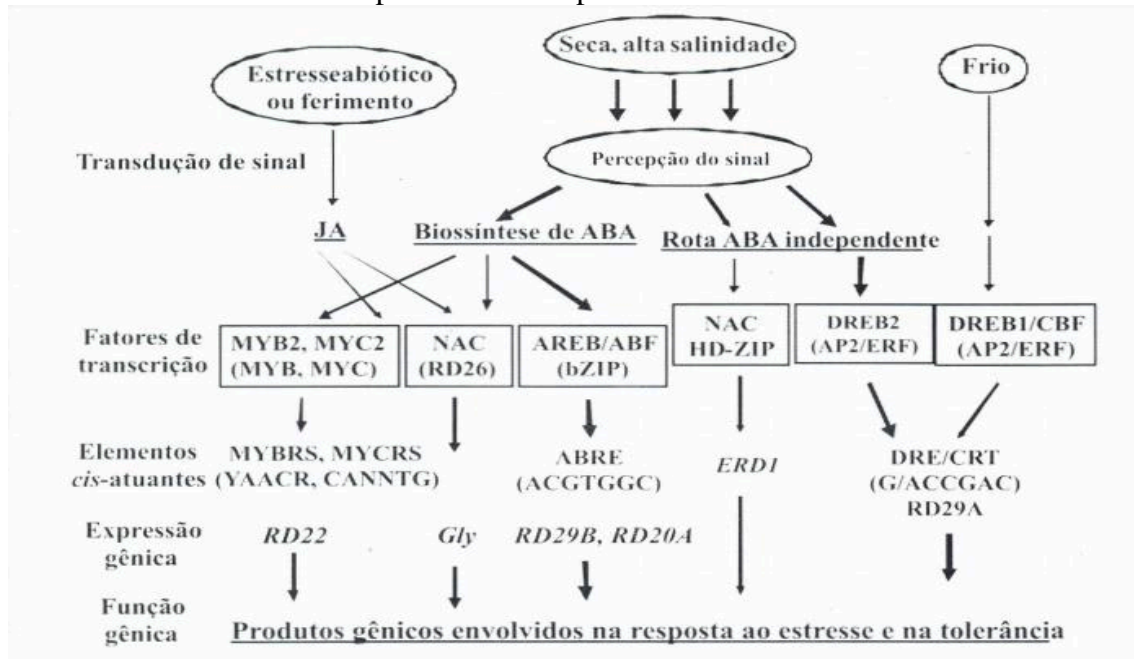
O primeiro grupo inclui proteínas que funcionam protegendo as células da desidratação, como as enzimas envolvidas na biossíntese de vários osmoprotetores, tais como prolina e betaínas, chaperonas e enzimas detoxificadoras. Já o segundo grupo inclui os fatores de transcrição (FTs), que desempenham um papel fundamental na resposta da planta ao estresse, por meio da ativação de genes específicos (SEKI et al., 2003; SHINOZAKI et al., 2003).

As respostas das plantas ao estresse abiótico podem ocorrer em vias dependente ou independente da presença de ácido abscísico (ABA) (NEPOMUCENO et al., 2001)

O ABA é um importante hormônio regulador do crescimento em plantas e desempenha papel fundamental nas respostas fisiológicas e de desenvolvimento das mesmas (LATA et al.,

2011). Tal hormônio atua na regulação de sinalização de estresse, em vias de transdução de sinal, com a capacidade de desencadear na planta uma resposta ao estresse, induzindo o fechamento dos estômatos, para que a perda de água através da transpiração seja reduzida, com a finalidade de melhorar a eficiência na utilização da água pelas plantas (LATA et al., 2011).

Figura 1 - Representação esquemática da regulação da transcrição pelos elementos de atuação em *cis*, e os fatores de transcrição envolvidos na resposta ao estresse abiótico, via dependente e independente de ABA



Fonte: (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2007).2.3 Fatores de transcrição (TFs)

Os fatores de transcrição são proteínas que se conectam ao DNA, em uma região específica do mesmo, nomeada domínio de ligação, de modo que tal ação culmina na regulação de genes, ocorrendo a ativação ou então a supressão da transcrição (PTASHNE, 1988; SALEH; PAGÉS, 2003).

De acordo com Nepomuceno et al. (2011) os TFs se encontram inseridos nas etapas iniciais da expressão e regulação de diversos grupos de genes, como resposta aos estresses ambientais.

Por controlar a expressão de vários grupos de genes, os fatores de transcrição, nos dias de hoje, são o foco de estudos, com o intuito de provocar o aumento da resistência de plantas a estresses abióticos (FARIA, 2009). Assim, entender como atuam os TFs em plantas expostas a condições de estresse hídrico, é de extrema relevância para o desenvolvimento de culturas que sejam mais tolerantes, quando submetidas a essas condições.

2.3.1 Família AP2/ERF

A família AP2/ERF é constituída por uma ampla quantidade de fatores de transcrição (SAKUMA et al., 2002) que ocorrem em múltiplas espécies de plantas monocotiledôneas e dicotiledôneas, com funções relevantes nas respostas a estresses bióticos e abióticos (SALEH; PAGÉS, 2003; SHARMA et al., 2010). Cada proteína dessa família de fatores de transcrição possui um domínio AP2/ERF extremamente conservado (SAKUMA et al., 2002), composto por aproximadamente 60 a 70 aminoácidos (JOFUKU et al., 1994; HAO et al., 1998; RIECHMANN et al., 2000; SHEN, 2013; NAKANO et al., 2006).

Os aminoácidos dos domínios AP2/ERF das subfamílias ERF, DREB, RAV e AP2 têm início pelo motivo YRG terminando com NFP (ou similar) (DAL CERO, 2010). Tem-se que a elevada especificidade de reconhecimento e de ligação ao DNA no domínio AP2/ERF deriva da presença de uma valina na posição 14 e de um ácido glutâmico na posição 19 no domínio de ligação (LATA; YADAV; PRASAD, 2011; SHEN, 2013).

Nakano et al. (2006), identificou genes da família AP2/ERF em *A. thaliana* que possuíam o domínio de ligação AP2/ERF, entretanto, as sequências desses genes exibiam baixa homologia, quando comparadas a outros genes AP2/ERF.

O domínio de ligação ao DNA dos fatores de transcrição da família AP2/ERF é demasiadamente conservado e se liga de modo específico aos elementos DRE/ CRT no promotor dos genes responsivos ao estresse (SHEN, 2013), ativando a transcrição de tais genes (AKHTAR et al., 2012).

O elemento DRE de atuação em *cis* está, sobretudo, relacionado à regulação de genes em via independente de ABA, em condições de seca, salinidade e baixas temperaturas (AKHTAR et al., 2012). Este elemento é constituído por nove nucleotídeos (TACCGACAT) (SAKUMA et al., 2002; THOMASHOW, 1999), enquanto CRT possui uma sequência de cinco nucleotídeos (CCGAC) (BAKER; WILHELM; THOMASHOW, 1994).

2.4 BIOINFORMÁTICA

A Bioinformática surgiu na década de 1980, a fim de superar as fronteiras das ciências por meio do desenvolvimento de novas abordagens, capazes de promover a análise e a apresentação de dados biológicos, bem como de garantir a pesquisa de novos métodos para a resolução de complexas perguntas (LESK, 2014).

As ferramentas de bioinformática são importantes no aperfeiçoamento e na integração de diversas áreas do conhecimento. Assim, essa capacidade tem inspirado uma grande quantidade de novos projetos e/ou pesquisas que visam o sequenciamento do genoma e a

descoberta de novas características entre as espécies (TREANGEN; SALZBERG, 2012). Diante tal fato, têm surgido diversas áreas do conhecimento, tais como a proteômica, responsável por analisar e identificar as proteínas expressas (WRIGHT et al., 2012) e a transcriptômica, que estuda a transcrição do DNA para o RNA (STUART et al., 2003).

Um dos benefícios em utilizar ferramentas de bioinformática na resolução de problemas biológicos se encontra nos bancos de dados públicos *online*, os quais oferecem informações detalhadas de moléculas ou funções específicas. O NCBI (*National Center for Biotechnology Information*), por exemplo, abriga em seu site um banco de dados de genes, genomas, RNAs, proteínas, artigos, livros, entre outros (CLARK et al., 2016).

Com os mais variados tipos de projetos de sequenciamento de genomas, geram-se também um grande número de dados. Assim, é necessária a utilização de ferramentas computacionais para análise destes dados gerando informações relevantes para o estudos dos organismos. As ferramentas de bioinformática são ferramentas usadas para extrair informações significativas da grande massa de dados biológicos auxiliando na resposta às perguntas biológicas (LUSCOMBE, 2001).

2.4.1 Sequenciamento de nova geração

O sequenciamento genético é uma série de métodos bioquímicos que têm como objetivo determinar a ordem das bases nitrogenadas presentes no material genético: adenina (A), guanina (G), citosina (C) e timina (T) (GRIFFITHS et al., 2008).

Várias tecnologias promovem o sequenciamento de segunda geração em plataformas capazes de gerar informação sobre milhões de pares de bases em uma única corrida (CARVALHO; SILVA, 2010).

Os crescentes números de sequenciadores de segunda geração causaram crescimento exponencial no número de genomas sequenciados. Isto se deve ao fato de que os novos métodos de sequenciamento aumentaram a velocidade de geração de dados em cerca de 100.000 vezes por década desde que o sequenciamento do genoma humano foi completado em 2003. Esses novos métodos são capazes de sequenciar todo o genoma humano em poucos dias (TREANGEN; SALZBERG, 2012).

