

UNIVERSIDADE FEDERAL DE PELOTAS
Programa de Pós-Graduação em Ciência e Tecnologia
Agroindustrial
Faculdade de Agronomia Eliseu Maciel



Dissertação

**Análise “*in silico*” e de expressão da família gênica
ETHYLENE RESPONSE FACTORS (ERF) no gênero
*Malus***

Joceani Dal Cero

Pelotas, fevereiro de 2010.

Joceani Dal Cero

**Análise “*in silico*” e de expressão da família
gênica *ETHYLENE RESPONSE FACTORS (ERF)*
no gênero *Malus***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia Agroindustrial da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciências e Tecnologia Agroindustrial.

Comitê de Orientação

César Luís Girardi

Cesar Valmor Rombaldi

Pelotas 2010

Dados de catalogação na fonte:
(Marlene Cravo Castillo – CRB-10/744)

D136a Dal Cero, Joceani

Análise “in silico” e de expressão da família gênica Ethylene response factors (ERF) no gênero *Malus* / Joceani Dal Cero ; orientador César Luís Girardi; co-orientador Cesar Valmor Rombaldi . - Pelotas, 2010. -126f. ; il. - Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência e Tecnologia Agroindustrial. Faculdade de Agronomia Eliseu Maciel. Universidade Federal de Pelotas. Pelotas, 2010.

1. Análise in silico 2 .*Malus* 3.ERF 4.Etileno 5. Maça Royal gala 6. Análise de expressão 7.*Venturia inaequalis* I Girardi, César Luís (orientador) II .Título.

CDD 664.8

BANCA EXAMINADORA:

Dr. César Luís Girardi

Prof. Dr. Cesar Valmor Rombaldi

Dra. Vera Maria Quecini

Dra. Luciane Arantes de Paula

AGRADECIMENTOS

A Cesar Valmor Rombaldi pelos preciosos ensinamentos.

A César Luis Girardi pela orientação e a oportunidade de realizar meu trabalho de dissertação junto à Embrapa Uva e Vinho.

A Vera Quecini, pela valiosa orientação e apoio

A Jorge Adolfo Silva, pelo apoio e amizade.

A Lucimara Antonioli pela simplicidade e amizade.

Aos amigos e companheiros do laboratório pós-colheita da Embrapa, em especial a Sayuri, Miqueli, Laís e demais estagiários. E a todos os laboratórios que disponibilizaram as instalações e equipamentos para a realização do experimento.

A Taís Letícia Bernardi e Rosane Giacomini pela paciência, força, palavras amigas e todos os ótimos momentos.

Ao Programa de Pós-graduação em Ciência e Tecnologia. E aos amigos do DCTA, especilamente Camila Pegoraro e Jocleita Ferrareze.

A CAPES pela concessão das bolsas de estudos.

Aos meus pais Nelson e Maria Dal Cero e irmãos Gelson, Jane e Jacó pela força, paciência, incentivo, amor e por nunca terem desistido de mim.

E finalmente a Deus por ter proporcionado tudo isso.

RESUMO

DAL CERO, Joceani. **Análise “*in silico*” e de expressão da família gênica *ETHYLENE RESPONSE FACTORS (ERF)* no gênero *Malus***. 2010. 126 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência e Tecnologia Agroindustrial. Universidade Federal de Pelotas, Pelotas.

Moléculas que participam dos processos regulatórios, como os fatores de transcrição, têm recebido atenção especial, pois uma das principais ações dos estímulos hormonais é a modulação da expressão gênica. Como a taxa de transcrição de um gene é o maior determinante da sua expressão, os mecanismos moleculares pelos quais a transcrição gênica é regulada têm se tornado um dos tópicos principais de estudos em genética molecular envolvendo o hormônio etileno. O objetivo deste trabalho foi realizar análises de bioinformática para a família ERF (Ethylene Response Factors), integrar bases de dados existentes na internet no modelo *Arabidopsis*, bem como análise filogenética que permitam avaliar os papéis dos diferentes membros da família. Este levantamento preliminar das seqüências ERF em *Malus* forneceu informação básica para estudos posteriores mais aprofundados, com relação aos mecanismos moleculares da família nesta importante cultura perene. A análise da expressão de *MdERF1* e *MdERF2* em frutos de maçã indica que outros fatores além do etileno estão envolvidos na regulação da transcrição dos *ERF* em *Malus*. O segundo capítulo refere-se à resposta dos *ERF* frente ao ataque de patógenos. Para isso, foram infectadas plantas de macieira provenientes de cultivo *in vitro* com o fungo *Venturia inaequalis* (sarna da maçã). As evidências desses estudos sugerem o envolvimento do gene *MdERF1* no processo de patogênese, enquanto que o gene *MdERF2* parece não estar envolvido no processo.

Palavras chaves: *Malus*, análise *in silico*, ERF, etileno, análise de expressão, maçã e *Venturia inaequalis*.

ABSTRACT

***In silico* and expression analysis of the *ETHYLENE RESPONSE FACTORS (ERF)* gene family in the genus *Malus*.**

Regulatory molecules, such as transcription factors, have been thoroughly investigated, especially in hormone-mediated responses that involve gene expression modulation. Frequently, the main determinant of gene expression is its transcriptional rate. Thus, molecular mechanisms underlying transcription regulation have become an important topic in genetic studies of ethylene signaling. The present work aimed to investigate the ERF (Ethylene Response Factor) family employing bioinformatic tools, integrating publicly available datasets from the model species *Arabidopsis thaliana* and phylogenetic analyses to help elucidating the biological roles of the family in apple. The preliminary survey of the ERF sequences in *Malus* has provided basic information to be incorporated in further studies of the functional role of ERFs in this perennial species. Expression analyses of *MdERF1* and *MdERF* in apple fruits suggest that other factors, besides ethylene, are involved in their transcriptional regulation in *Malus*. The second chapter reports the investigation of the transcriptional profiling of those *ERF* genes in response to pathogen attack, using a biological assay of *in vitro* propagated plants inoculated with the fungus *Venturia inaequalis* (apple scab disease). The study has provided evidences of the involvement of *MdERF1* in eliciting the plant response; whereas, *MdERF2* does not appear to be participate in the pathogenesis.

Keywords: *Malus*, *in silico* analysis, ERF, ethylene, expression analysis, apple and *Venturia inaequalis*

LISTA DE FIGURAS

Página

REVISÃO BIBLIOGRÁFICA	15
Figura 1 - Via de biossíntese do etileno descoberto por Yang em 1980.....	16
Figura 2 - Modelos de sinalização do etileno em <i>Arabidopsis</i> . Adaptado de McCourt, 1999.....	17
Figura 3 - Ilustração esquemática da via de tradução de sinais do etileno.....	18
Figura 4 - Ilustração esquemática da organização do domínio de AP2, DREB, ERF e proteína RAV.....	19
Figura 5 - Função fisiológica dos ERFs em resposta aos diferentes estresses.....	21
 Capítulo I	 28
 Figure 1 – Sequence alignment and structure prediction of the AP2/ERF domain of proteins from the ERF sub-family in <i>Malus</i> sp.. <i>Malus</i> deduced amino acid sequences were aligned with ClustalX. Black and light gray shading indicate identical and conserved amino acid residues in the Clustal Consensus. The divergent sequences are highlighted in red boxes. The black bar and arrows represent predicted α -helix and β -sheet regions within the AP2/ERF domain, respectively.	 42
Figure 2 - AP2/ERF-type transcription factors from the ERF sub-family in <i>Malus</i> . Neighbor-joining tree for <i>Malus</i> deduced amino acid and <i>Arabidopsis</i> full length sequences were aligned with ClustalX. Bootstrap values are indicated above each branch. Cluster nomenclature according to Nakano et al. (2006) is exhibited as roman numerals and classification by Sakuma et al. (2002) is indicated in parentheses.....	43
Figure 3 - Phylogenetic relationships among the <i>Arabidopsis</i> ERF genes, from group I (A), group II (B), group III (C), group IV (D), group V (E), group VI-L (F), group VII (G), group VIII (H), group IX (I) and group X (J) in <i>Arabidopsis</i> and <i>Malus</i> ERF families. Bootstrap values from 1000 replicates are shown above the branches. The phylogenetic tree and a schematic diagram of the protein structures of groups I to X are shown in A to J, respectively. Each colored box represents the AP2/ERF domain and conserved motifs, as indicated in legends below the trees. The amino acid sequences of the conserved motifs are summarized in Supplementary Table IV. Classification by Nakano et al. (2006) is shown above each tree and by Sakuma et al. (2002), indicated in parentheses.....	45
Figure 4 - Prediction of the folding state of six ERF sequences from <i>Malus</i> : A1LM1, A1LM2, A5YRQ8, A6XA55, Q2LMD4 and Q8VWW8, in comparison to <i>Arabidopsis thaliana</i> At1g50680 RAV1 and <i>Populus deltoides</i> PdTC131879. The fold index calculated by the software FoldIndex (Prilusky et al., 2005) is represented by the Y axis and the amino acids of the protein are represented by their position in the sequence in the X axis. Folded protein regions are represented in green and, unfolded in red.....	47

Figure 5- Expression profile of four ERF sequences from *Malus*. The normalized number of reads for the transcripts in each library is represented as a grayscale. The names of the libraries and its tissue type are listed at left and right, respectively. Detailed description of the libraries is provided elsewhere (http://compbio.dfci.harvard.edu/tgi/gi/mdgi/searching/xpress_search.html). Hierarchical clustering of the expression patterns by *k*-means using Spearman Rank correlation is represented by tree.49

Capítulo II.....70

Figura 1 - Plantas controle (A) e com sintomas de sarna (B).....79

Figura 2 - Acúmulo de transcritos dos genes *MdActina*, *MdERF1* e *MdERF2* em plantas de maçã, cv Royal Gala não inoculadas (1) e inoculadas (2) com fungo *Venturia inaequalis*.....80

LISTA DE TABELAS

Página

Table 1 Summary of the AP2/ERF superfamily in *Arabidopsis thaliana* and *Malus* sp.....35

Table 2 *Malus* ESTs with homology to genes of the ERF family in *Arabidopsis thaliana*...37

LISTA DE ANEXOS

Página

Supplemental Table I Unigene list of AP2/ERF superfamily in <i>Malus</i>	87
Supplemental Table II Conserved motifs in the ERF family in <i>Malus</i>	92
Supplementary Table III Summary of conserved motifs(CMs) within <i>Malus</i> ERF family in comparison to <i>A. thaliana</i> orthologs.....	96
Supplementary Table IV Prediction of the molecular and cellular properties of <i>Malus</i> ERF-like sequences.....	113
Figure 3 - Phylogenetic relationships among the <i>Arabidopsis</i> ERF genes, from group I (A), group II (B), group III (C), group IV (D), group V (E), group VI-L (F), group VII (G), group VIII (H), group IX (I) and group X (J) in <i>Arabidopsis</i> and <i>Malus</i> ERF families. Bootstrap values from 1000 replicates are shown above the branches. The phylogenetic tree and a schematic diagram of the protein structures of groups I to X are shown in A to J, respectively. Each colored box represents the AP2/ERF domain and conserved motifs, as indicated in legends below the trees. The amino acid sequences of the conserved motifs are summarized in Supplementary Table IV. Classification by Nakano et al. (2006) is shown above each tree and by Sakuma et al. (2002), indicated in parentheses.....	117

SUMÁRIO

Página

1 INTRODUÇÃO	14
2 REVISÃO BIBLIOGRÁFICA	15
2.1- Biossíntese do Etileno	15
2.2- Via de sinalização do etileno	16
2.3- Fatores de transcrição ERF	17
2.4- Classificação de fatores de transcrição AP2/ERF	19
2.5- Estrutura do domínio AP2/ERF	19
2.6- Diferenças entre ERF e DREB	20
2.7- Etileno e mecanismos de defesa contra patógenos	22
3 REFERÊNCIAS BIBLIOGRÁFICA	23
Capítulo I	28
Artigo: <i>In silico</i> analysis of the Ethylene Response Factor (ERF) gene family in <i>Malus</i>	
ABSTRACT	29
1 INTRODUCTION	30
2 MATERIAL AND METHODS	32
2.1- Database Searches and Alignments	32
2.2- Phylogenetic Analysis	32
2.3- Motif analysis and <i>in silico</i> characterization	33
2.4- Prediction of cellular and molecular parameters	33
2.5- <i>In silico</i> gene expression analysis	33
3 RESULTS	34
3.1- The family of ethylene response factors in apple	34
3.2- Conserved motifs outside the AP2/ERF domain	44
3.3- Cellular and molecular characteristics of apple ERF-like proteins	46
3.4- <i>In silico</i> expression profiling of ERF genes in apple	48
4 DISCUSSION	50
4.1- Conserved motifs outside the AP2/ERF domain	51
4.2- Cellular and molecular characteristics of apple ERF-like proteins	52
4.3- <i>In silico</i> expression profiling of ERF genes in apple	54
5 ACKNOWLEDGMENTS	56
6 RESUMO DO ARTIGO EM PORTUGUÊS	56