Assim com o aumento da quantidade de dados, se fez necessário que as informações sequenciadas fossem gerenciadas por um Banco de dados, pois além das informações genômicas armazenadas em arquivos e bancos de dados, a manipulação, observação e a análise

desses dados, produzem novas informações que são armazenadas em formato de textos ou em novos bancos de dados (NCBI, 2015).

Existem diversos bancos de dados biológicos destacando-se, o NCBI (<http://www.ncbi.nlm.nih.gov/>), o EBI (European Bioinformatics Institute) (<http://www.ebi.ac.uk/>) e o DDBJ (*DNA Data Bank of Japan*) (<http://www.ddbj.nig.ac.jp/>).

2.4.2 Sequenciamento genômico *coffea*

O Projeto Brasileiro do Genoma Café foi desenvolvido em parceria entre CBP & D café (Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café - PBGC), FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo), e Embrapa Recursos Genéticos e Biotecnologias. Foi desenvolvido para acrescentar e implantar ferramentas úteis para a descoberta de genes e análises genéticas funcionais em espécies de café e relacionados, e para auxiliar no avanço do conhecimento sobre a estrutura e evolução do genoma do café (MONDEGO et al., 2011; VIEIRA et al., 2006).

Para isso, 214.964 clones foram aleatoriamente escolhidos de 37 bibliotecas de cDNA de *C. arabica* e *C. canephora* e então sequenciados. O sequenciamento resultou na anotação de 153.739 sequências. Os ESTs foram agrupados em 17.982 *contigs* e em 32.155 *singlets* (VIEIRA et al., 2006).

Os dados obtidos no PBGC foram armazenados em duas plataformas de bioinformática. Uma delas desenvolvida pelo Laboratório de Genômica e Expressão e está disponível desde o segundo semestre de 2004. Já a outra foi disponibilizada em março de 2007, foi criada pela Embrapa Recursos Genéticos e Biotecnologia – Cenargen (EMBRAPA, 2012).

No ano de 2010 iniciou-se o sequenciamento estrutural completo do *coffea*, o qual teve como objetivo sequenciar 700 milhões de pares de bases (Mb), utilizando-se tecnologias de sequenciamento de nova geração associadas a informações já disponíveis para o gênero, como de marcadores moleculares, bancos de sequências parciais e mapas genéticos (BERTRAND, 2010).

A diferença entre o sequenciamento completo do genoma (estrutural) para aquele obtido há 10 anos (funcional) é que o estrutural permite aos cientistas ter conhecimento sobre a ordem dos genes dentro das sequências de DNA e das regiões intergênicas que compõem o genoma, o que não é possível ver no sequenciamento funcional. É como se o sequenciamento permitisse a "radiografia" do genoma de cada planta, o que permite acelerar o melhoramento genético do

cafeeiro de características relacionadas à produtividade, à precocidade, à tolerância a estresses climáticos e à resistência a doenças, por exemplo (INFOCAFÉ, 2014; EMBRAPA, 2014).

2.2.4 Sequenciamento do genoma estrutural de *Coffea canephora*

O sequenciamento completo do genoma do café (*Coffea canephora*) foi publicado na revista Science (DENOEUDE et al., 2014). Este projeto foi fruto de um consórcio internacional composto por 11 países – Brasil, França, Itália, Canadá, Alemanha, China, Espanha, Indonésia, Austrália, Índia e Estados Unidos. Os inéditos resultados encontrados possibilitarão prever o desenvolvimento de algumas características de interesse agrônomo e acelerar o melhoramento genético do cafeeiro (INFOCAFÉ, 2014)

O sequenciamento do *C. canephora* foi realizado por Denoeud et al. (2014). Os autores sequenciaram 54,4 milhões de *reads* pelo método do *Shotgun*, com sequenciador Roche 454, acrescido de 143.605 *reads* de cromossomo bacteriano artificial, geradas por sequenciamento com Sanger dupla haplóide, o que representou uma cobertura de aproximadamente 30x o genoma de 710 Mb. Dados adicionais de sequenciamento Illumina (60 x) foram utilizados para melhorar a montagem (DENOEUDE, et al. 2014).

Os resultados foram 25.574 genes codificadores de proteínas, 92 precursores de microRNA e 2573 transferências de genomas organelares para nucleares. Os elementos transponíveis representam aproximadamente 50% do genoma, dos quais aproximadamente 85% são retrotransposons de repetição longa terminal (LTR). A comparação em larga escala entre os retrotransposons LTR de *C. canephora* e os genomas de plantas de referência mostra excelente conservação de vários grupos em genomas relacionados a distância, sugerindo que as transferências horizontais de elementos móveis podem ser mais frequentes do que o geralmente reconhecido (DENOEUDE, et al. 2014).

O sequenciamento completo do genoma de *C. canephora* resultou em uma redução significativa em tempo e custo na seleção de plantas com características de interesse. Assim, o genoma de referência juntamente com as tecnologias de sequenciamento de nova geração constitui uma ferramenta de referência para a pesquisa de genética molecular e genética de populações em espécies de café (BIANCO et al., 2016).

Referências

- ALONSO-SALCES R.M. SERRA F, RENIERO F, HÉBERGER K. Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): chemometric evaluation of phenolic and methylxanthine contents. **Journal of Agricultural and Food Chemistry**. Easton, v. 57, n. 10, p. 4224-4235, 2009. <https://doi.org/10.1021/jf8037117>
- ANA. Agência Nacional de Águas. **Atlas da irrigação. Uso da água na agricultura irrigada**. Superintendência de Planejamento de Recursos Hídricos (SPR). Brasília, DF, 2017.
- ANA. Agência Nacional de Águas. **Situação da Água no Mundo**. Brasília, DF, 2010.
- BERTRAND, B. **Decryption of the coffee genome has begun**. 2010. Disponível em: <https://www.cirad.fr/en/news/all-news-items/articles/2010/science/coffee-genome-sequencing>. Acesso em: jul. 2018.
- CARVALHO, M. C. C.G; SILVA, D.C.G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**. Santa Maria, v.40, n.3, p.735-744, 2010 <https://doi.org/10.1590/S0103-84782010000300040>
- CENCI, A.; COMBES, M. C.; LASHERMES, P. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. **Plant Molecular Biology**, Berlin, v. 78, n. 1/2, p. 135-145, 2012. <https://doi.org/10.1007/s11103-011-9852-3>
- CHALFOUN, S. M.; REIS, P. R. A história da cafeicultura no Brasil. In: REIS, P. R.; CUNHA, R. L. (Ed). **Café Arábica do plantio á colheita**. Lavras: U.R. EPAMIG SM, 896 p., 2010.
- CLARK, K.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; SAYERS, E. W. GenBank. **Nucleic Acids Res**, v. 44, n. D1, p. D67-72, Jan 4 2016. <https://doi.org/10.1093/nar/gkv1276>
- CONAB - Companhia Nacional de Abastecimento **Acompanhamento as Safra Brasileira: Café safra 2017, terceira estimativa, Abril/2017**. Disponível em: <<http://www.conab.gov.br/>> Acesso: Acesso em: jul. 2018
- CONAB – Companhia Nacional de Abastecimento. **Acompanhamento as Safra Brasileira 2015**. Disponível em: <<http://www.conab.gov.br/>> Acessado em: 20 jan 2015
- CUBRY, P; MUSOLI, P; LEGNATÉ, H. P. D; BELLIS, F; PONCET, V; ANTHONY, F; DUFOUR, M; LEROY, T. Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. **Genome**, Ottawa, v. 51, n. 1, p. 50-63, 2008. <https://doi.org/10.1139/G07-096>
- DAVIS, A. P; TOSH, J; RUCH, N; FAY, F. M. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data: implications for the size, morphology, distribution and evolutionary history of *Coffea*. **Botanical Journal of the Linnean Society**. v. 167, p. 357-377, 2011. <https://doi.org/10.1111/j.1095-8339.2011.01177.x>

DENOEUDE, F., CARRETERO-PAULET, L., DEREPPER, A., DROC, G., GUYOT, R., PIETRELLE, M., et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, n. 345, p.1181-1184, 2014. <https://doi.org/10.1126/science.1255274>

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Cenargen**. Disponível em: <https://alanine.cenargen.embrapa.br/cafEST/>. Acesso em: 17 jun. 2018.

EMBRAPA. EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA **Biotecnologia e biossegurança.**, 2014. Disponível em: < <https://www.embrapa.br/busca-de-noticias/-/noticia/2018380/genoma-completo-do-cape-e-sequenciado-pela-primeira-vez-no-mundo>>. Acesso junho de 2018.

GEROMEL, C; FERREIRA, L. P; GUERREIRO S. M; CAVALARI, A. A; POT, D; PEREIRA, L.F; LEROY, T; VIEIRA, L. G; MAZZAFERA, P; MARRACCINI, P. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. **Journal of Experimental Botany**, Oxford, v. 57, n. 12, p. 3243-3258. 2006. <https://doi.org/10.1093/jxb/erl084>

GRIFFITHS, A; WESSLER, S.R; LEWONTIN, R; GELBART, W; SUZUKI, D; MILLER, J. (2008). **Introdução à Genética**. Editora Guanabara, 9ª Edição.

HASEGAW, A. P. M; BRESSAN, R. A; ZHU, J. K; BOHNERT, H. J; Plant cellular and molecular responses to high salinity. **Annual Review of Plant Physiology and Plant Molecular**. v. 51 p. 463-499, 2000. <https://doi.org/10.1146/annurev.arplant.51.1.463>

INFOCAFÉ. Disponível em http://www.revistacafeicultura.com.br/index.php?tipo=ler&mat=55114&&utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+RevistaCafeicultura+%28Revista+Cafeicultura%29. Acesso em 29 de setembro de 2014.