6.1- INTRODUÇÃO.....	56
6.2- RESULTADOS.....	59
6.2.1- A família ERF em maçã.....	59
6.2.2- Motivos conservados fora do domínio AP2/ERF.....	60
6.2.3- Características moleculares e celulares.....	60
6.2.4- Perfis <i>in silico</i> da expressão gênica.....	61
7 REFERENCES.....	62
Capítulo II.....	70
Expressão dos genes <i>MdERF1</i> e <i>MdERF2</i> em plantas infectadas com fungo <i>Venturia inaequalis</i>	
1 INTRODUÇÃO.....	70
2 MATERIAL E MÉTODOS.....	72
2.1- Material Vegetal.....	72
2.1.1- Cultura de tecidos.....	72
2.1.2- Inoculação.....	73
2.2- Estudos da expressão do <i>MdERF1</i> e <i>MdERF2</i>	74
2.3- Extração de RNA total.....	74
2.3.1-Tratamento com DNase e RNase-free.....	75
2.3.2- Precipitação dos RNAs.....	75
2.3.3- Análise quantitativa e qualitativa dos RNAs totais extraídos.....	76
2.4- Síntese de cDNA.....	77
2.5- PCR semi-quantitativa.....	78
3 RESULTADO E DISCUSSÃO.....	78
3.1- Avaliação das plantas inoculadas com <i>Venturia inaequalis</i>	78
3.2- Estudo de expressão ERF.....	80
4 CONCLUSÃO.....	81
5 REFERÊNCIAS BIBLIOGRÁFICAS.....	81
CONSIDERAÇÕES FINAIS.....	85
ANEXOS.....	86

1 INTRODUÇÃO

Nos últimos anos muitos esforços têm sido empregados para elucidar os mecanismos envolvidos na regulação da transcrição gênica. Moléculas que participam desses processos regulatórios, como os fatores de transcrição, têm recebido atenção especial. Uma das principais ações dos estímulos hormonais é a modulação da expressão dos genes. Como a taxa de transcrição de um gene é o maior determinante da sua expressão, os mecanismos moleculares pelos quais a transcrição gênica é regulada têm ganhado interesse crescente e se tornado um dos tópicos principais de estudos em genética molecular envolvendo o hormônio etileno. O etileno é importante em vários processos fisiológicos que ocorrem nas plantas, sendo indutor de respostas relacionadas ao estresse. A cascata de sinais que ele regula termina pela indução de fatores de transcrição específicos que controlam as diferentes respostas relacionadas a este hormônio. Os ERF (ETHYLENE RESPONSE FACTOR) são, portanto, os últimos elementos da cascata de sinais, sendo que constituem uma grande família multigênica específicas das plantas, caracterizada por apresentar um domínio envolvido na ligação do DNA, o domínio ERF. Muitas proteínas contendo o domínio ERF foram isoladas e caracterizadas nos últimos anos em várias espécies de plantas. Em *Arabidopsis*, 124 genes foram identificados como pertencentes à família ERF, sendo a terceira família de fator de transcrição, pelo número de membros, após os fatores *MYB* e *bHLH* (Riechmann et al., 2000). Em maçã, apenas 2 *ERF* (*MdERF1* e *MdERF2*) foram estudados até o momento (Wang et al. 2007).

Assim, o objetivo deste trabalho foi realizar análises de bioinformática para esta família, integrar bases de dados existentes ao modelo *Arabidopsis*. Finalmente, comparar o conhecimento disponível em *Arabidopsis* a fim de identificar as estruturas da família comuns ao gênero *Malus*. Um segundo capítulo, refere-se à resposta de ERF frente ao ataque de patógenos. Para isso, plantas infectadas com o fungo *Venturia inaequalis*

(sarna da maçã) foram estudadas, procurando verificar o funcionamento e resposta de alguns desses fatores de transcrição frente a esse tipo de estresse, discutindo as bases moleculares da percepção desse sinal pela planta, assim como muitas outras funções desse hormônio em diferentes respostas fisiológicas ou mecanismos de defesa.

2 REVISÃO BIBLIOGRÁFICA

2.1- Biossíntese do Etileno

Etileno foi evidenciado pela primeira vez por Gane (1934) durante o amadurecimento de maçã. Desde então, tem sido demonstrado que todos os tecidos vegetativos e reprodutivos são capazes de sintetizar etileno (Yang e Hoffman, 1984). As etapas de sua biossíntese são atualmente bem conhecidas e representadas na figura 1 (Yang e Hoffman, 1984). Sua biossíntese inicia a partir do aminoácido metionina, que é convertido em SAM (S-adenosil L-metionina). As enzimas ACCsintase (ACCS) e ACCoxidase (ACCO) catalisam respectivamente, a formação de ACC a partir de SAM, e etileno a partir da ACC (Yang e Hoffman, 1984). Apesar da simplicidade estrutural, o etileno desempenha um importante papel na regulação de diversos processos fisiológicos como senescência foliar, abscisão de órgãos, amadurecimento de frutos, entre outros. Além disso, constitui resposta ao estresse causado por vários tipos de patógenos, por danos mecânicos e pela seca.

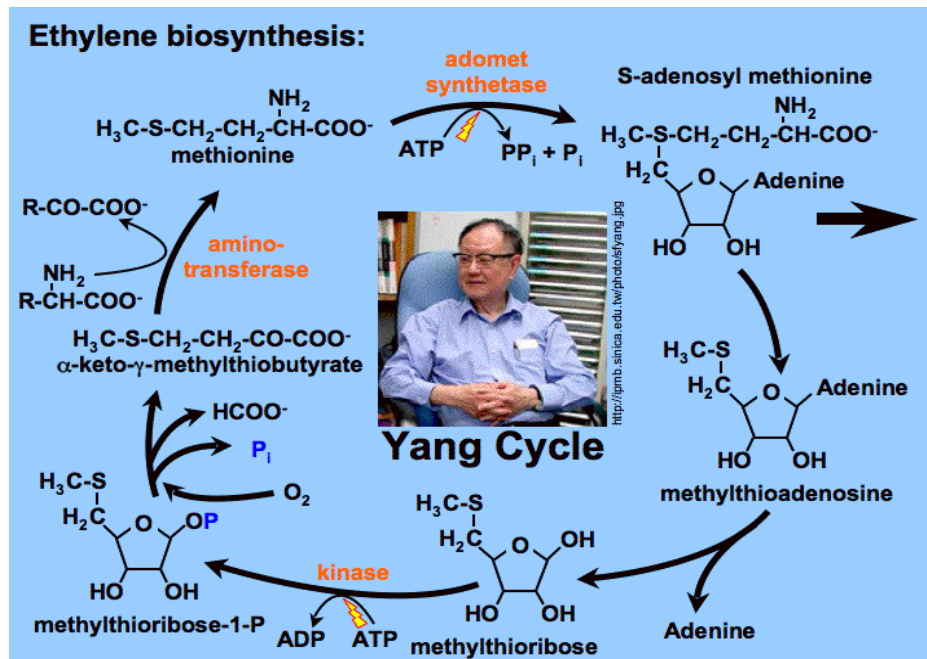


Figura 1: Via de biossíntese do etileno descoberto por Yang em 1980 (http://plantphys.info/plant_physiology/ethylene.shtml)

2.2- Via de sinalização do etileno

O ETR1 foi o primeiro receptor de etileno a ser descoberto, ele é um sensor do sistema de dois componentes da tradução histidina quinase de bactérias, que consiste de um sensor histidina quinase e um regulador de resposta, que pode agir como um fator de transcrição. O genoma de *Arabidopsis* codifica quatro proteínas adicionais semelhantes ao ETR1, são eles ETR2, ERS1, ERS2 e EIN4, que também funcionam como receptores de etileno (Taiz & Zeiger, 2004). O domínio transmembrana do receptor é o sinal de ligação ao etileno, através de um co-fator cobre (Rodriguez et al., 1999). Segundo McCourt (1999) existem dois possíveis modelos de sinalização de etileno, no primeiro, na ausência de etileno o gene *ETR1* ativaria o gene *CTR1*, que inibiria o gene *EIN3*, bloqueando a resposta. Na presença de etileno o gene *ETR1* seria bloqueado, o gene *CTR1* ficaria inativo e o gene *EIN3* expressaria a resposta ao etileno (Figura 2A). No segundo modelo, na ausência de etileno o *ETR1*, estaria inativo, o *CTR1* ativo bloquearia o gene *EIN3* sem resposta ao etileno. A presença de etileno ativaria *ETR1*, que

bloquearia o *CTR1* e assim o gene *EIN3* ficaria ativo, permitindo a resposta ao etileno (Figura 2B).

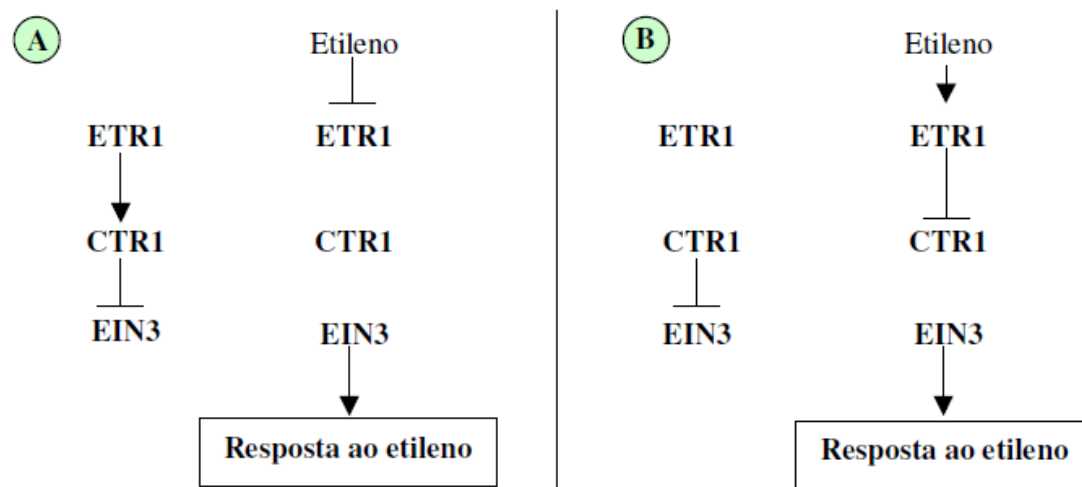


Figura 2: Modelos de sinalização do etileno em *Arabidopsis*. Adaptado de McCourt, 1999.

2.3- Fatores de transcrição ERF

Os fatores de transcrição são proteínas que se ligam a genes em locais específicos, tais como promotores, e, portanto, desempenham um papel na ativação ou repressão da transcrição de genes aumentando ou diminuindo (induzir ou reprimir) a expressão gênica (Zhang 2003). Os fatores de transcrição são responsáveis pela seletividade na regulação genética e muitas vezes são expressos em tecido específico, estágio específico de desenvolvimento ou via caminhos dependentes de estímulo exógenos (por exemplo, do ambiente) e estímulos endógenos durante todo o ciclo de vida do organismo. (Liu et al., 1999; Zhang 2003; Davuluri et al., 2003). O domínio de ligação ao DNA (DNA-binding domain) é responsável pela união do fator de transcrição a seqüências *cis* reguladora do DNA do promotor específicas dos genes que eles regulam. A maior parte de fatores de transcrição apresentam somente um tipo de ligação ao DNA, ocasionalmente em múltiplas cópias, mas alguns contêm dois tipos distintos de domínio (Herr et al., 1988).