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

LEROY, T. MARRACCINI, P; DUFOUR, M; MONTAGNON. C; LASHERMES. P; SABAU, X; FERREIRA, L. P; JOURDAN, I. POT, D; ANDRADE, A.C; GLASZMANN, J.C; VIEIRA, L. G; PIFFANELLI, P. Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. **Theoretical and Applied Genetics**, Berlin, v. 111, n. 6, p. 1032-1041. 2005. <https://doi.org/10.1007/s00122-005-0018-z>

LESK, A. M. **Introduction to Bioinformatics**. Oxford University Press, 2014.

LUSCOMBE, N. M., GREENBAUM, D., GERSTEIN, M; (2001). What is bioinformatics? a proposed definition and overview of the field. **Methods of information in medicine**, 40(4):346–358. <https://doi.org/10.1055/s-0038-1634431>

MONDEGO J. M; VIDAL, R. O; CARAZZOLLE, M. F; TOKUDA, E. K; PARIZZI, L. P; COSTA, G.G; PEREIRA, L.F; ANDRADE, A.C; COLOMBO, C.A; VIEIRA, L.G; PEREIRA, G. A. An EST-based analysis identifies new genes and reveals distinctive gene expression

features of *Coffea arabica* and *Coffea canephora*. **BMC Plant Biology**, London, v. 11, p. 30-34, Feb. 2011. <https://doi.org/10.1186/1471-2229-11-30>

NCBI (2015). **National center for biotechnology information**. <http://www.ncbi.nlm.nih.gov>.

NEPOMUCENO, A. L. NEUMAIER, N.; FARIAS, J. R. B.; OYA, T. Tolerância à seca em plantas: mecanismos fisiológicos e moleculares. *Biotecnologia, Ciência e Desenvolvimento*, v. 4, n. 23, p. 12-18, 2001.

ORGANIZAÇÃO INTERNACIONAL DO CAFÉ (OIC). Disponível em: <http://consorcioesquisacafe.com.br/arquivos/consorcio/publicacoes_tecnicas/Relatorio_sobre_o_mercado_de_cafe_Novembro_2018.pdf>. Acesso em abril 2018.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (UNESCO), **Relatório Mundial das Nações Unidas sobre Desenvolvimento dos Recursos Hídricos: Água para um mundo Sustentável**, 2015. Disponível em: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/SC/images/WWDR2015ExecutiveSummary_POR_web.pdf. Acesso em maio. 2018.

PINHEIRO, P. A; MATTA, F. M; CHAVES, A. R. M; FONTES, E. P. B. E; LOUREIRO, M. E. Drought tolerance in relation to protection against oxidative stress in clones of *Coffea canephora* subjected to long-term drought. **Plant Science**, v. 167, p. 1307-1314, 2004 <https://doi.org/10.1016/j.plantsci.2004.06.027>

SAES, M. S. M; NAKAZONE, D. O agronegócio café do Brasil no mercado internacional. **Fae Business**. N, 9, 2004.

SANTOS, V. C; STEVAUX, J. C. Estimativa e regionalização hidrológica de vazões médias e produção de sólidos suspensos para a bacia hidrográfica do Rio Ivaí – Estado do Paraná. **Geographia Meridionalis**, Pelotas. v.1, n.02. p.384–404, 2015. <https://doi.org/10.15210/gm.v1i2.6879>

SEKI, M.; KAMEI, A.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Molecular responses to drought, salinity and frost: common and different paths for plant protection. **Current Opinion of Biotechnology**. v. 14, p.194-199, 2003. [https://doi.org/10.1016/S0958-1669\(03\)00030-2](https://doi.org/10.1016/S0958-1669(03)00030-2)

SHINOZAKI, K. YAMAGUCHI-SHINOZAKI, K. Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58:221-227, (2007) <https://doi.org/10.1093/jxb/erl164>

SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K.; SEKI, M. Regulatory network of gene expression in the drought and cold stress responses. **Current Opinion in Plant Biology**, v. 6, p. 410-417, out. 2003. [https://doi.org/10.1016/S1369-5266\(03\)00092-X](https://doi.org/10.1016/S1369-5266(03)00092-X)

SOUZA, F. F; FERRÃO, L. F. V; CAIXETA, E. T; SAKIYAMA, N. S; PEREIRA, A. A; OLIVEIRA, A. C. B. Aspectos gerais da biologia e da diversidade genética de *Coffea canephora*. (Ed.). **Café na Amazônia**. Brasília, DF: Embrapa, 2015. p.85-95

STUART, J. M.; SEGAL, E.; KOLLER, D.; KIM, S. K. A gene-coexpression network for global discovery of conserved genetic modules. **Science**, v. 302, n. 5643, p. 249-55, 2003. <https://doi.org/10.1126/science.1087447>

TREANGEN, T. J. AND SALZBERG, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. **Nature Reviews Genetics**, 13(1):36–46. (2012). <https://doi.org/10.1038/nrg3117>

VIEIRA, L. G. E. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, Piracicaba, v. 18, n. 1, p. 95-108, Mar. 2006.

WANG, C; Yin-Mao Dong. Overexpression of maize ZmDBP3 enhances tolerance to drought and cold stress in transgenic Arabidopsis plants. **Biologia, Bratislava** v. 64, n. 6, p. 1108-1114, 2009. <https://doi.org/10.2478/s11756-009-0198-0>

WRIGHT, P. C., NOIREL, J., OW, S. Y, FAZELI, A. A review of current proteomics technologies with a survey on their widespread use in reproductive biology investigations. **Theriogenology**, v. 77, n. 4, p. 738- 765, 2012. <https://doi.org/10.1016/j.theriogenology.2011.11.012>

ARTIGO CIENTÍFICO

Titulo: Characterization of AP2/ERF family transcription factors involved with mechanisms associated to water-stress in *Coffea canephora*

Periodico: Canadian Science Publishing: Genome

14 **ABSTRACT**

15

16 Coffee is one of the most valuable agricultural exports in the world and the second most traded
17 commodity. In this scenario, Brazil stands out, being considered the largest coffee producer and
18 exporter in the whole world. However, abiotic stresses such as drought, significantly affect its
19 production, which requires the expression of stress-responsive genes, regulated by a network
20 of transcription factors. The AP2 / ERF family plays an important role in the development and
21 expression of genes that regulates the defense response in plants. Thus, the aim of this work
22 was to identify and characterize, through *in silico* analysis, the transcription factors of the AP2
23 / ERF family in the genome of *Coffea canephora*. Analyses using public database of plant
24 species were performed identifying and characterizing 39 putative proteins similar to the the
25 AP2 / ERF family in *Coffea canephora* genome. All putative proteins identified in the predicted
26 *Coffea canephora* proteome had the Apetala 2 (AP2) domain conserved in their structures.
27 These putative AP2 / ERF proteins found in *Coffea canephora* showed conserved amino acids
28 mainly in YRG- and RAYD-motifs that have been located within AP2 domain. Phylogenetic
29 distribution of the *C. canephora* proteins showed high similarity when compared to orthologous
30 proteins from each clade. Thus, the identification and characterization of the putative AP2 /
31 ERF family proteins in coffee genome may help to develop an interesting approach in
32 improvement strategies for drought resistance in *C. canephora*.

33 .

34

35

36 **Keywords:** water stress, *coffee canephora*, transcription factors, *in silico* analysis.

37

38 3 INTRODUCTION

39 Coffee is one of the most important agricultural commodities in the world (DAVIS
40 et al. 2011) and the main livelihood for over 125 million people, being produced in over 60
41 countries. Furthermore, coffee plants are responsible for almost half of total tropical product
42 exports, being the second largest in the world market of natural products, after oil. In this
43 scenario, Brazil has been considered the worldwide largest coffee producer and exporter (IOC,
44 2018). Thereby, coffee has become one of the most important agricultural commodities in the
45 country (CONAB, 2017). Coffee belongs to the genus *Coffea* and has two species of greater
46 economic importance, *Coffea arabica* and *Coffea canephora* (IOC, 2018).

47 However, much of coffee production is limited by abiotic stresses, such as drought,
48 considered the main stress that affects production in Brazil (PINHEIRO et al. 2005). Drought
49 is responsible for a variety of plant responses, including changes in gene expression, metabolite
50 accumulation and specific protein synthesis (SOUZA et al. 2015). *C. canephora* has wide
51 genetic variability and greater resistance to pests and drought tolerance (SOUZA et al., 2004).

52 Innumerable signaling pathways regulate stress responses in plants. Thus,
53 modifications to increase drought tolerance are related to activation and regulation of genes that
54 protect and maintain cellular structure functions and components (WANG et al. 2009). Stress-
55 induced gene products can be classified into two groups (SHINOZAKI et al. 2003).

56 The first group includes proteins that function to protect cells from dehydration,
57 such as enzymes involved in the biosynthesis of several osmoprotectors, such as prolines,
58 betaines, chaperones, and detoxifying enzymes. The second group includes transcription factors
59 (FT), which play a key role in plant stress response through activation of specific genes (SEKI
60 et al. 2003; SHINOZAKI et al. 2003).

61 Transcription factors are proteins that attach to DNA in a specific region known as
62 a binding domain (SALEH & PAGÉS, 2003). This connection results in gene regulation leading

63 to activation or suppression of transcription. TF are present in early stages of expression and
64 regulation of several gene groups, whether they are involved in response to environmental
65 stresses or not (NEPOMUCENO et al. 2001).

66 The AP2 / ERF family is one of the largest families of transcriptional plant factors.
67 This family plays some key roles, not only in the response of plants to various stresses, but also
68 in plant growth and development (HUANG et al. 2016).

69 Up to date several studies were conducted focusing on the identification of family
70 AP2 / ERF in plant species, such as: cabbage (*Brassica oleracea*) (THAMILARASAN et al.
71 2014), eucalyptus (*Eucalyptus grandis*) (CAO et al. 2015), potato (*Solanum tuberosum*)
72 (CHARFEDDINE et al. 2015), grape (*Vitis vinifera*) (LICAUSI et al. 2010) and *Arabidopsis*
73 *thaliana* (NAKANO et al. 2006).

74 One of the ways to understand how transcriptional factors respond to abiotic stress in
75 plants is with the use of bioinformatics tools. These are used to extract significant information
76 from the large mass of biological data to assist in answering biological questions (LUSCOMBE,
77 2001).

78 The complete sequencing of the *C. canephora* genome was performed by an
79 international consortium composed by researchers from 11 countries (DENOEUDE et al., 2014).
80 The results showed the presence of 25.574 genes in the genome. To analyze this large number
81 of data, it is necessary to use computational tools, which will provide an understanding of how
82 the transcriptional factors considered resistant respond to abiotic stress in *C. canephora*.

83 Thus, the aim of this paper was to identify and characterize the transcription factors
84 of the AP2 / ERF family involved with the mechanisms associated with the response to water
85 stress in public databases of *C. canephora*, using *in silico* analyses.

86 4 MATERIAL AND METHODS

87 The putative AP2 / ERF family transcription factors were identified and selected by
88 mining *C. canephora* protein sequences from public *coffee* database such as CAFEST
89 (<http://bioinfo04.ibi.unicamp.br>), *Coffee Genome Hub* (<http://coffee-genome.org/>) and NCBI
90 (*National Center for Biotechnology Information*) using BLASTp and plant reference proteins
91 as queries [24]. Blastp *C. canephora* hit showing cutoff values $< e^{-10}$ were selected. Reference
92 AP2 / ERF family transcription factor proteins from those considered model organisms such as
93 *A. thaliana* and other plant species such as *Camellia sinensis*, *Hevea brasiliensis*, *Vitis vinifera*,
94 *Manihot esculenta*, *Lactuca sativa*, *Nicotiana tabacum* and *C. arabica*, were searched in the
95 NCBI (National Center for Biotechnology Information) using BLASTp tool to obtain a full set
96 of putative orthologue proteins to compare with the *C. canephora* putative proteins. Analyses
97 of protein family domains and distribution of functionally active amino acids in the putative
98 proteins were performed using the PFAM (version 32.0 <http://pfam.sanger.ac.uk>) and
99 Conserved Domains Database (CDD) ([http:// www.ncbi.nlm.nih.gov/cdd/](http://www.ncbi.nlm.nih.gov/cdd/)). The putative
100 protein sequences were used to perform multiple sequence alignments - MSA using
101 CLUSTALX 2.1 with the default settings (Larkin et al. 2007). A phylogenetic tree was used to
102 reconstructs phylogenies using the neighbour-joining method by iteratively joining pairs of
103 nodes method and the Jones-Taylor-Thornton model (Tamura et al. 2013). A bootstrap
104 consensus tree with 2,000 replicates was used to represent the evolutionary history of the taxa
105 analyzed. The molecular phylogenetic analyses were conducted using MEGA X 5.2 software
106 (Tamura et al. 2011). All positions containing gaps and missing data were eliminated from the
107 dataset. The protein domain logos were generated using WebLogo – 2.8.2
108 (<http://weblogo.berkeley.edu/logo.cgi>).

109 5 RESULTS

110 39 putative *C. canephora* transcription factors that encode AP2/ERF proteins which
111 contain a complete AP2 domain, were identified by mining available *C. canephora* databases
112 and Blastp tool.

113 The predicted proteins identified in the *C. canephora* database and their orthologue
114 proteins from plant species, as well as their coding transcript ID followed by their respective
115 sizes (in amino acids), the Conserved Domains and Blastp e-value were reported in Table 1.

116 For characterization of the predicted AP2/ERF proteins, sequences of putative
117 orthologous of *C. canephora* proteins were obtained from NCBI database . The amino acid
118 sequences of *C. canephora* AP2/ERF putative proteins were used as a query in the analysis
119 against the NCBI RefSeq database using Blastp tool selecting the best hits. The fasta sequence
120 of those proteins were retrieved and used to perform conserved domain analysis, phylogeny
121 analysis and global alignment analysis.

122 The analyses revealed that *C. canephora* AP2/ERF putative proteins are conserved at
123 the amino acid level compared to their orthologue AP2/ERF proteins from diverse plant species
124 such as *A. thaliana*, *Camellia sinensis*, *Hevea brasiliensis*, *Vitis vinifera*, *Manihot esculenta*,
125 *Lactuca sativa*, *Nicotiana tabacum* and *Coffea arabica*.

126 *C. canephora* AP2/ERF putative proteins displayed a highly conserved AP2 domain
127 with approximately 60-70 amino acid residues in their consensus sequences, as shown in Figure
128 1.

129 *C. canephora* AP2/ERF putative proteins showed 2 conserved sequence groups within
130 each AP2 domain. The first group named as YRG element (Figure 2) and the second group
131 RAYD element (Figure 2). Also, in this context, several invariant amino acid residues in the
132 YRG and RAYD elements had been found and may also play a critical role in the structure or
133 function of these proteins.

134 Phylogenetic analysis was performed to evaluate the evolutionary relationship of the
135 putative proteins found in *C. canephora* against their orthologue proteins Phylogenetic tree
136 (Figure 3) displayed 32 distinct clades of the transcription factor AP2/ERF proteins. This
137 inferred that the putative proteins found in *C. canephora*, probably refer to sequences of
138 transcriptional factors involved in the plant response to water stress.

139

140 **6 DISCUSSION**

141 39 putative *C. canephora* transcription factors that encode AP2/ERF proteins which
142 contain a complete AP2 domain, were identified by mining available *C. canephora* databases
143 and Blastp tool. (Table 1). It is composed of a large number of transcription factors (SAKUMA
144 et al., 2002) that are present in many species of monocotyledonous and dicotyledonous plants
145 (Saleh and Pagés, 2003).

146 Expression analyzes of the AP2 / ERF family genes in tomato (Sharma et al 2010), rice
147 (Sharoni et al. 2011), soybean (Zhang et al. 2008) and pepper (*Capsicum annuum L.*) JIN et al.,
148 2018, reveal that many genes in this family are induced by high or low temperature or
149 dehydration.

150 The protein sizes of *C. canephora* AP2/ERF putative proteins corroborated to the
151 orthologue protein sizes in model organisms, such as *A. thaliana* (Table 1). For example, the
152 putative ERF020 protein Cc10_g09120 showed 137 amino acids and its orthologue, found in
153 *A. thaliana* (NP_177307.1), presented 143 amino acid residues. The putative ERF011 protein
154 Cc07_g06220 showed 164 amino acids and its orthologous, found in *A. thaliana*
155 (NP_190595.1), showed 153 amino acid residues. The putative ERF098 protein Cc01_g18400
156 showed 155 amino acids and its orthologous, found in *A. thaliana* (NP_188964.2), showed 139
157 amino acid residues. The consistency between the sizes of *C. canephora* proteins and their
158 orthologous is an indication of the high conservation of this pathway in plant organisms.

159 Some of the ERFs have acted as transcription activators such as ERF98 and ERF011
160 (ÇEVIK et al., 2012). Transcriptional activators are DNA-binding proteins that interact with
161 the Mediator complex and subsequently recruit RNA polymerase II (RNAP II) and other
162 components to form the pre-initiation complex (PIC) in the promoter region of targeted genes
163 (CONAWAY & CONAWAY, 2011). In this work we found both ERFs considered as
164 transcription activators, ERF98 and ERF011, in the *C. canephora* database.

165 The ERF8 found in *C. canephora* (Cc01_g18710) showed 184 amino acids,
166 corroborating the sizes of the orthologue proteins from *Prunus persica* and *Nicotiana attenuate*.
167 The *C. canephora* ERF8 showed the same conserved domains (AP2) compared to orthologue
168 proteins. Proteins within a subgroup that share these conserved domain are likely to have similar
169 functions. ERF8 is a part of the complex network of protein-protein interactions that are
170 transcriptionally regulated by ABA (LUMBA et al., 2017). ABA is closely related to tolerance
171 to several stress conditions. Many genes expressed during different abiotic stress conditions
172 depend on the presence of ABA (TAIZ; ZEIGER, 2004). ERF8 was also recently identified as
173 part of an ERF transcription factor network, along with ERF98, which controls osmotic stress
174 (VAN DEN BROECK et al., 2017).

175 All proteins in this study identified in the predicted proteome of *C. canephora* presented
176 the conserved domain Apetala 2 (AP2) (Figure 1). The conserved Apetala 2 (AP2) motif was
177 found in all 19 proteins studied from *Brassica napus L* (Colza) (Wang et al 2019). This highly
178 conserved domain have shown in its consensus sequence approximately 60-70 amino acid
179 residues, which confer a typical three-dimensional conformation organized in a layer of three
180 antiparallel beta sheets followed by a parallel alpha helix (Dinh et al. 2012; (Allen et al., 1998).

181 The AP2 family is one of the largest families of transcriptional plant factors (Riechmann
182 and Meyerowitz, 1998). They play a role in regulating transcription, which involves complex

183 growth processes, biological stresses, seed germination, flower growth, aging, fruit arrival, salt
184 response, drought, low temperature, and pathogen attack (Aukerman et al., 2003).