A via de sinalização do etileno é parcialmente conhecida, sabendo-se, porém que este hormônio regula a expressão gênica a nível transcricional. Sabe-se também que esse caminho de sinalização termina ativando fatores de transcrição específicos que controlam diferentes respostas. Os mutantes de *Arabidopsis ein3*, caracterizados pela insensibilidade parcial ao etileno, tem sido clonado. Seu gene correspondente codifica um fator de transcrição pertencente a uma família exclusiva de plantas superiores (Chao et al. 1997; Solano et al., 1998). O *EIN3* tem três homólogos em *Arabidopsis* nomeados *EIL 1,2* e *3* (*EIN3 Like*) que são reguladores positivos da resposta ao etileno. Foi demonstrado que a proteína EIN3 se liga no promotor do gene *ERF1*, a um elemento regulador cis chamado PERE (Primary Ethylene Regulator Element), similar ao elemento presente no promotor do gene *E4* do tomate. O *ERF1* por sua vez também é um fator de transcrição da grande família de ERF (Ethylene Response Factor), os quais se ligam a elementos GCC no promotor de genes regulados por etileno (elementos de resposta secundários). Tem sido demonstrado que os ERF estão envolvidos em diversos processos de respostas a estresses bióticos e abióticos, sendo os mesmos considerados como bons candidatos para explicar a diversidade de resposta ao etileno (Chao et al., 1997).

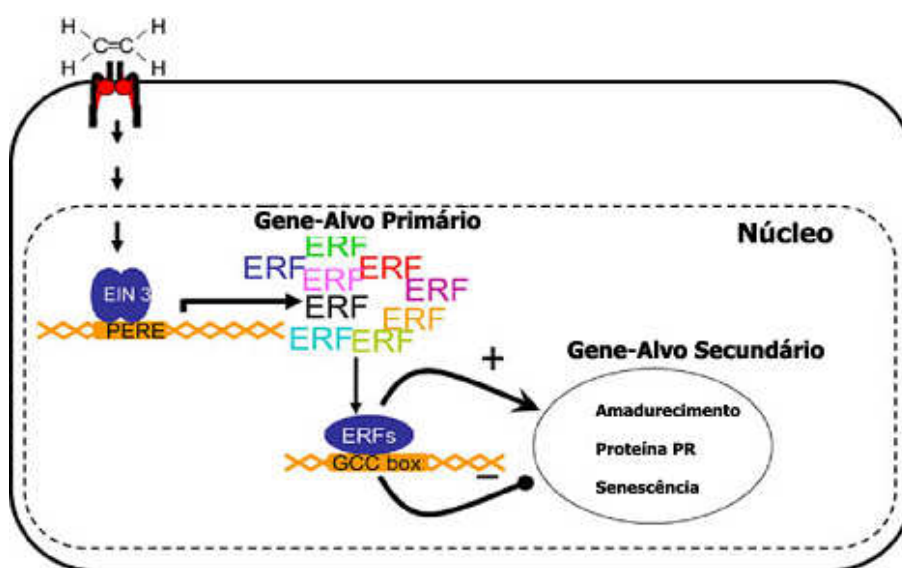


Figura 3. Ilustração esquemática da via de tradução de sinais do etileno. (<http://gbf.ensat.fr>).

2.4- Classificação de fatores de transcrição AP2/ERF

Baseado no número de cópias do domínio AP2/ERF e a sua semelhança de sequência, a proteína AP2/ERF foi subdividida em cinco subfamílias (Sakuma et al., 2002): subfamília AP2, subfamília DREB, subfamília ERF, subfamília RAV e outros. Os membros da subfamília AP2 contêm dois domínios AP2/ERF unidos por um linker conservado de 25 aminoácidos, sendo que os membros do DREB, ERF e outros subgrupos AP2/ERF contêm um único domínio.

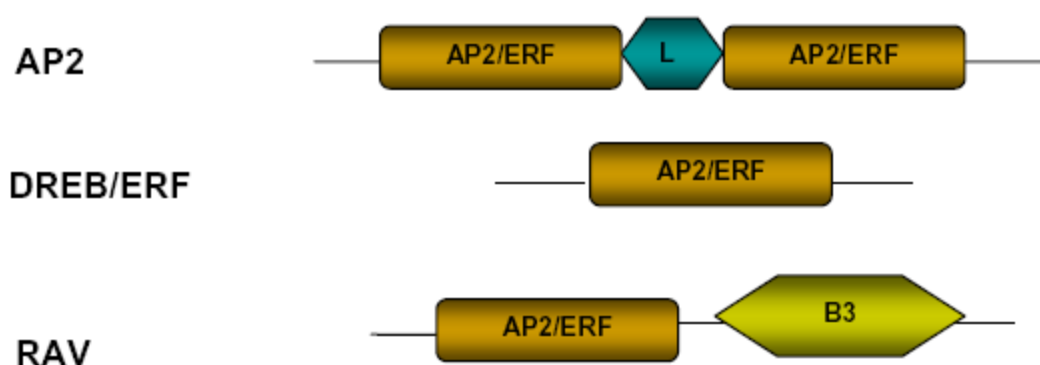


Figura 4. Ilustração esquemática da organização do domínio de AP2, DREB, ERF e proteína RAV. (Giraudat et al., 1992; Kagaya et al., 1999)

2.5- Estrutura do Domínio AP2/ERF

O domínio AP2/ERF exibe aproximadamente 60-70 aminoácidos sendo altamente conservado e específico para as plantas superiores. (Riechmann, 2000). Os aminoácidos dos domínios ERF, DREB, RAV e APETALA2 iniciam pelo motivo YRG (elemento de YRG) terminando com NFP (ou similar). A recente resolução da estrutura cristalina do domínio ERF tem levado a uma melhor compreensão do seu funcionamento (Allen et al., 1998). Ela é composta de três folhas β antiparalelas na região N- terminal montada em torno de uma α hélice no domínio C-terminal. O domínio ERF liga-se ao DNA através das três folhas β interagindo com 9 bases consecutivas no DNA. Os ERF foram isolados pela sua capacidade de ligar-se a caixa de GCC, um elemento regulador cis de 11 pares de

bases (TAAGAGCCGCC) encontrado em promotores de inúmeras proteínas PR de tabaco (Ohme-Takagi & Shinshi, 1995). Este elemento conservado é responsável pela indução por etileno em um amplo espectro de genes de defesa (Zhou et al. 1997, Fujimoto et al., 2000).

2.6- Diferenças entre ERF e DREB

Nakano et al. (2006) classificaram os genes da família em doze grupos, sendo que alguns divididos em subgrupos. Os domínios conservados entre os membros do mesmo grupo, ou mesmo entre os subgrupos, podem potencialmente desempenhar um papel semelhante. Vários genes *ERF* que regulam a resposta a estresses abióticos / bióticos pertencem a dois grandes grupos: Grupo III (dividido em 4 subgrupos) que contém membros que respondem ao frio, ao estresse hídrico e estresse salino e os membros do grupo IX (dividido em 3 subgrupos) que respondem à infecção por patógenos. Sakuma et al. (2002) concluíram que as proteínas ERF que se ligam ao motivo GCC (AGCCGCC) estão envolvidas na regulação de genes de resposta ao estresse biótico e proteínas DREB ligam-se ao motivo DRE (A / GCCGAC), regulando genes de resposta ao estresse abiótico. Embora contendo uma sequência muito próxima do domínio do ERF, fatores DREB (Dehydration Responsive Element Binding Factors) caracterizados em *Arabidopsis* têm propriedades semelhantes, tendo à particularidade de estar envolvido em resposta a estresses diferentes, sendo dependentes do etileno para a maioria dos ERF e independente desse hormônio para os DREBs (figura 5). Eles podem definir dois elementos cis perto da caixa GCC envolvidos na resposta ao frio e estresse hídrico: o elemento DRE (5'-TACCGACAT-3') encontrado no promotor do gene de resposta à seca (Yamaguchi-Shinozaki e Shinozaki, 1994), e paralelamente o elemento CRT (5'-TGCCGAC-3') que foi identificado no promotor do gene de resposta ao frio (Baker et al., 1994). Várias proteínas que codificam fatores de transcrição DREB foram identificadas em

Arabidopsis e outras plantas, que especificamente ligam-se com a sequência DRE/CRT. Um elemento A/GCCGAC, nomeado DRE leva essas duas seqüências é determinada pelas mesmas proteínas DREB1A/B/C (= CBF3/1/2) e DREB2A/B (Stockinger et al., 1997, Liu et al. 1998, Sakuma et al., 2002).

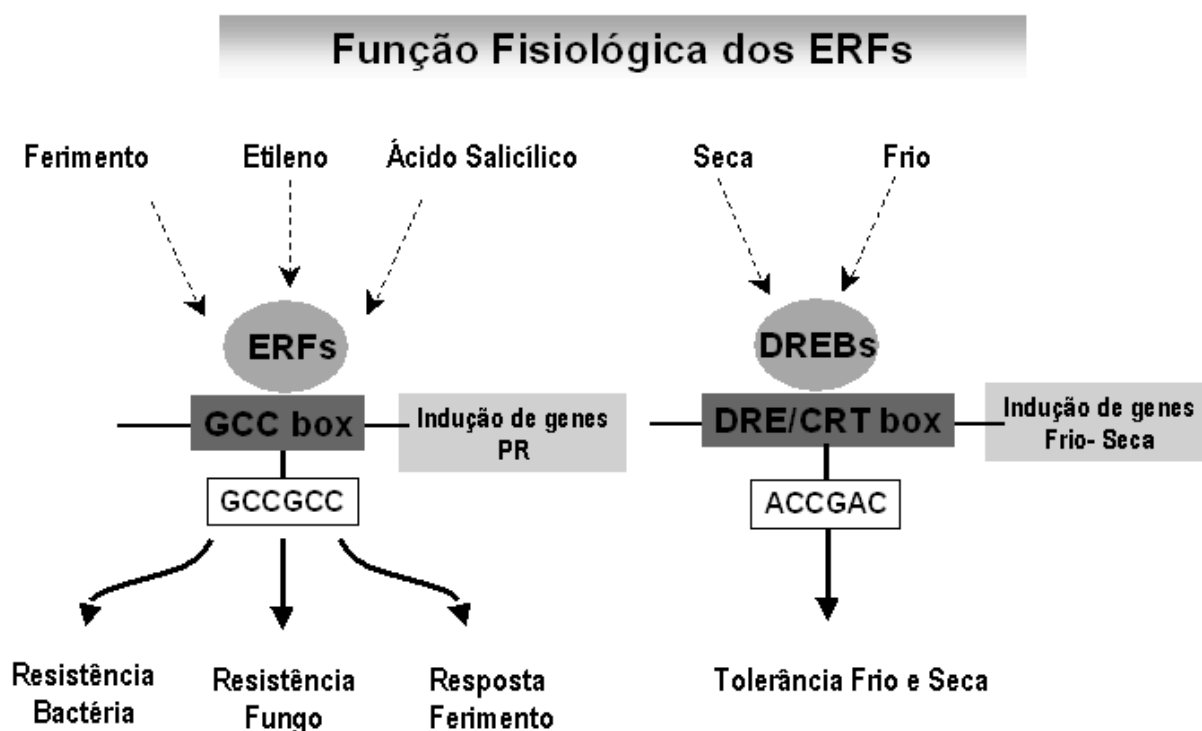


Figura 5. Função fisiológica dos ERF em resposta aos diferentes estresses. (Adaptado de Singh et al., 2002)

Apesar de numerosos estudos descritos sobre alguns ERFs utilizando estratégia específica, há poucos dados sobre o funcionamento específico desta grande família de fatores de transcrição. Poucas comparações entre espécies têm sido realizadas, não estando claro se os resultados obtidos a partir de uma via de sinalização ou processos fisiológicos são conservados em outras espécies (Gu et al., 2002). Lesões, estresse hídrico, osmótico, calor e os ataques de alguns patógenos induzem a produção de etileno e a expressão de proteínas PR (Chang e Shockey, 1999; Kiziz et al., 2001). Os ERF estudados responderam a esses estímulos de formas diferentes, mas todos respondem ao etileno (Ohme-Takagi e Shinshi, 1995; Gu et al., 2000).

A super expressão desses fatores de transcrição ERF tem levado a múltiplos fenótipos relacionados aos mecanismos de aumento da resistência ao estresse. Por outro lado, plantas “antisenso” nunca produziram efeitos, provavelmente devido à redundância de função nesta grande família. A primeira transformação genética super expressando o gene *ERF1* foi realizada em *Arabidopsis* e tabaco. Plantas de *Arabidopsis* tiveram um fenótipo de resposta constitutiva ao etileno, com uma expressão permanente de genes de defesa (Solano et al., 1998). Os trabalhos mais recentes têm demonstrado que essas plantas apresentam maior resistência a certos fungos (Berrocal-Lobo et al., 2002). Os primeiros ERF caracterizados em tabaco foram os *ERF1* a 4, os quais se ligam a caixa GCC presentes no promotor de muitas proteínas PR (Ohme-Takagi e Shinshi, 1995). Esta caixa, chamada Etileno Responsive Element (ERE) em resposta ao etileno é responsável pela expressão de proteínas PR (Sato et al. 1989). A super expressão do ERF *Tsi* em tabaco confere um fenótipo de maior resistência ao estresse hídrico, osmótico e a patógenos, com ativação constitutiva de várias proteínas PR (Park et al., 2001). No entanto, todas as proteínas PR ativadas não são idênticas, sugerindo que tenham um efeito diferente para os promotores de cada uma.

2.7- Etileno e mecanismos de defesa contra patógenos

O etileno desempenha um papel importante na regulação da resposta das plantas a estresses bióticos em combinação com outras moléculas, como o ácido salicílico e metil jasmonato (Thomma et al., 2001). O papel do ácido salicílico tem sido amplamente estudado em resposta a patógenos e na resistência sistêmica adquirida (SAR) (Lawton et al. 1995). Mutantes que não sintetizam ácido salicílico apresentaram um fenótipo de maior sensibilidade aos agentes patogênicos, e um enfraquecimento do SAR. No entanto, esses mutantes permanecem resistentes a alguns desses agentes patogênicos, sugerindo que o ácido salicílico abrange apenas um grupo de respostas ao ataque de patógenos. Existe

certamente uma sinalização independente do ácido salicílico, que envolve tanto o etileno como o metil jasmonato (Penninckx et al. 1996; Knoester et al. 1998). Esta sugestão foi especialmente evidenciada com o estudo da expressão do gene antifúngico *PDF1.2* (codifica peptídio defensina) em *Arabidopsis* que é induzido localmente e sistemicamente pelo jasmonato, independente do ácido salicílico (Penninckx et al. 1996). O etileno também está envolvido na indução de resistência sistêmica (SAR), um mecanismo independente do ácido salicílico, que permite que a planta induza resistência a patógenos em tecidos aéreos, em resposta à presença de certas bactérias patogênicas de solo (Pieterse et al., 1998). O papel destes três hormônios tem sido extensivamente estudado e mostram que o ácido salicílico está envolvido nos mecanismos de resistência a fungos biotrófico (que crescem em tecidos vivos) e certas bactérias, enquanto jasmonato metílico e o etileno participam de resposta envolvendo fungos necrofitófagos (crescem na necrose de tecidos vegetais), como *Botrytis cinerea* (Thomma et al. 1998).

3 REFERÊNCIAS BIBLIOGRÁFICAS

Allen M.D; 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, 5485–5496.

Baker S.S; Wilhelm K.S; Thomashow M.F (1994): The 5'-region of *Arabidopsis thaliana* cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant Mol Biol., 24 (5), 701-713.

Berrocal-Lobo M; Molina A; Solano R (2002) Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in *Arabidopsis* confers resistance to several necrotrophic fungi. Plant J 29: 23-32.

Chang C; Shockey JA (1999) The Ethylene-Response Pathway: Signal Perception to Gene Regulation Curr. Opin. Plant Biol. 2 352-358.

Chao Q; Rothenberg M; Solano R; Roman G; Terzaghi W; Ecker J.R (1997) Activation of the ethylene gas response pathway in *Arabidopsis* by the nuclear protein ethylene-insensitive3 and related proteins. *Cell* 89:1133-1144

Davuluri R.V; Sun H; Palaniswamy S.K; Matthews N; Molina C; Kurtz M; Grotewold E (2003) AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinformatics* 4: 25

Fujimoto S.Y; Ohta M; Usui A; Shinshi H; Ohme-Takagi M (2000) *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* 12: 393-404

Gane R (1934) Production of ethylene by some ripening fruits. *Nature*, n. 134, p. 1008

Giraudat J; Hauge B.M; Valon C; Smalle J; Parcy F; Goodman H.M (1992): Isolation of the *Arabidopsis ABI3* gene by positional cloning. *Plant Cell*, 4, 1251-1261.

Gu Y.Q; Yang C; Thara V.K; Zhou J; Martin G.B (2000) Pti4 is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. *Plant Cell* 12: 771–786

Herr W; Sturm R.A; Clerc R.G; Corcoran L.M; Baltimore D; Sharp P.A; Ingraham H.A; Rosenfeld M.G; Finney M; Ruvkun G. (1988): The POU do-main: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.*, 2 (12A), 1513-1516.

Kagaya Y; Ohmiya K; Hattori T (1999): RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Research*, 27 (2), 470-478.

Kizis D; Lumberras V; Pages M (2001): Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. *FEBS Lett.* 8, 498 (2-3), 187-189.

Knoester M; Van Loon LC; Van den Heuvel J; Hennig J Bol JF. (1998). Ethylene-insensitive tobacco lacks nonhost resistance against soil-borne fungi. *Proc Natl Acad Sci USA* 95:1933–1937.

Lawton K; Weymann K; Friedrich L; Vernooij B; Uknes S;(1995). Systemic acquired resistance in *Arabidopsis* requires salicylic acid but not ethylene. *Mol Plant Microbe*

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

McCourt P.(1999) Genetic analysis of hormone signaling. *Annual Review of Plant Physiology*, v.50, p. 219 – 243.

Nakano T; Suzukî K; FujimuraT; Shinshi H (2006) Genome-wide analysis of the ERF gene family in *Arabidopsis* and *Rice*. *Plant Physiology*, 140:411-432

Ohme-Takagi M; Shinshi H (1995): Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell.*, 7 (2), 173-182.

Park J.M; Park C.J; Lee S.B; Ham B.K; Shin R; Pae K.H (2001): Overexpression of the tobacco Tsi1 gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. *Plant Cell*, 13 (5), 1035-46.

Penninckx I.A; Eggermont K; Terras F.R; Thomma B.P; De Samb Penninckx I.A; Eggermont K; Terras F.R; Thomma BP; De Samblanx G.W and others. (1996) Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway. *Plant Cell* 8:2309–2323.

Pieterse C.M; Van Wees S.C; Van Pelt JA; Knoester M; Laan R. (1998). A novel signaling pathway controlling induced systemic resistance in *Arabidopsis*. *Plant Cell* 10:1571–1580

Riechmann J.L; Heard J; Martin G; Reuber L; Jiang C.Z; Keddie J; Adam L; Pineda O; Ratcliffe O.J; Samaha R.R; Creelman R; Pilgrim M; Broun P; Zhang J.Z; Ghandehari D; Sherman B. K; Yu G.L (2000): *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, 290, 2105–2110.

Rodriguez F.I; Esch J.J; Hall A.E; Binder B.M; Schaller G.E; Bleecker A.B (1999) A copper cofactor for the ethylene receptor ETR1 from *Arabidopsis*. *Science* 283:996-998

Sakuma Y; Liu Q; Dubouzet J.G; Abe H; Shinozaki K; Yamaguchi I-Shino-Zaki 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.* 25, 290 (3), 998-1009.

Sato T; Theologis A. (1989). Cloning the mRNA encoding 1-aminocyclopropano 1-carboxylate synthase, the key enzyme for ethylene biosynthesis in plants. *Proc Natl Acad Sci USA* 86:6621–6625.

Singh K.B; Foley R; Onate-Sanchez L.(2002) Transcription factors in plant defense and stress response. *Curr Opin Plant Physiol* 116:1145 – 1150

Solano R; Stepanova A; Chao Q; Ecker J.R. (1998): Nuclear events in ethylene signaling: A transcriptional cascade mediated by ethylene-insensitive3 and ETHYLENE-RESPONSIVE-FACTOR1. *Genes Dev.*, 12, 3703–3714.

Stockinger E.J; Gilmour S.J; Thomashow M.F (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. *Proc Natl Acad Sci USA* 4, 94 (3), 1035-1040.

Taiz L. & Zeiger E. (2004) *Fisiologia Vegetal*. 3 ed, Porto Alegre, Artmed, 719p.

Thomma B.P; Penninckx I.A; Broekaert W.F; Cammue B.P (2001) The complexity of disease signaling in *Arabidopsis*. *Curr Opin Immunol* 13:63–68.

Thomma B; Eggermont K; Penninckx I; Mauch-Mani B; Vogel Thomma B; Eggermont K; Penninckx I; Mauch-Mani B; Vogelsang R; and others. (1998) Separate jasmonate-dependent and salicylate-dependent defense-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. *Proc Natl Acad Sci USA* 95:15107–15111.

Wang A; Tan D; Takahashi A; Li T.Z; Harada T (2007) *MdERFs*, two ethylene-response factors involved in apple fruit ripening. *J Exp Bot* 58:3743-3748

Yang S.F; Hoffman N.E (1984) Ethylene biosynthesis and its regulation in higher plants. *Annual Review of Plant Physiology* 35:155–189

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 (2), 251–264.

Zhang J.Z. (2003): Overexpression analysis of plant transcription factors. *Current Opinion in Plant Biology*, 6, 1-11.

Zhou J; Tang X; Martin G.B (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. *EMBO J* 16:3207–3218

Capítulo I

Artigo enviado para a revista Genetics and Molecular Biology

In silico* analysis of the Ethylene Response Factor (ERF) gene family in *Malus

César Luís Girardi^{1*}, Vera Quecini¹, César Valmor Rombaldi², Joceani Dal Cero², Paula M. Nobile³, François Laurens³, Julien Pirrello⁴, Mondher Bouzayen⁴

¹Embrapa Uva e Vinho - Caixa Postal 130, 95700-000, Bento Gonçalves, RS - Brazil

²Universidade Federal de Pelotas, Faculdade de Agronomia Eliseu Maciel, Departamento de Ciência e Tecnologia Agroindustrial –s/nº caixa postal 354, 96010-900, Pelotas, RS - Brazil

³UMR 1259 Génétique et Horticulture (GenHort) Angers, France

⁴UMR 990, Génomique et Biotechnologie des Fruits, INP, ENSAT, INRA, F-31320 Castanet Tolosan, France

* Corresponding author: girardi@cnpuv.embrapa.br

ABSTRACT

We have investigated *Malus* genome databases of expressed sequence tags (EST) in order to identify genes coding for functionally characterized proteins sharing sequence similarity to ethylene response factors. Approximately 300,000 EST sequences from 120 libraries were investigated and 66 orthologs of ERF proteins were identified. In *Vitis* and *Populus*, perennial species with complete genome sequence, 197 and 57 predicted ERF orthologs were identified, respectively; thus, suggesting that our analyses have uncovered a significant number of apple ERFs. We found evidences of differential sequence evolution in the ERF families from model species and *Malus*. The presence of conserved motifs distinct from the AP2/ERF DNA binding sequence in apple ERFs allowed clustering in ten groups. Evidence from model species suggests that distinct groups perform distinct ethylene-mediated functions in plant physiology and metabolism. Prediction analysis of apple proteins uncovered further differences between *Malus* versus *Arabidopsis* and *Populus* ERF proteins including folding status, isoelectric point and sub-cellular localization. *In silico* transcriptional profiling was performed with a sub-set of *Malus* putative proteins and suggests the involvement of apple ERFs in distinct ethylene-mediated biological functions, such as developmental control and stress responses. Our results demonstrate the power of comparative genomics between model systems and economically important crop species to help elucidating the molecular mechanisms underlying plant physiology and metabolism.

Keywords: apple, climacteric fruit, ethylene, fruit ripening, physiological conditions

1 INTRODUCTION

Hormones are responsible for the control of plant growth and development and are also involved in mediating responses to several biotic and abiotic stresses. Besides its important role in regulating distinct physiological processes; such as seed germination, cell elongation, wounding, ripening, senescence and abscission, ethylene also regulates the plant responses to biotic and abiotic stresses (Abeles et al. 1992; Chen et al. 2005). Ethylene exerts its action through complex regulatory steps of its biosynthesis, perception and signal transduction, leading to dramatic changes in gene expression (Chang and Bleecker 2004; Chen et al. 2005; Kendrick and Chang 2008). The promoter sequences of several genes induced by ethylene were found to contain a *cis*-regulatory element known as the Ethylene-Responsive Element (ERE) (Broglie et al. 1989). Sequence analysis of various ERE regions identified a short motif, rich in G/C nucleotides, labeled the GCC-box, essential for ethylene-mediated responses. The ERE motif is recognized by a family of transcription factors: the ERE binding factors (ERF) (Fujimoto et al. 2000).