185 The binding of transcription factors occurs by the interaction of the promoter region of
186 the target genes with the proteins responsible for the activation or repression of gene expression.
187 This connection happens in a specific way to each stimulus and also due to the conserved AP2
188 /ERF domain (LI et al., 2018). With the interaction of transcription factors AP2 / ERF and gene
189 promoters, mechanisms that lead to adaptation of plants to stress are activated, enabling them
190 to survive to abiotic stress conditions (Xie et al., 2019; Okamuro et al., 1997).

191 Two protein motifs were observed in the AP2 domain of *C. canephora* ERFs, YRG and
192 RAYD, corroborating with literature. These two motifs are critical in activating DNA binding
193 to modulate the expression of target genes and protein-protein interaction (Okamuro et al.
194 (1997). Weblogo showed the level of conservation of each amino acid residues within both
195 motifs. YRG and RAYD motifs were highly conserved highlighting the glycine residue (G) in YRG
196 showed with 100% of identity among all *C. canephora* proteins and the residue aspartic acid
197 (D) in RAYD motif.

198 The YRG motive (Figure 2) is about 19 and 22 amino acids, is highly basic and contains
199 the conserved amino acid motive located at the N-terminal portion. The RAYD site (Figure 2)
200 is about 42 and 43 amino acids long and contains a highly conserved 18 amino acid central
201 region, located in the C-terminal portion, which forms an amphipathic α -helix in the AP2
202 domains (Kizis, 2001; Okamuro et al., 1997).

203 The amino acids of the AP2 / ERF domains begin with the YRG motif ending with NFP.
204 The high recognition and binding specificity for DNA in the AP2 / ERF domain is due to the
205 presence of a valine (V) at position 14 and a glutamic acid (E) at position 19 in the connection
206 domain (Lata et al., 2011). In the same way, there are several invariant amino acid residues in

207 the YRG and RAYD elements that may also play a critical role in the structure or function of
208 these proteins.

209 All the studied proteins presented a conserved glycine residue (G), in the active sites
210 YRG. These residues within the YRG element are invariable throughout the AP2 domain and
211 have shown to be important for the AP2 function. In the AP2 domain, two amino acids, a 14
212 valine (V) and a 19 glutamate (E), play a key role in determining DNA binding specificities
213 (Sakuma et al., 2002).

214 The phylogenetic distribution (Figure 3) of *C. canephora* AP2/ERF putative proteins
215 showed these proteins clustering with their orthologue proteins, due to their evolutionary
216 proximity. The analysis demonstrated that all ERF groups can be classified based on AP2
217 domains, corroborating with Sakuma et al. (2002).

218 AP2 / ERF protein members were classified into three main families (AP2, ERF and
219 RAV) based on the number of AP2 domains present (Nakano et al., 2006). The AP2 family
220 proteins contain two repeated AP2 / ERF domains. RAV family proteins, besides to a single
221 AP2 / ERF domain, contains a B3 domain, which is a DNA binding domain conserved in other
222 TFs, and ERF family proteins contain a single AP2 / ERF domain (Sakuma et al., 2002; Nakano
223 et al., 2006).

224 The phylogenetic analysis of *C. canephora* AP2/ERF putative proteins and their
225 orthologous allowed the distribution in the three different classes of AP2 / ERF proteins
226 corroborating as previously described by Nakano et al., 2006.

227 Transcription factors represent ideal targets for plant breeding, whether traditional or
228 assisted by genetic engineering, with specific characteristics related to stress tolerance or higher
229 yield. In particular, ERF genes are among the most interesting FTs, as they have been selected
230 through evolution to regulate a number of stress response pathways (Licausi et al. 2013).

231 In conclusion, 39 AP2 / ERF genes were identified in *C. canephora* genome using public
 232 plant databases. All proteins identified in the predicted proteome of *C. canephora* presented the
 233 conserved domain Apetala 2 (AP2). The results obtained demonstrated that the ERF proteins
 234 from *C. canephora* were conserved at the amino acid level mainly in the AP2 domain and the
 235 active sites YRG and RAYD. Therefore, identifying and characterizing the drought resistance
 236 genes in *C. canephora* contributes to improving the understanding of the AP2 / ERF family
 237 genes, which despite being composed of a large number of transcription factors, few are still
 238 studied. The identification of these genes, can contribute to the development of crops more
 239 tolerant to drought.

240

241 REFERÊNCIAS

242

243 AUKERMAN M.J., SAKAI, H: 2003. Regulation of flowering time and floral organ identity
 244 by a microRNA and its APETALA2-like target genes. *Plant Cell*.
 245 doi.org/10.1105/tpc.016238.

246

247 CAO, P.B., AZAR, S., SAN, C.H., MOUNET, F., DUNAND, C., MARCA, G., MARCA, C;
 248 TEULIÈRES, C. 2015. Genome-Wide Analysis of the AP2/ERF Family in *Eucalyptus*
 249 *grandis*: An Intriguing Over-Representation of Stress-Responsive DREB1/CBF Genes. *Plos*
 250 *One*, v. 10, n. 4, p. 1-22. DOI: 10.1371 / journal.pone.0121041.

251

252 CHARFEDDINE, M., SAIDI, M.N., CHARFEDDINE, S., HAMMAMI, A., GARGOURI.
 253 BOUZID, R. 2015. Genome-wide analysis and expression profiling of the ERF transcription
 254 factor family in potato (*Solanum tuberosum* L.). *Molecular Biotechnology*. P. 348-358. DOI:
 255 10.1007 / s12033-014-9828-z.

256

257 CONAB – Companhia Nacional de Abastecimento. Acompanhamento as Safra Brasileira
 258 2015. Disponível em: <<http://www.conab.gov.br/>> Acessado em: 20 abr 2018.

259

260 CONAWAY, RC., CONAWAY, J.W. Function and regulation of the Mediator complex
 261 *Current Opinion in Genetics & Development*, 21(2), 225–230.doi:10.1016/j.gde.2011.01.013

262

263 ÇEVIK, V., KIDD, B. N., ZHANG, P., HILL, C., KIDDLE, S., DENBY, K. J., KAZAN, K.
 264 HOLUB, E.B., CAHILL, D.M., MANNERS, J.M., SCHENK, P.M. BEYNON, J (2012).
 265 MEDIATOR25 Acts as an Integrative Hub for the Regulation of Jasmonate-Responsive Gene

- 266 Expression in Arabidopsis. *Plant Physiology*, 160(1), 541–555. DOI:
267 <https://doi.org/10.1104/pp.112.202697>
268
- 269 DAVIS, A., P; TOSH, J., RUCH, N., FAY, M.F. 2011. Growing coffee: *Psilanthus*
270 (*Rubiaceae*) subsumed on the basis of molecular and morphological data; implications for the
271 size, morphology, distribution and evolutionary history of *Coffea*. *Bot J Linn Soc* 167:357–
272 377. doi.org/10.1111/j.1095-8339.2011.01177.x.
273
- 274 DINH. T.T., GIRKE, T., LIU, X. 2012. The floral homeotic protein APETALA2 recognizes
275 and acts through an AT-rich sequence element. *Development*; 1978-1986. DOI:
276 10.1242/dev.077073.
277
- 278 FINN, R.D., BATEMAN, A., CLEMENTS, J., COGGILL, P., EBERHARDT, R.Y., EDDY,
279 S.R., HEGER, A., HETHERINGTON, K., HOLM, L., MISTRY, J., SONNHAMMER,
280 E.L.L., TATE, J., PUNTA, M. 2014 Pfam, the protein families database. *Nucleic Acids*
281 *Research*. 222 – 230. DOI: 10.1093/nar/gkt1223.
282 .
- 283 HUANG, P.Y., CATINOT, J., AND ZIMMERLI, L. 2016. Ethylene response factors in
284 Arabidopsis immunity. *Journal of Experimental Botany*.1231–1241. DOI: 10.1093 / jxb /
285 erv518.
286
- 287 INTERNATIONAL COFFEE ORGANIZATION. Total production by all exporting
288 **countries**. ICO. 2018. Disponível em: <<http://www.ico.org/prices/po-production.pdf>>.
289 Acesso maio de 2018.
290
- 291 JIN, J.-H., WANG, M., ZHANG, H.-X., KHAN, A., WEI, A.-M., LUO, D.-X. E GONG, Z.-
292 H. 2018. Genome-wide identification of the AP2/ERF transcription factor family in pepper
293 (*Capsicum annuum* L.). *Genoma*. 663-674, DOI: 10.1139 / gen-2018-0036.
294
- 295 LARKIN, M.A., BLACKSHIELDS, G., BROWN N.P., CHENNA, R., MCGETTIGAN,
296 P.A., WILLIAM, H., VALENTIN, F., WALLACE, I.M., WILM, A., LOPEZ, R.,
297 THOMPSON, J.D., GIBSON, T.J., HIGGINS D. G. 2007. Clustal W e Clustal X versão 2.0.
298 *Bioinformatics*. 2947–2948. DOI: 10.1093 / bioinformática / btm404.
299
- 300 LATA, C.; YADAV, A.; PRASAD, M. 2011. Role of Plant Transcription Factors in Abiotic
301 Stress Tolerance. In *Tech*. 269-296. DOI: 10.5772 / 23172.
302
- 303 LI, X., TAO, S., WEI, S., MING, M. HUANG, S., ZHANG, S., WU, J. 2018. The mining and
304 evolutionary investigation of AP2/ERF genes in pear (*Pyrus*). *BMC Plant Biology*. 18, 46.
305 doi: 10.1186 / s12870-018-1265-x
306