The ERF is a large gene family of transcription factors that constitute a sub-group of the APETALA2 (AP2)/ERF superfamily, which also contains the AP2 and RAV families (Riechmann et al. 2000). The superfamily is defined by the presence of the AP2/ERF domain, a conserved DNA-binding sequence consisting of approximately 60 to 70 amino acids (Weigel 1995). The AP2 family consists in proteins with two AP2/ERF domains, whereas the ERF family is constituted of proteins containing a single AP2/ ERF domain, and the RAV family consists of proteins characterized by the presence of a B3 domain, a DNA-binding domain conserved in other plant-specific transcription factors, in addition to a single AP2/ERF domain. Previous studies have proposed a sub-division of the ERF family into two major sub-families; the ERF subfamily and the CBF/ DREB subfamily (Sakuma et

al. 2002). More recent data suggest the existence of ten groups containing distinct conserved motifs within the family rather than distinct sub-families (Nakano et al. 2006).

The role of the ERF proteins in stress-related, important agronomical responses and the accumulation of genomic and expressed sequence data have prompted in-depth bioinformatic investigation of the ERF sub-family in several plant species, including *Arabidopsis*, rice, *Populus* and soybean (Nakano et al. 2006; Zhang et al. 2008; Zhuang et al. 2008). In fruit species, ERFs are associated to flavor biosynthesis and texture modification during ripening (Alba et al. 2005; DaSilva et al. 2005; Ziliotto et al. 2008). In apple (*Malus x domestica* Borkh. cv. Golden Delicious), two ERF transcripts, *MdERF1* and *MdERF2*, demonstrated to be predominantly and exclusively expressed in ripening fruits, respectively, were isolated from ripening fruit (Wang et al. 2007). Similarly, Newcomb et al. (2006) identified expressed sequences of ethylene response factors in the tag collection.

Currently, over 400,000 ESTs are available for Rosaceae species via the National Center for Biotechnology Information (NCBI) EST and the Genome Database for Rosaceae (GDR) (www.rosaceae.org) databases. Additionally, about 150,000 5'-end ESTs from apple have been deposited in public access databases by HortResearch in New Zealand (Newcomb et al. 2006). The combined number of publicly available apple ESTs adds up to over 260,000 sequences, resulting in a 17,000 unigene set. The role of ERF proteins in the regulation of the expression of genes responsible for changes in fruit quality, texture and physiological post-harvest disorders has prompted us to investigate the family in *Malus* using bioinformatic tools and comparative genomic approaches. The preliminary information obtained by *in silico* analyses will be further validated by *in vivo* gene expression studies.

2 MATERIAL AND METHODS

2.1-Database Searches and Alignments

Orthologs of *MdERF1* and *MdERF2* and orthologs of *Arabidopsis thaliana* ethylene response factors (ERF) were identified in BLAST searches (Altschult et al. 1997) against open-access *Malus* EST databases at The Institute of Genome Research (TIGR), The Genome Database for Rosaceae (GDR), GenBank and The Plant Transcription Factor Databases, comprising approximately 300,000 ESTs obtained from sequencing of 120 specific libraries. Data validation was performed by tBLASTx and tBLASTn searches of the retrieved sequence against GenBank database. Sequences failing to retrieve the original sequence used to query the database were eliminated from the projects. The resulting alignments were filtered by a threshold e-value of $1e-25$ and the hits were further analyzed. Validated sequences were translated and protein (deduced amino acid) alignments were performed using ClustalX (Thompson et al. 1997). When necessary, alignments were manually adjusted using Lasergene MegAlign (DNASTAR, Madison, WI, USA).

2.2-Phylogenetic Analysis

The functionality of *Malus* genes in comparison to their *Arabidopsis* counterparts and to functionally characterized apple *ERFs* was assessed by genetic distance and phylogenetic studies. Phylogenetic analyses were performed using distance and parsimony methods in the software PAUP* 4.0b10 (<http://paup.csit.fsu.edu/>), using the software default parameters. Resampling bootstrap trees containing 1000 random samples were constructed using PSIGNFIT software (<http://www.bootstrap-software.org/>). Modular functional domains were employed for genetic distance studies for genes previously characterized as having divergent regions and conserved blocks.

2.3-Motif analysis and *in silico* characterization

Conserved motifs were further investigated by multiple alignment analyses using ClustalX and the MEME version 4.0 suite (Bailey and Elkan 1994). The presence and sequence conservation of recognizable functional domains was studied employing protein analysis and gene function tools from databases (European Bioinformatics Institute-European Molecular Biology Laboratory –EMBL-EBI; Expert Protein Analysis System – ExPaSy from Swiss Institute of Bioinformatics - SIB; Gene Ontology database – GO; Protein Families database – Pfam).

2. 4- Prediction of cellular and molecular parameters

The physical characteristics; isoelectric point (IP) and predicted molecular weight, of apple ERF sequences were calculated from the deduced amino acid using default parameter of the Lasergene MegAlign software (DNASTAR, Madison, WI, USA). The proteins folding state was predicted by the FoldIndex program (<http://bioportal.weizmann.ac.il/fldbin/findex>). The prediction of the sub-cellular compartment location of the apple ERF-like proteins was performed using the plant algorithm of the software WoLF PSORT (<http://wolfsort.org/>) (Horton et al. 2007).

2. 5- *In silico* gene expression analysis

Qualitative gene expression profiling was performed by *in silico* analyses of the *Malus* EST databases using virtual northern blot analyses. The gene of interest was used in queries against reference sequence databases, generating an alignment of the input gene to its orthologs. The resulting alignment was used to find sequences in the entire mRNA input that are specific to the gene (probe). The resulting alignments were collectively used to query the EST database again using BLAST. This heuristic was critical to avoid false-positives, or ESTs from a paralog of the input gene rather than the gene itself. The

identity numbers of the ESTs matching the probes were recovered and the databases were used to find the names of the libraries from which those ESTs were derived. The frequency of reads of each EST contig in a given library was calculated and normalized according to the total number of reads from the investigated library and the total number of reads in all libraries. A correlation matrix between EST contigs and libraries was then generated and gene expression patterns among ESTs and libraries were obtained by hierarchical clustering based on Spearman Rank correlation matrix using Cluster v.2.11 software (Eisen et al. 1998), by substituting the clusters by their average expression pattern. Graphic outputs were generated using Tree View v.1.6 software (<http://rana.lbl.gov/EisenSoftware.htm>) and presented in grayscale.

3 RESULTS

3.1- The family of ethylene response factors in apple

We have employed bioinformatic tools to identify the ERF family genes in open-access *Malus* sequence databases. We have identified 15 apple sequences sharing sequence similarity to proteins containing two AP2/ERF domains that were attributed to the *Malus* AP2 family (Table 1, Supplementary Table I). Four of those exhibit a single, complete AP2/ERF domain distinct from the ERF type, being instead more closely related to AP2 type. We have identified seven ESTs whose deduced amino acid sequences share significant functional domain conservation to members of the RAV family (Table 1, Supplementary Table I). Sixty-six sequences from *Malus* databases were clustered in the ERF family (Table 1, Table 2, Supplementary Table I). Thus, to date, the *Malus* AP2/ERF superfamily consists of 80 genes, clustered into three families; the ERF family (66 genes), the AP2 family (15) and the RAV family (seven genes) (Table 1, Supplementary Table I).

Table 1. Summary of the AP2/ERF superfamily in *Arabidopsis thaliana* and *Malus sp.*

Classification	Group	<i>Arabidopsis thaliana</i>		<i>Malus sp.</i>	
		Number	Reference	Number	Reference
AP2 family		18		15	This study
	double AP2/ERF domain	14	Sakuma et al. (2002), Nakano et al. (2006)	11	This study
	single AP2/ERF domain	4	Nakano et al. (2006)	4	This study
At4g13040		1	Sakuma et al. (2002), Nakano et al. (2006)	0	This study
RAV family		6	Sakuma et al. (2002), Nakano et al. (2006)	7	This study
ERF family		122		66	This study
	groups I to IV	57	Nakano et al. (2006)	19	This study
	group V to X	58	Nakano et al. (2006)	45	This study
	groups VI-L and Xb-L	7	Nakano et al. (2006)	2	This study
Total		147		80	

No *Malus* orthologs of the soloist *Arabidopsis* gene, At4g13040, were present in the investigated databases, although, the lack of full genome data prevents further conclusions about the existence of this member in apple. The identity of the ERF proteins in apple to their *Arabidopsis* counterparts ranged from 23 to 70%, being higher for members of the WIN/SHN1 family, where it ranged from 53 to 70% (Table 2). The smallest degree of sequence conservation in comparison to *Arabidopsis* was observed in the *Malus* ERFs belonging to groups VI and X (Table 2).

Table 2. *Malus* ESTs with homology to genes of the ERF family in *Arabidopsis thaliana*.

<i>Arabidopsis thaliana</i>		<i>Malus</i> sequence			Protein motifs ^c and	References
Name	Gene	Identifier (coverage %) ^a	% ^b	e value	biological process	
At1g78080	RAP2.4	PTMx00041.1 (90)	57.1	9e-107	CMI-1 to CMI-7	Lin et al. (2008)
		PTMx00050.1 (FL)	58.2	2e-109	drought tolerance, ethylene response, light response	
		PTMx00057.1 (85)	55.2	3e-104		
		PTMx00088.1 (FL)	48.0	2e-97		
		PTMx00089.1 (98)	45.5	1e-95		
At1g46768	RAP2.1	PTMx00026.1 (FL)	52.9	3e-108	CMII-1 to CMII-3	Nakano et al. (2006)
		PTMx00032.1 (87)	43.9	5e-94	unknown	
		PTMx00054.1 (FL)	54.4	1e-103		
At4g25490	CBF1/DREB1B	A5YRQ8 (FL)	55.2	1e-105	CMII-1 to CMIII-7	Gilmour et al. (2000); Haake et al. (2002)
		Q2LMD4 (FL)	54.9	5e-103	drought tolerance, cold tolerance	
		PTMx00030.1 (FL)	54.8	2e-105		
		PTMx00036.1 (98)	32.5	1e-81		
		PTMx00045.1 (FL)	52.5	8e-102		
		PTMx00059.1 (95)	50.8	9e-100		
		PTMx00060.1 (82)	45.1	1e-80		
		PTMx00084.1 (FL)	56.9	5e-96		
		PTMx000100.1 (85)	47.8	7e-98		

At4g40220	ABI4	PTMx00037.1 (90)	53.4	2e-100	CMIV-1 to CMIV-3	Liu et al. (1998); Nakashima et al. (2000); Song et al. (2005)
		PTMx00086.1 (97)	62.6	1e-119	drought tolerance, ABA responses	
At1g15360	WIN/SHN1	PTMx00031.1 (FL)	57.9	3e-108	CMV-1 to CMV-4	Ahroni et al. (2004); Broan et al. (2004)
		PTMx00058.1 (90)	52.6	7e-99	drought tolerance, wax biosynthesis	
		PTMx00065.1 (FL)	69.8	6e-125		
At1g25470	AtERF#116	PTMx00077.1 (80)	29.7	7e-79	CMVI-L-1 to CMVI-L-5	Zhou et al. (1997)
		PTMx00078.1 (95)	28.6	4e-68	biotic and abiotic stress responses	Park et al. (2001)
At3g16770	AtEBP/RAP2.3	A1LM1 (FL)	39.8	2e-85	CMVII-1 to CMVII-8	Büttner and Singh (1997)
		CO576318 (84)	48.9	1e-99	unknown	
		CV997565 (87)	42.7	8e-89		
		EB114677 (80)	35.6	9e-83		
		EB120281 (80)	31.3	1e-80		
		EB135117 (90)	45.7	5e-96		
		EB151494 (95)	51.1	7e-98		
		EB176565 (95)	48.2	1e-95		
		TC4497 (FL)	43.0	2e-94		
		PTMx00020.1 (FL)	51.1	3e-99		
		PTMx00029.1 (FL)	56.1	7e-99		
		PTMx00033.1 (FL)	56.7	1e-104		
		PTMx00067.1 (FL)	52.0	8e-99		

At1g12980	ESR1/DRN	PTMx00068.1 (FL)	43.0	7e-93		
		TC6242 (FL)	45.2	6e-94	CMVIII-1 to CMVIII-3	McGrath et al. (2005)
		PTMx00022.1 (FL)	50.0	7e-96	ABA, ethylene and jasmonate responses	Yang et al. (2005)
		PTMx00023.1 (FL)	45.2	4e-94	organ differentiation and development	Song et al. (2005)
		PTMx00024.1 (80)	39.2	9e-91		Kirch et al. (2003)
		PTMx00061.1 (82)	59.9	1e-112		Banno et al. (2001)
		PTMx00062.1 (FL)	38.6	8e-91		van der Graaff et al. (2000)
		PTMx00063.1 (FL)	41.7	9e-92		
		PTMx00064.1 (95)	47.5	1e-97		
		PTMx00070.1 (97)	54.2	5e-98		
At4g17500	AtERF1	PTMx00076.1 (90)	32.5	7e-75		
		A1LM2 (FL)	42.0	1e-93	CMIX-1 to CMIX-6	Berrocal-Lobo et al. (2002)
		CO865301 (98)	47.3	5e-96	pathogen responses	Gu et al.(2000)
		TC14438 (FL)	42.0	9e-92	ethylene, salycilic acid and jasmonate responses	Gu et al. (2002)
		PTMx00021.1 (85)	41.2	7e-91		Oñate-Sánchez and Singh (2002)
		PTMx00044.1 (FL)	41.5	1e-91		
		PTMx00073.1 (99)	39.0	3e-88		
		PTMx00095.1 (FL)	54.7	6e-98		
At5g64750	ABR1	PTMx00096.1 (95)	34.8	9e-87		
		PTMx00027.1 (92)	38.1	7e-90	CMX-1 to CMX-3	Pandey et al. (2005)
		PTMx00028.1 (98)	42.0	4e-92	unknown	

PTMx00038.1 (95)	44.9	3e-95	repressor of ABA responses
PTMx00048.1 (82)	33.0	2e-85	
PTMx00049.1 (FL)	27.3	2e-32	
PTMx00056.1 (80)	40.3	9e-90	
PTMx00066.1 (95)	48.1	1e-98	

Multiple sequence alignments of the deduced amino acid sequences of the AP2/ERF domains were performed to establish the phylogenetic relationships between the genes in the *Malus* ERF family. The conserved residues Gly-4, Arg-6, Glu-16, Trp-28, Leu-29, Gly-30, and Ala-38 are present in all 66 ERF proteins in apple (Figure 1 and Figure 2). The C-terminal regions of the AP2/ERF domains of two proteins (MdERF#41, MdERF#42) are divergent from the consensus sequence (Figure 1).



Figure 1 – Sequence alignment and structure prediction of the AP2/ERF domain of proteins from the ERF sub-family in *Malus* sp.. *Malus* deduced amino acid sequences were aligned with ClustalX. Black and light gray shading indicate identical and conserved amino acid residues in the Clustal Consensus. The divergent sequences are highlighted in red boxes. The black bar and arrows represent predicted α -helix and β -sheet regions within the AP2/ERF domain, respectively.

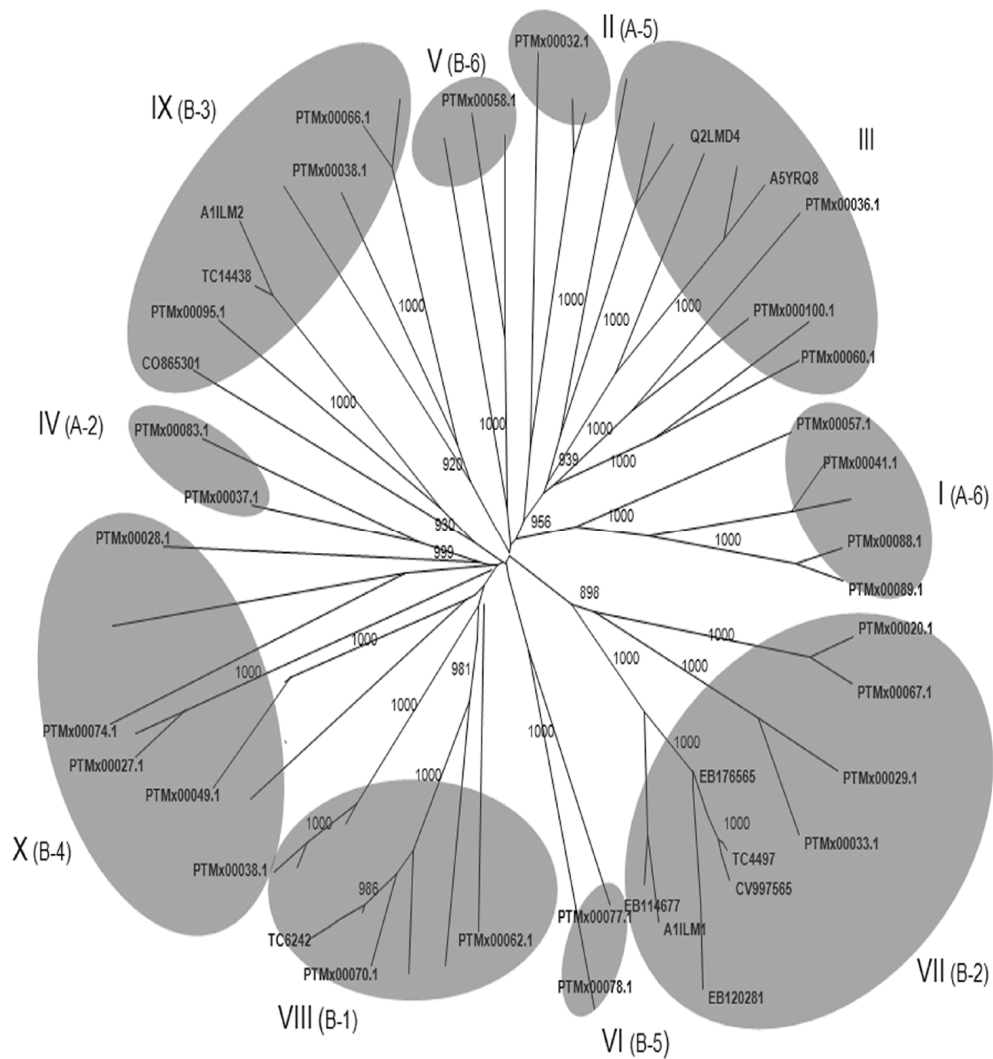


Figure 2 – AP2/ERF-type transcription factors from the ERF sub-family in *Malus*. Neighbor-joining tree for *Malus* deduced amino acid and *Arabidopsis* full length sequences were aligned with ClustalX. Bootstrap values are indicated above each branch. Cluster nomenclature according to Nakano et al. (2006) is exhibited as roman numerals and classification by Sakuma et al. (2002) is indicated in parentheses.

3.2- Conserved motifs outside the AP2/ERF domain

In order to associate the unknown putative apple ERFs to biological functions, we have investigated additional conserved motifs within the deduced amino acid sequences employing multiple sequence analysis and the MEME v.4.0 suite (Bailey and Elkan 1994). The vast majority of the ERF-like sequences identified in apple shared one or more motifs outside the AP2/ERF domain with their *Arabidopsis* counterparts (Figure 3, Table 2, Supplementary Tables III and IV). A great number of the identified *Malus* sequences (21.2%, 14) were clustered with *Arabidopsis* sequences from group VII, followed by group VIII and X (15.1%, 10 members in each group) and groups III and IX (13.6% and 12.1%, 9 and 8, respectively) (Figure 2, Figure 3, Table 2). The ERF groups with smaller number of members in apple were II, IV, V and VI (4.5%, 3.0%, 4.5% and 3.0% corresponding to 3, 2, 3 and 2 members, respectively). Apple sequences in group VII exhibit a *Malus*-exclusive conserved motif CMVII-9 (KALFSLKVSTRGNIN) (Figure 3). Although absent from functional site prediction tools, the novel motif is hydrophobic and consists of polar, non-charged, non-aliphatic residues. Apple ERFs in group VIII exhibit the CMVIII-1 (Figure 3), which is identical to an ERF-associated amphiphilic repression (EAR) motif.

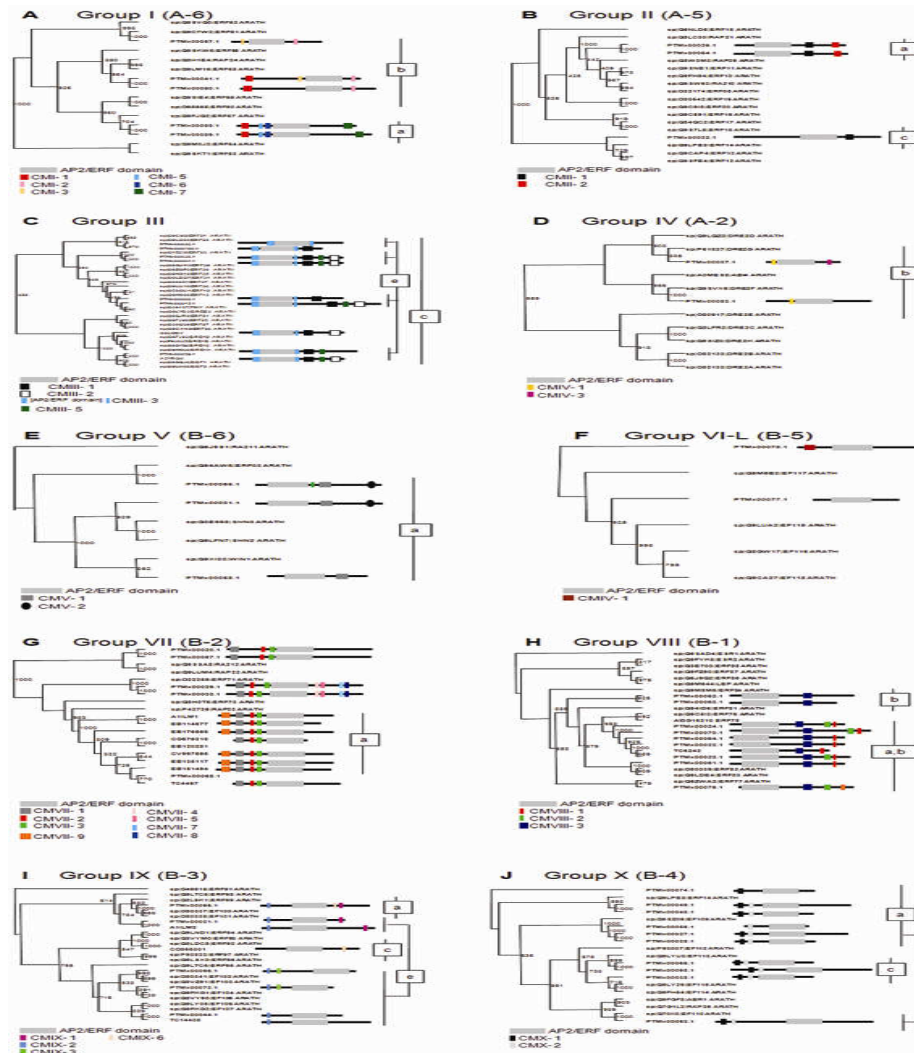


Figure 3 - Phylogenetic relationships among the *Arabidopsis* ERF genes, from group I (A), group II (B), group III (C), group IV (D), group V (E), group VI-L (F), group VII (G), group VIII (H), group IX (I) and group X (J) in *Arabidopsis* and *Malus* ERF families. Bootstrap values from 1000 replicates are shown above the branches. The phylogenetic tree and a schematic diagram of the protein structures of groups I to X are shown in A to J, respectively. Each colored box represents the AP2/ERF domain and conserved motifs, as indicated in legends below the trees. The amino acid sequences of the conserved motifs are summarized in Supplementary Table IV. Classification by Nakano et al. (2006) is shown above each tree and by Sakuma et al. (2002), indicated in parentheses.