- 307 LICAUSI, F., GIORGI, F.M., ZENONI, S., OSTI, F., PEZZOTTI, M., PERATA, P. 2010.
308 Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. BMC
309 Genomics. 1 - 15. doi.org/10.1186/1471-2164-11-719.
310
- 311 LUMBA S, TOH S, HANDFIELD LF, SWAN M, LIU R, YOUN J.Y. CUTLER, S.R.,
312 SUBRAMANIAM R., PROVART, N., MOSES, A., DESVEAUX, D., MCCOURT, P. A
313 mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in
314 Arabidopsis. Dev Cell. 2014; 29: 360–72. DOI: 10.1016/j.devcel.2014.04.004.
315
- 316 KIZIS, D.; PAGES, M. 2002. Maize DRE-binding proteins DBF1 and DBF2 are involved in
317 rab17 regulation through the drought-responsive element in an ABA-dependent pathway. The
318 Plant Journal. 679-689. DOI: 10.1046 / j.1365-313x.2002.01325.x.
319
- 320 NCBI (2015). National center for biotechnology information. <http://www.ncbi.nlm.nih.gov>.
321 jun 2018.
322
- 323 NAKANO, T., SUZUKI, K., FUJIMURA, T., AND SHINSHI, H. 2006. Genome-wide
324 analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol. 411-432. DOI:
325 <https://doi.org/10.1104/pp.105.073783>.
326
- 327 NEPOMUCENO, A.L., NEUMAIER, N., FARIAS, J.R.B., OYA, T. 2001. Tolerância à seca
328 em plantas: mecanismos fisiológicos e moleculares. Biotecnologia, Ciência e
329 Desenvolvimento. 12-18. doi.org/10.1046/j.1365-313X.2002.01325.x.
330
- 331 PINHEIRO, P.A., MATTA, F.M., CHAVES, A.R.M., FONTES, E.P.B.E., LOUREIRO, M.
332 E. 2004. Drought tolerance in relation to protection against oxidative stress in clones of
333 *Coffea canephora* subjected to long-term drought. **Plant Science**, 167, 1307-1314.
334 doi.org/10.1016/j.plantsci.2004.06.027.
335
- 336 OKAMURO, J.K., CASTER B., VILLARROEL R., VAN, MONTAGU, M., JOFUKU K.D:
337 1997. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins
338 in Arabidopsis. Proc Natl Acad Sci U S A., 94: 7076-7081. 10.1073/pnas.94.13.7076.
339
- 340 RIECHMANN, J.L., MEYEROWITZ, E.M. 1998 The AP2/EREBP family of plant
341 transcription factors. Biol Chem. 379: 633-646. DOI: 10.1515 / bchm.1998.379.6.633.
342
- 343 SAKUMA, Y., LIU, Q., DUBOUZET, J.G., ABE, H., SHINOZAKI, K., AND
344 YAMAGUCHI-SHINOZAKI, K. 2002. DNA-binding specificity of the ERF/AP2 domain of
345 Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene
346 expression. Biochemical and Biophysical Research Communications 290. 998-1009. doi:DOI
347 10.1006/bbrc.2001.6299.
348

- 349 SALEH, A.; PAGÉS. M. 2003. Plant AP2/ERF transcription factors. *Genetika*, v. 35, n. 1. 37-
350 50. <https://doi.org/10.3389/fpls.2019.00228>.
351
- 352 SEKI, M.; KAMEI, A.; YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. 2003. Molecular
353 responses to drought, salinity and frost: common and different paths for plant protection.
354 *Current Opinion of Biotechnology*. 14.194-199. DOI: 10.1016 / s0958-1669 (03) 00030-2.
355
- 356 SHARONI, A.M., NURUZZAMAN, M., SATOH, K., SHIMIZU, T., KONDOH, H.,
357 SASAYA, T., CHOI, I.R., OMURA, T., AND KIKUCHI, S. 2011. Gene structures,
358 classification and expression models of the AP2/EREBP transcription factor family in rice.
359 *Plant Cell Physiol* 52(2): 344-360. doi:10.1093/pcp/pcq196.
360
- 361 SHARMA, R. KUMAR, A. SOLANKE, R. SHARMA, A. TYAGI, A. SHARMA, (2010)
362 Identification, phylogeny, and transcript profiling of ERF family genes during development and
363 abiotic stress treatments in tomato, *Mol. Genet. Genomics* 284 455–475
364 <https://doi.org/10.1007/s00438-010-0580-1>
365
- 366 SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K.; SEKI, M. 2003. Regulatory network of
367 gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology*. 6.
368 410-417. DOI: 101016 / s1369-5266 (03) 00092-x.
369
- 370 SILVA, C A; ANTÔNIO M. C; COELHO, G; REZENDE, F C; SATO, F. 2008.
371 Produtividade e potencial hídrico foliar do cafeeiro ‘Catuai’ em função da época de irrigação.
372 *Revista Brasileira de Engenharia Agrícola e Ambiental*. 01. 21-25. doi.org/10.1590/S1415-
373 43662008000100003.
374
- 375 SOUZA, F. F; FERRÃO, L. F. V; CAIXETA, E. T; SAKIYAMA, N. S; PEREIRA, A. A;
376 OLIVEIRA, A. C. B. Aspectos gerais da biologia e da diversidade genética de *Coffea*
377 *canephora*. (Ed.). *Café na Amazônia*. Brasília, DF: Embrapa, 2015. p.85-95.
378
- 379 TAMURA, K., STECHER, L., PETERSON, D., FILIPSKI, U., KUMAR S. 2013. MEGA6:
380 Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*. 30,
381 12, 2725-2729. 2013. DOI: 10.1093 / molbev / mst197.
382
- 383 TAIZ, L.; ZEIGER, E. *Fisiologia vegetal*. 3. ed. Porto Alegre: Artmed, 2004.720p.
384
- 385 THAMILARASAN, S.K; PARQUE, J. I; JUNG, H. J; NOU, I. S. 2014. Genome-wide
386 analysis of the distribution of AP2/ERF transcription factors reveals duplication and CBFs
387 genes elucidate their potential function in *Brassica oleracea*. *BMC Genomics*. 15, 422. 1-13.
388 DOI: 10.1186 / 1471-2164-15-422.
389

- 390 VAN DEN BROECK L, DUBOIS M, VERMEERSCH M, STORME V, MATSUI M, INZÉ
391 D. 2017. Da rede ao fenótipo: a fiação dinâmica de uma rede transcricional de Arabidopsis
392 induzida por estresse osmótico. *Mol Syst Biol* ; 13: –961. DOI: 10.15252/msb.20177840
393
- 394 WANG, C; YIN-MAO DONG. 2009. Overexpression of maize ZmDBP3 enhances tolerance
395 to drought and cold stress in transgenic Arabidopsis plants. *Biologia*. 64, 6,1108-1114.
396 doi.org/10.2478/s11756-009-0198-0.
397
- 398 XIE, Z., NOLAN, T.M., JIANG, H., YIN, Y. AP2/ERF Transcription Factor Regulatory
399 Networks in Hormone and Abiotic Stress Responses in Arabidopsis. *Frontiers in Plant*
400 *Science*, v. 10, 1-17. doi.org/10.3389/fpls.2019.00228
401
- 402 ZHANG, M. CHEN, X. CHEN, Z. XU, S. GUAN, L.C. LI, A. LI, J. GUO, L. MAO, Y. MA,
403 Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean
404 (*Glycine max* L.), *J. Exp. Bot.* 59 (2008) 4095–4107.
405
- 406

407 **FIGURE LEGENDS**

408

409 Table 1 – *Coffea canephora* transcriptional factors and their orthologs involved in water stress
410 response.

411

412 Figure 1 - Evolutionary conservation of proteins in *Coffea canephora*. Each domain is
413 represented by a green rectangle between the beginning and end of protein amino acid
414 residues. The size of each protein is indicated at the end of each sequence and the comparative
415 values of each domain are indicated to the right of each sequence.

416

417 Figure 2 - Amino acid residues involved in active site formation identified in *Coffea*
418 *canephora*. The cysteine residue (YRG and RAYD) appears to be the active site (indicated by
419 blue arrows), while the other amino acid residues are probably involved in other ways.

420

421 Figure 3 - Phylogenetic distribution of AP2 / EFR sequences found in *Coffea canephora* data
422 against its orthologous organisms.

423

424 Table 1 – *Coffea canephora* transcriptional factors and their orthologs involved in water stress
 425 response.

Gene	ID organism	Organism	Size (aa)	Conserved Domains	Blastp (e-value)
ERF017	Cc06_g12520	<i>Coffea canephora</i>	218	AP2 domain	3.3e ⁻⁰⁹
	Cc08_g07780	<i>Coffea canephora</i>	228	AP2 domain	1.2e ⁻¹⁰
	NP_173355.3	<i>Arabidopsis thaliana</i>	185	AP2 domain	6.6e ⁻¹²
ERF020	Cc10_g09120	<i>Coffea canephora</i>	137	AP2 domain	3.8e ⁻⁰⁷
	NP_177307.1	<i>Arabidopsis thaliana</i>	143	AP2 domain	3.9e ⁻⁰⁸
ERF011	Cc07_g06220	<i>Coffea canephora</i>	164	AP2 domain	1.9e ⁻¹⁰
	NP_190595.1	<i>Arabidopsis thaliana</i>	153	AP2 domain	3.1e ⁻¹¹
ERF014	Cc10_g10960	<i>Coffea canephora</i>	249	AP2 domain	5.6e ⁻¹¹
	NP_175104.1	<i>Arabidopsis thaliana</i>	211	AP2 domain	4.2e ⁻¹¹
ERF118	Cc11_g10210	<i>Coffea canephora</i>	381	AP2 domain	2.4e ⁻¹¹
	NP_177022.1	<i>Arabidopsis thaliana</i>	324	AP2 domain	2.8e ⁻¹²
ERF23	Cc02_g39490	<i>Coffea canephora</i>	263	AP2 domain	2.0e ⁻¹²
	NP_563624.1	<i>Arabidopsis thaliana</i>	192	AP2 domain	1.0e ⁻¹¹
ERF039	Cc02_g24810	<i>Coffea canephora</i>	296	AP2 domain	4.7 e ⁻¹³