3. 3- Cellular and molecular characteristics of apple ERF-like proteins

We have employed bioinformatic tools to determine the physical properties of the ERF-like sequences from apple, such as the molecular weight (MW), pH value of isoelectric point (pI) and folding index. The vast majority of the apple ERF-like sequences were predicted to have low molecular weight, ranging from 1.5 to 4.48 kDa and an average of 233 amino acids, going from 111 to 413 amino acids (Supplementary Table IV). In apple, most of the sequences sharing sequence similarity to *Arabidopsis* ERF proteins (74.2%, 49) were predicted to be basic to neutral, whereas only 25.8% (17) displayed predominance of acidic amino acids (Supplementary Table IV).

The folding states of ERF family proteins in apple were predicted by FoldIndex program (Prilusky et al. 2005). The majority of the identified apple sequences sharing sequence similarity to ERFs were predicted to be unfolded (59.1%, 39) under physiological conditions (Supplementary Table IV, Figure 4). The predicted sub-cellular localization of the ERF proteins identified in *Malus* was investigated using the software WoLF pSORT (Horton et al. 2007). Most of the ERF-like sequences from apple (83.3%, 55) was predicted to be exclusively nuclear or with ambiguous sub-cellular location for the nucleus and organelle or cytoplasm (Supplementary Table IV).

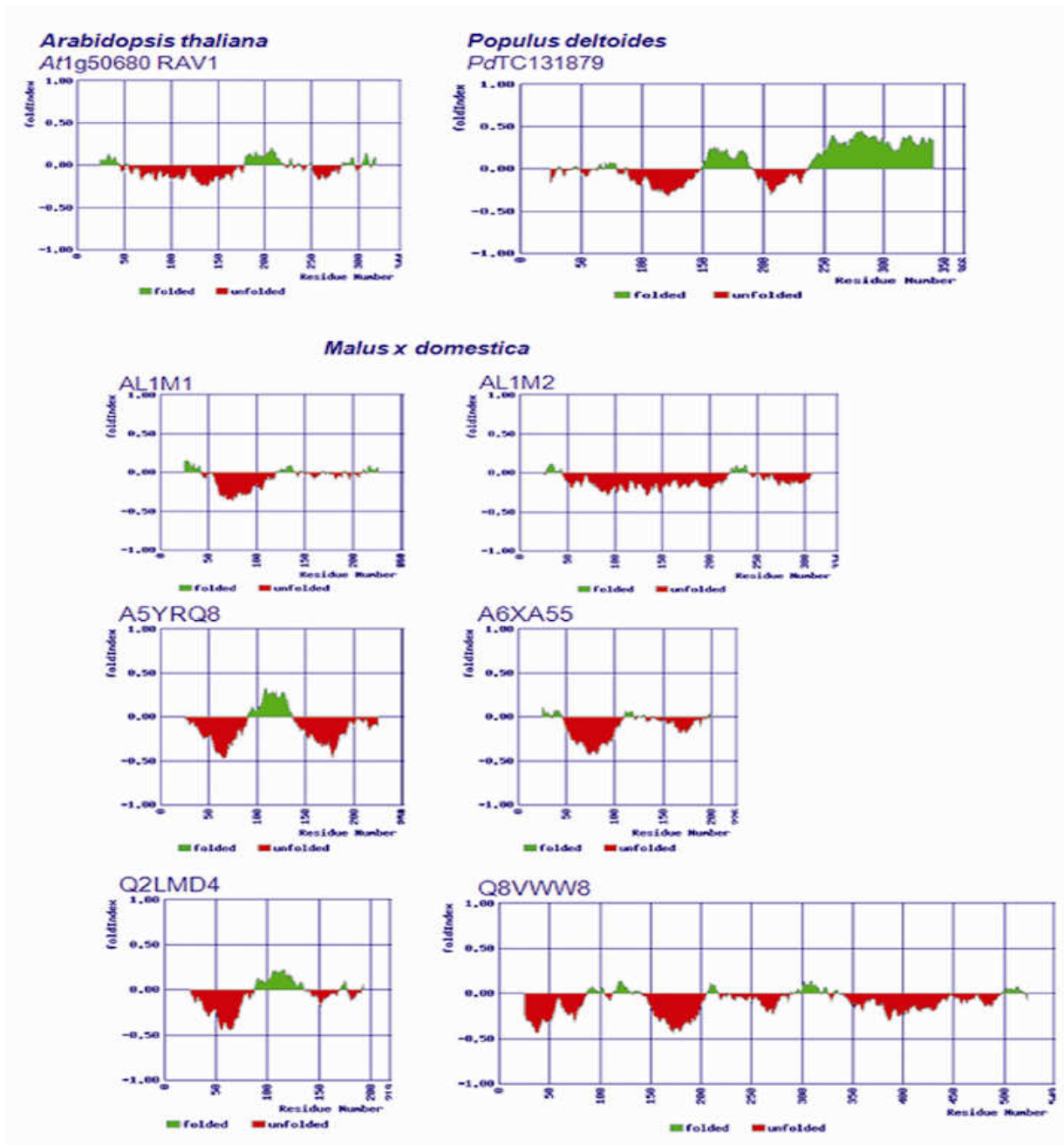


Figure 4 –Prediction of the folding state of six ERF sequences from *Malus*: A1LM1, A1LM2, A5YRQ8, A6XA55, Q2LMD4 and Q8VWW8, in comparison to *Arabidopsis thaliana* At1g50680 RAV1 and *Populus deltoides* PdTC131879. The fold index calculated by the software FoldIndex (Prilusky et al., 2005) is represented by the Y axis and the amino acids of the protein are represented by their position in the sequence in the X axis. Folded protein regions are represented in green and, unfolded in red.

3. 4- *In silico* expression profiling of ERF genes in apple

The expression profile of the *ERF* genes in *Malus* was preliminarily investigated by *in silico* virtual northern blot. The investigated sequences correspond to members of the AP2 (Q8VWW8) and ERF family (A1LM1, A1LM2, Q2LMD4 and TC6242), including members of the groups VII, IX, III and VIII, respectively.

Higher levels of expression of the *ERF* sequences from apple were observed in libraries derived from bud, root, xylem, phloem and fruit tissues (Figure 5). Differential expression libraries from pathogen-challenged tissues exhibited low frequency of transcripts derived from the investigated *ERF* sequences (Figure 5). *In silico* expression profile of Q2LMD4 was divergent from the ERF family members, although the higher induction in libraries derived from apple buds, xylem and phloem tissues was retained (Figure 5). Transcripts similar to A1LM1 were frequent in fruit-related libraries whereas those similar to A1LM2 were also found in cell cultures and bud sample libraries (Figure 5). The *Malus* AP2 sequence Q8VWW8 exhibited higher frequencies in libraries obtained from flower, vascular and fruit tissues.

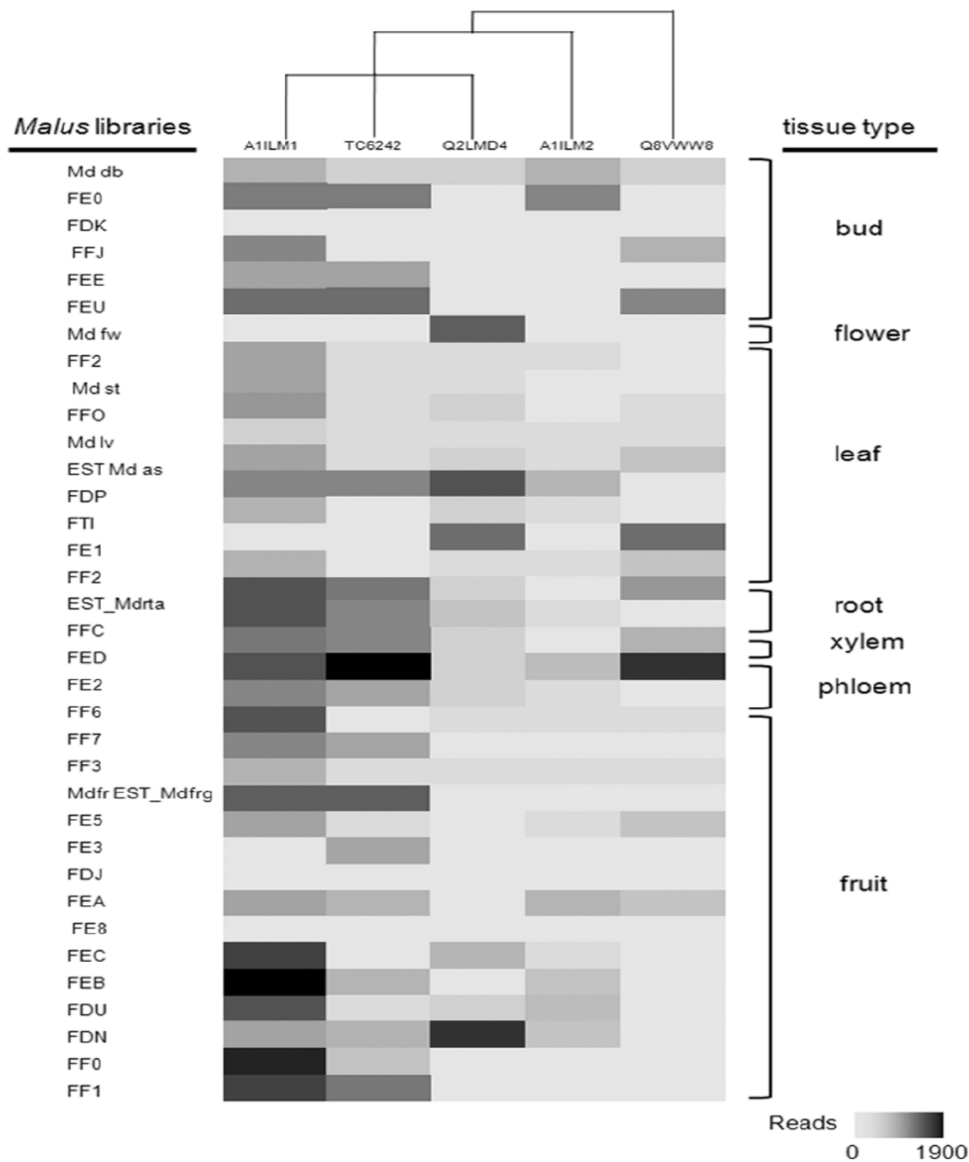


Figure 5– Expression profile of four ERF sequences from *Malus*. The normalized number of reads for the transcripts in each library is represented as a grayscale. The names of the libraries and its tissue type are listed at left and right, respectively. Detailed description of the libraries is provided elsewhere (http://compbio.dfci.harvard.edu/tgi/gi/mdgi/searching/xpress_search.html). Hierarchical clustering of the expression patterns by *k*-means using Spearman Rank correlation is represented by tree.

4 DISCUSSION

In silico analyses of open-access databases allowed us to perform an initial characterization of the *Malus* AP2/ERF superfamily; it consists of 80 genes, clustered into three families; the ERF family (66 genes), the AP2 family (15) and the RAV family (seven genes). The soloist gene, At4g13040, found in *Arabidopsis* was absent from the searched databases. In soybean, a genome-wide analysis of the AP2/ERF superfamily also failed to identify orthologs of the soloist (Zhang et al. 2008). However, in a similar study of the perennial species *Populus trichocarpa*, a single ortholog was identified (Zhuang et al. 2008). However, the lack of full genome data prevents further conclusions about the existence of this member in apple. In tomato, a large number of public available sequences display extensive homology to the ERF family from model species, among them some are involved in the defense system (Zhou et al. 1997). In *Vitis* and *Populus*, perennial species with complete genome sequence, 197 and 57 predicted ERF-like sequences, respectively, were identified (Zhuang et al., 2008). Thus, the 66 apple ERF sequences hereby identified are likely to represent a significant portion of the family. However, at this point, the existence of additional ERF-like sequences in *Malus* cannot be ruled out.

In average, the identity of the ERF proteins in apple to their *Arabidopsis* counterparts was high, with higher levels of identity found for the WIN/SHN1 family, which may be indicative of functional conservation. In *Arabidopsis*, proteins of the WIN/SHN1 have been demonstrated to be involved in wax and cuticle biosynthesis (Aharoni et al. 2004; Broun et al. 2004). The smallest degree of sequence conservation in comparison to *Arabidopsis* was observed in the *Malus* ERFs belonging to groups VI and X. Similarly, these groups have been reported to be the least conserved ones in soybean and *Populus* (Zhang et al. 2008; Zhuang et al. 2008).