	Cc08_g12160	<i>Coffea canephora</i>	223	AP2 domain	1.4 e ⁻¹²
	NP_193408.1	<i>Arabidopsis thaliana</i>	179	AP2 domain	1.1 e ⁻¹¹
ERF24	Cc06_g05340	<i>Coffea canephora</i>	244	AP2 domain	2.0 e ⁻¹²
	NP_181186.1	<i>Arabidopsis thaliana</i>	184	AP2 domain	5.0 e ⁻¹²
ERF098	Cc03_g05400	<i>Coffea canephora</i>	170	AP2 domain	2.5 e ⁻¹⁴
	Cc01_g18400	<i>Coffea canephora</i>	155	AP2 domain	2.0 e ⁻¹⁴
	NP_188964.2	<i>Arabidopsis thaliana</i>	139	AP2 domain	1.6 e ⁻¹⁴
ERF54	Cc05_g06840	<i>Coffea canephora</i>	493	AP2 domain	4.2 e ⁻¹²
	NP_194543.1	<i>Arabidopsis thaliana</i>	292	AP2 domain	2.5 e ⁻¹³
ERF12	Cc02_g13680	<i>Coffea canephora</i>	146	AP2 domain	3.1e ⁻¹³
	XP_002279760.1	<i>Vitis vinifera</i>	152	AP2 domain	3.4e ⁻¹³
ERF003	Cc08_g00110	<i>Coffea canephora</i>	193	AP2 domain	1.3e ⁻¹¹
	Cc04_g04880	<i>Coffea canephora</i>	182	AP2 domain	3.1 e ⁻¹¹
	XP_016462558.1	<i>Nicotiana tabacum</i>	199	AP2 domain	3.6 e ⁻¹¹
ERF062	Cc03_g07870	<i>Coffea canephora</i>	426	AP2 domain	3.7e ⁻¹²
	XP_002268004.1	<i>Vitis vinifera</i>	442	AP2 domain	7.3e ⁻¹¹

ERF61	Cc08_g15980	<i>Coffea canephora</i>	310	AP2 domain	5.0e ⁻¹³
	NP_176620.1	<i>Arabidopsis thaliana</i>	335	AP2 domain	5.7e ⁻¹³
RAP2-11	Cc06_g23190	<i>Coffea canephora</i>	287	AP2 domain	2.7e ⁻¹³
	NP_197480.1	<i>Arabidopsis thaliana</i>	253	AP2 domain	7.9e ⁻¹³
RAP2-12	Cc04_g00840	<i>Coffea canephora</i>	404	AP2 domain	5.0e ⁻¹¹
	NP_001031185.1	<i>Arabidopsis thaliana</i>	358	AP2 domain	1.0e ⁻¹²
ERF071	Cc01_g21090	<i>Coffea canephora</i>	328	AP2 domain	1.5e ⁻¹¹
	NP_182274.1	<i>Arabidopsis thaliana</i>	171	AP2 domain	1.4e ⁻¹²
ERF3	Cc05_g14220	<i>Coffea canephora</i>	241	AP2 domain	2.2e ⁻¹²
	NP_175479.1	<i>Arabidopsis thaliana</i>	225	AP2 domain	6.2e ⁻¹³
ERF8	Cc01_g18710	<i>Coffea canephora</i>	184	AP2 domain	1.2e ⁻¹²
	XP_019257965.1	<i>Nicotiana attenuata</i>	173	AP2 domain	1.6e ⁻¹²
ERF12	Cc02_g13680	<i>Coffea canephora</i>	146	AP2 domain	3.1 e ⁻¹³
	XP_002279760.1	<i>Vitis vinifera</i>	152	AP2 domain	3.4 e ⁻¹³
ESR2	Cc11_g15790	<i>Coffea canephora</i>	394	AP2 domain	2.0 e ⁻¹⁰
	XP_022775602.1	<i>Durio zibethinus</i>	433	AP2 domain	1.4 e ⁻¹⁰

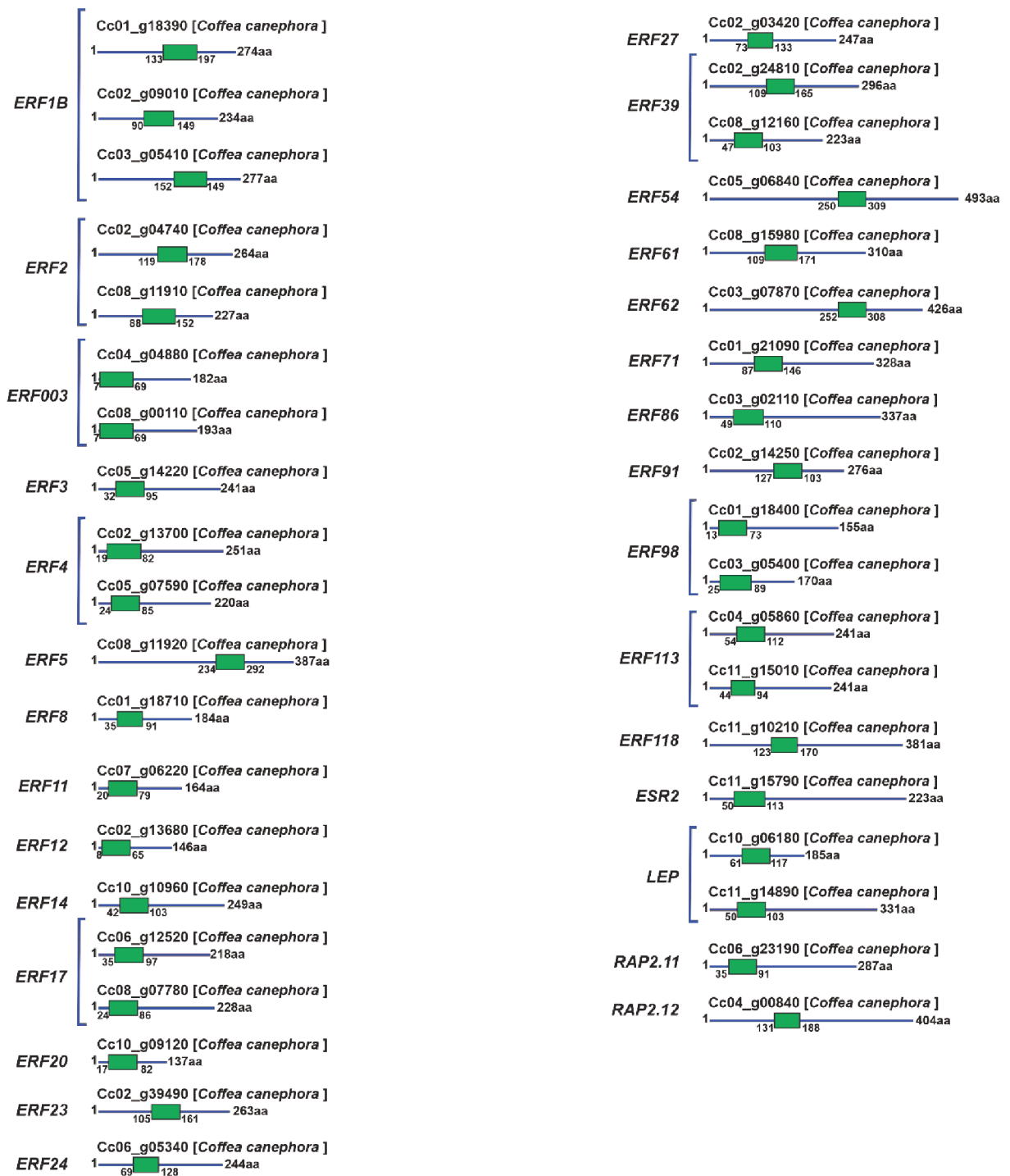
ERF-LEP	Cc10_g06180	<i>Coffea canephora</i>	185	AP2 domain	1.0 e ⁻⁸
	Cc11_g14890	<i>Coffea canephora</i>	331	AP2 domain	8.0 e ⁻¹²
	XP_004238505.1	<i>Solanum lycopersicum</i>	160	AP2 domain	1.3 e ⁻¹¹
ERF86	Cc03_g02110	<i>Coffea canephora</i>	337	AP2 domain	6.1 e ⁻¹¹
	NP_197357.2	<i>Arabidopsis thaliana</i>	348	AP2 domain	6.1 e ⁻¹¹
ERF091	Cc02_g14250	<i>Coffea canephora</i>	276	AP2 domain	5.3 e ⁻¹³
	NP_193580.1	<i>Arabidopsis thaliana</i>	303	AP2 domain	4.5 e ⁻¹¹
ERF1B	Cc01_g18390	<i>Coffea canephora</i>	274	AP2 domain	1.5 e ⁻¹³
	Cc02_g09010	<i>Coffea canephora</i>	234	AP2 domain	6.9 e ⁻¹³
	Cc03_g05410	<i>Coffea canephora</i>	277	AP2 domain	1.2 e ⁻¹²
	NP_188965.1	<i>Arabidopsis thaliana</i>	218	AP2 domain	1.8 e ⁻¹³
ERF4	Cc02_g13700	<i>Coffea canephora</i>	251	AP2 domain	7.9 e ⁻¹⁴
	Cc05_g07590	<i>Coffea canephora</i>	220	AP2 domain	5.6 e ⁻¹⁴
	XP_002279692.1	<i>Vitis vinifera</i>	249	AP2 domain	1.7 e ⁻¹³
ERF5	Cc08_g11920	<i>Coffea canephora</i>	387	AP2 domain	3.4 e ⁻¹⁴
	NP_568679.1	<i>Arabidopsis thaliana</i>	300	AP2 domain	7.5 e ⁻¹⁵

ERF113	Cc11_g15010	<i>Coffea canephora</i>	241	AP2 domain	1.0 e ⁻¹³
	Cc04_g05860	<i>Coffea canephora</i>	241	AP2 domain	1.6 e ⁻¹³
	NP_196837.1	<i>Arabidopsis thaliana</i>	212	AP2 domain	4.3 e ⁻¹⁴
ERF4	Cc02_g13700	<i>Coffea canephora</i>	251	AP2 domain	7.9 e-14
	XP_002279692.1	<i>Vitis vinifera</i>	249	AP2 domain	1.7 e-13
ERF2	Cc08_g11910	<i>Coffea canephora</i>	227	AP2 domain	9.1 e-13
	NP_199533.1	<i>Arabidopsis thaliana</i>	243	AP2 domain	4.2 e-13
ERF027	Cc02_g03420	<i>Coffea canephora</i>	247	AP2 domain	7.3 e-10
	NP_172723.2	<i>Arabidopsis thaliana</i>	192	AP2 domain	4.2 e-13

426

427

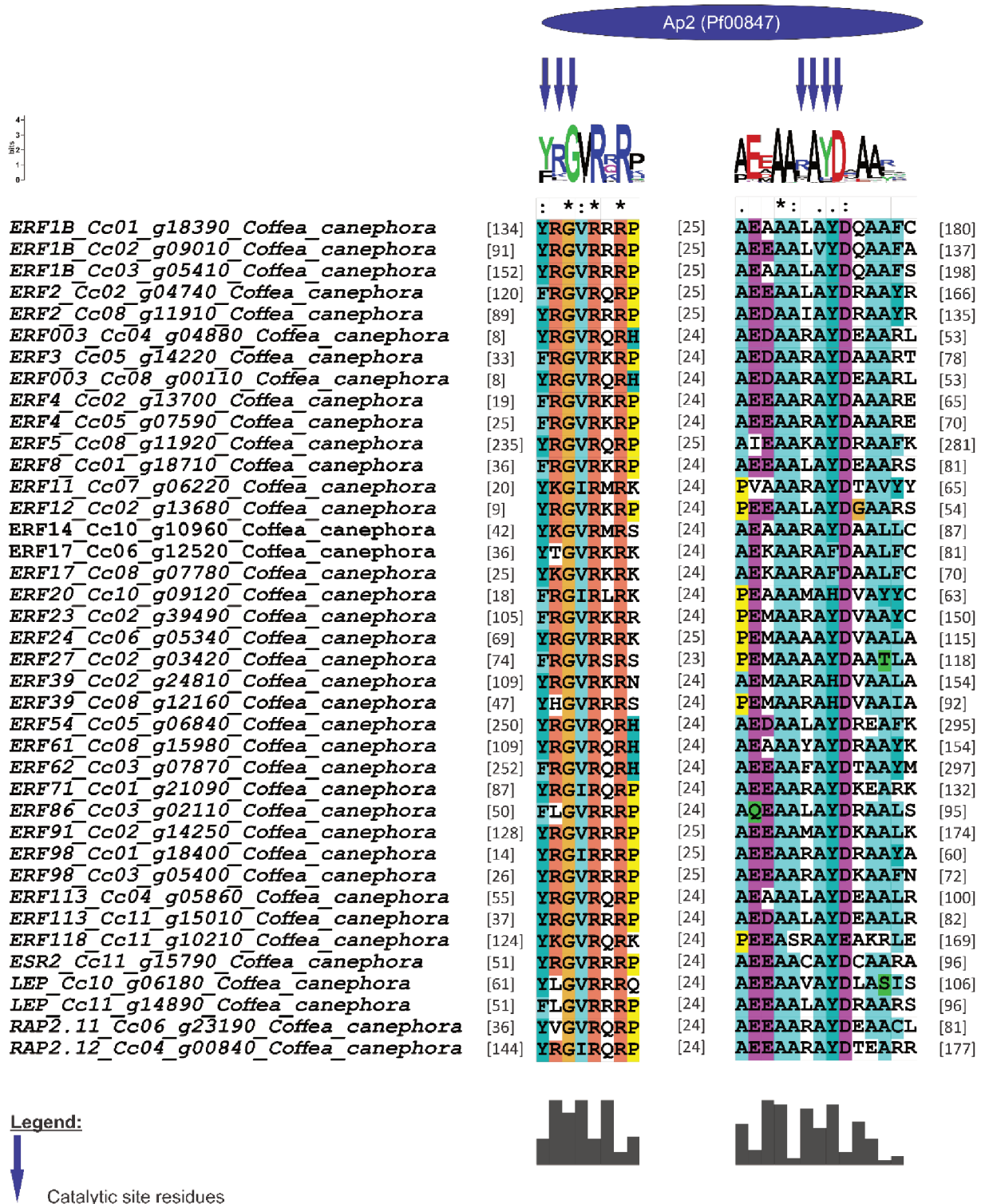
428 **Figure 1** - Evolutionary conservation of proteins in *Coffea canephora*. Each domain is
 429 represented by a green rectangle between the beginning and end of protein amino acid
 430 residues. The size of each protein is indicated at the end of each sequence and the comparative
 431 values of each domain are indicated to the right of each sequence.



432

433

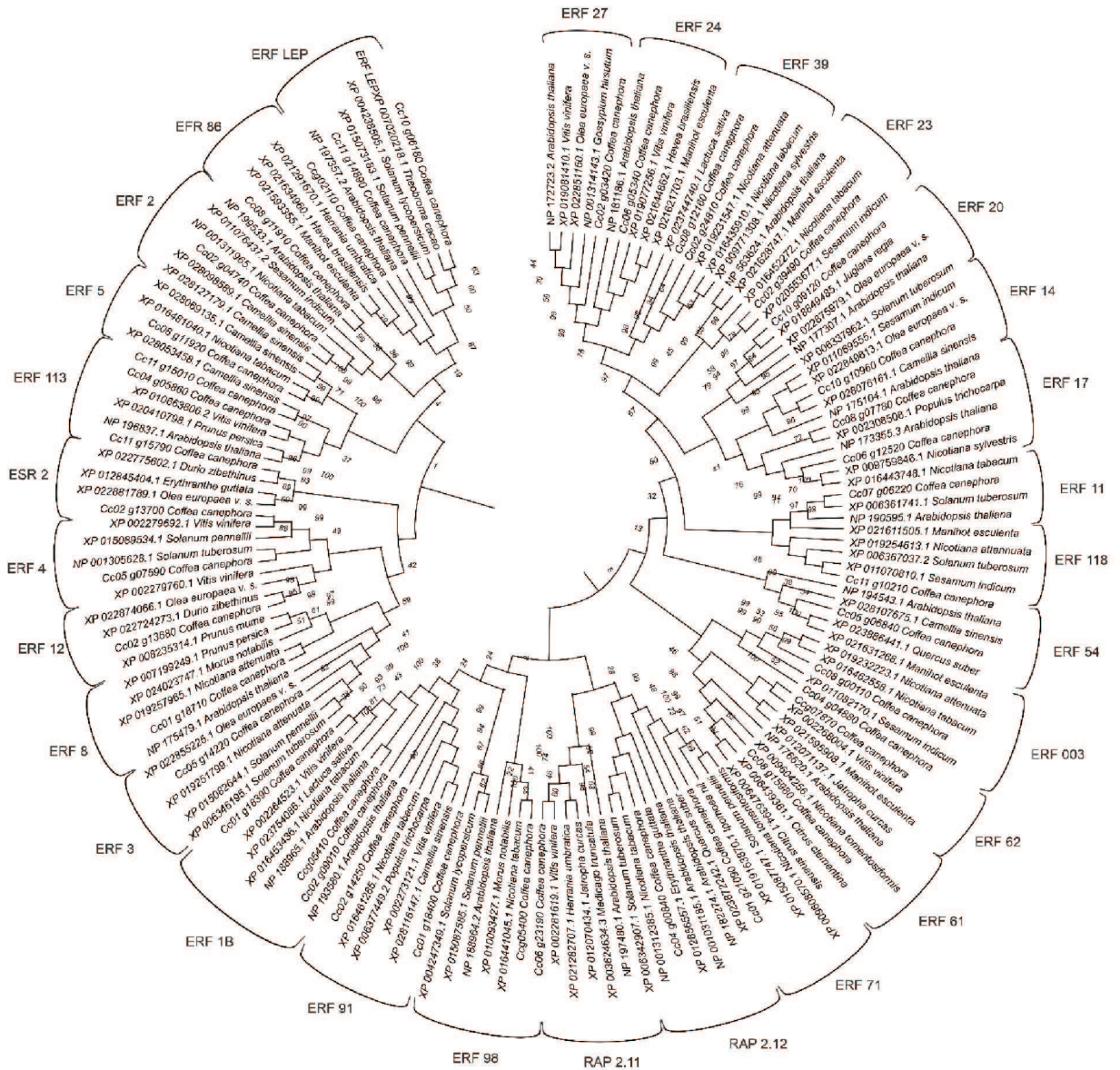
434 Figure 2 - Amino acid residues involved in active site formation identified in *Coffea*
 435 *canephora*. The cysteine residue (YRG and RAYD) appears to be the active site (indicated by
 436 blue arrows), while the other amino acid residues are probably involved in other ways.



437

438

439 **Figure 3** - Phylogenetic distribution of AP2 / EFR sequences found in *Coffea canephora* data
440 against its orthologous organisms.



441

442