As observed in *Arabidopsis*, rice, soybean and *Populus* (Nakano et al. 2006; Zhang et al. 2008; Zhuang et al. 2008), the conserved residues Gly-4, Arg-6, Glu-16, Trp-28, Leu-29, Gly-30, and Ala-38 are present in all 66 ERF proteins in apple. Two proteins (MdERF#41, MdERF#42) are divergent at the C-terminal region of the AP2/ERF domains of from the consensus sequence, as observed for two *Arabidopsis* ERFs (Nakano et al. 2006). This region corresponds to the C-terminal half of the α -helix (Allen et al. 1998), consisting of the highly conserved residues Asp-43 and Asn-57. The overall sequence of the *Malus* proteins with divergent C-terminus is related to *Arabidopsis* ERFs from group VI-like (VI-L) (Nakano et al. 2006). In soybean and *Populus*, a divergent group of sequences was also observed in group VI (Zhang et al. 2008; Zhuang et al. 2008), which suggests an association to functional divergence in evolutionary distant species.

4. 1- Conserved motifs outside the AP2/ERF domain

Transcription factors often exhibit regions outside the DNA-binding domain that are involved in regulation of the transcriptional activity, protein-protein interactions, and nuclear localization (Liu et al. 1999). These motifs, frequently associated to specific functions or regulation patterns, are characteristic of large transcription factor families in plants, such as MYB, WRKY, NAC, Dof, GATA, and GRAS (Eulgem et al. 2000; Kranz et al. 1998; Lijavetzky et al. 2003; Ooka et al. 2003; Reyes et al. 2004; Tian et al. 2004). The vast majority of the ERF-like sequences identified in apple shared one or more motifs outside the AP2/ERF domain with their *Arabidopsis* counterparts, as observed in rice, *Populus* and soybean (Nakano et al. 2006; Zhang et al. 2008; Zhuang et al. 2008). The conserved domains outside the AP2/ERF sequence were used to cluster the *Malus* orthologs in ten groups: with the majority of apple sequences in group VII, followed by VIII and X, whereas, only a few sequences clustered to group II, IV, V and VI, although without completion of the apple genome sequencing, it would be premature to attribute biological

meanings to the number of ERF proteins per group. However, the group VII represents only approximately 4% of the family in *Arabidopsis* (Nakano et al. 2006), 3% in *Populus* (Zhuang et al. 2008) and 10% in soybean (Zhang et al. 2008). In apple, it represents more than 21% of the single AP2/ERF-containing proteins found in open-access databases; thus indicating a significant biological role of the group in the species and differential evolution of the gene family in *Malus*. The functionally characterized gene *MdERF1* (A1LM1), associated to fruit ripening in apple (Wang et al. 2007) clustered in the most abundant group VII, whereas *MdERF2* (A1LM2), which is exclusively expressed in ripening fruits was clustered in group IX. The sequences from group VII exhibit a *Malus*-exclusive conserved motif CMVII-9 (KALFSLKVSTRGNIN), absent from functional site prediction tools. The novel motif found in *Malus* is hydrophobic and consists of polar, non-charged, non-aliphatic residues, suggesting a role in protein-protein interactions. The biological significance of these findings remains to be further investigated.

Apple ERFs exhibit the CMVIII-1 in group VIII, which is identical to an ERF-associated amphiphilic repression (EAR) motif, which has been shown to function as a repression domain (Fujimoto et al. 2000; Ohta et al. 2001). The EAR motif was identified as a conserved sequence, (L/F)DLN(L/F)xP, in the C-terminal regions of the repressor-type ERF proteins and also in TFIIIA-type C2H2 zinc-finger proteins (Ohta et al. 2001). Similar motifs are found in AUX/IAA proteins and AtERF4 (Tiwari et al. 2004) and within the repression domains of SUPERMAN and its related TFIIIA-type C2H2 zinc-finger proteins (Hiratsu et al. 2004).

4. 2- Cellular and molecular characteristics of apple ERF-like proteins

Proteins biochemical and physical properties are responsible for several characteristics associated to their biological function, such as molecular structure, ligand binding and subcellular location. Similar to *Arabidopsis*, rice, *Populus* and soybean (Nakano et al.

2006; Zhang et al. 2008; Zhuang et al. 2008), apple ERF-like proteins are predicted to have low molecular weight. In apple, most of the ERF-like sequences were predicted to be basic to neutral, as observed for *Arabidopsis* and *Populus*, although the proteins from the annual and perennial model plants exhibit slightly lower pI values. These observations are consistent with the presence of a basic region mediating sequence-specific DNA-binding in families of transcriptional regulators from evolutionarily distant organisms (Miller et al. 2003; Panne et al. 2004). Thus, the predicted chemical nature of the proteins identified in apple is compatible with a functional role in transcriptional control.

Folding state predictions are important tools to the analysis of protein structure and function and were, therefore, performed for apple ERF. The majority of the identified apple sequences sharing sequence similarity to ERFs were predicted to be unfolded under physiological conditions and a similar pattern was also observed for *Populus deltoides*, where the majority of the AP2/ERF proteins were predicted to be unfolded (Zhuang et al. 2008). In *Arabidopsis*, the ERF protein with the lowest percentage of disordered amino acid residues is At1g50680 (RAV1) with 19.8%, whereas in *Populus*, the DREB3-A subgroup protein PtERFB2-4 exhibited 99.2% of unfolded amino acid residues (Zhuang et al. 2008). The unfoldability prediction of *Malus* ERF sequences was intermediate between *Populus* and *Arabidopsis*. Consistently with observations from *Populus* (Zhuang et al. 2008), *Malus* ERF sequences clustered in the same group also exhibited distinct folding indices.

The predicted sub-cellular localization of the ERF proteins identified in *Malus* was investigated using the software WoLF pSORT (Horton et al. 2007). The sub-cellular location is an important clue for the biological function of identified sequences, especially for those hypothesized to be involved in gene expression regulation. The majority of the ERF-like sequences from apple were predicted to be exclusively nuclear or exhibiting ambiguous sub-cellular location for the nucleus and organelle or cytoplasm. At this point,

the incomplete nature of some of the sequences and the limitations of bioinformatic analyses prevent us from establishing how accurate the predictions are *in vivo*. Further functional analyses will be necessary to determine the sub-cellular location of the apple ERF-like sequences.

4. 3- *In silico* expression profiling of ERF genes in apple

The expression profile of the *ERF* genes in *Malus* was preliminarily investigated by *in silico* virtual northern blot. In order to avoid false-positives derived from orthologs of the input gene rather than the gene itself the databases were queried by an alignment corresponding to the queried gene and its orthologs. The heuristic allowed us to identify distinct expression patterns of five apple *ERF-like* sequences; the functionally characterized A1LM1 and A1LM2 (Wang et al. 2007), the SwissProt ERF validated match Q2LMD4, and the TIGR assembled contig TC6242. The transcription factor AHAP2 (Zhou et al. 2002), member of the AP2 family and corresponding to the Swissprot entry Q8VWW8, was also included. Phylogenetically, the sequences correspond to members of the AP2 (Q8VWW8) and ERF family (A1LM1, A1LM2, Q2LMD4 and TC6242), including members of the groups VII, IX, III and VIII, respectively.

Expression profiles of the *Malus* ERF genes indicate that the family is differentially regulated by development and biotic stress. The presence of distinct expression patterns observed in apple ERFs by *in silico* analysis is associated to the phylogenetic distance between the sequences, that is, phylogenetically related ERF proteins exhibited more similar expression patterns than divergent sequences. Thus, gene expression patterns suggest the existence of some extent of functional specialization for the apple ERF sequences investigated, although it is expected a high degree of function overlapping in large families of plant genes (Soltis et al. 2008; Sémon and Wolfe 2007).

The expression pattern of A1LM1, a protein associated to fruit ripening (Wang et al. 2007), was more similar to the profile observed for TC6242; both were highly expressed in libraries derived from buds, roots, vascular and fruit tissues, indicating a potential functional equivalence between the ERF-like sequences A1LM1 and TC6242.

In silico expression profile of Q2LMD4 that shares sequence similarity to *Arabidopsis* proteins involved in modulating dehydration and cold stress responses (Sun et al. 2008; Yamaguchi-Shinozaki and Shinozaki 2006), was divergent from the pattern of ripening-associated ERFs. Recently, it has been demonstrated that the 15th amino acid in the AP2/ERF domain of the *Arabidopsis* TINY protein, an ERF transcription factor clustered in the group III, is essential for its specific binding to the ethylene-responsive element (ERE) and that the 14th and 19th amino acids are responsible for the binding to the dehydration-responsive element (DRE) (Sun et al. 2008). *In silico* analysis indicates the possibility of a similar biological function linking ethylene- and stress-mediated gene regulation for *Malus* ERF Q2LMD4. Consistently, extensive analysis of A1LM1 and A1LM2 expression in apple fruits indicate that other factors besides ethylene are involved in regulating the transcription of *Malus* ERFs (Wang et al. 2007). Transcripts similar to A1LM2 were frequent libraries derived from cell cultures and bud samples; thus, suggesting a role for A1LM2 in developmental control. However, gene expression analyses by northern blot and RT-PCR were unable to detect A1LM2 expression outside fruit tissue (Wang et al. 2007). The observed differences in the expression pattern of A1LM2 could be due to methodological limitations of *in silico* and hybridization- and PCR-based approaches. High throughput sequencing technologies could help to elucidate the expression profile of A1LM2.

The *Malus* AP2 sequence Q8VWW8, which corresponds to the transcription factor AHAP (Zhou et al. 2002), exhibited higher expression in libraries obtained from flower, vascular and fruit tissues which is consistent with the role of AP2 proteins in

developmental control in *Arabidopsis* (Chandler et al. 2007) and other plant species (Chuck et al. 2008). An association between sequence conservation and gene expression pattern was detected by our *in silico* analyses.

In this study, comparative transcriptome and phylogenetic analysis were employed to investigate the family of ethylene response factors of apple. This preliminary survey of the ERF sequences in *Malus* has provided basic information for further in-depth studies relating to ethylene signaling mechanisms in this important perennial crop. Further studies of the genes identified here will also give a new perspective on the role of the gene family for economically important physiological conditions, such as mealiness and cork disease. Our results highlight the potential and the power of comparative genomics and *in silico* transcriptional profiling in agronomically important perennial fruit species.

5 ACKNOWLEDGMENTS

The authors would like to acknowledge the excellent technical assistance of Daniela Dal Bosco and Iraci Sinski. This research is funded by an Embrapa grant number 02.07.05.001.00 Macroprograma 2 - Agrofuturo.

6 RESUMO DO ARTIGO EM PORTUGUÊS

Análise *in silico* da família gênica *ETHYLENE RESPONSE FACTORS* (ERF) no gênero *Malus*

6.1- INTRODUÇÃO

Nas plantas os hormônios são responsáveis pelo controle do crescimento e desenvolvimento, funcionando também na intermediação da geração de respostas a vários fatores de estresses bióticos e abióticos. O etileno tem uma função importante na regulação de vários processos fisiológicos, tais como: germinação de sementes,

elongação celular, respostas a danos mecânicos, amadurecimento, senescência e abscisão; além de estar envolvido nas respostas a estresses abióticos e bióticos (Abeles et al., 1992; Chen et al. 2005). A ação do etileno é exercida por etapas complexas de regulação da biossíntese, percepção e transdução de sinais, que, finalmente, induzem alterações dramáticas na expressão gênica (Chang & Bleecker, 2004, Chen et al., 2005; Kendrick & Chang, 2008). As sequências dos promotores de vários genes induzidos por etileno foram caracterizadas e demonstrou-se que elas contêm um elemento de regulação em *cis*, denominado elemento de resposta ao etileno (ERE) (Broglie et al., 1989). As análises de várias regiões do elemento ERE identificaram um motivo curto, rico em nucleotídeos GC, o motivo GCC, essencial para respostas mediadas por etileno. Uma família conservada de fatores de transcrição se associa ao elemento ERE nos promotores dos genes regulados por etileno, sendo a família gênica conhecida como FATORES DE RESPOSTA AO ETILENO (ERF, do inglês ETHYLENE RESPONSE FACTORS) (Fujimoto et al., 2000).

Os reguladores transcricionais ERF constituem um subgrupo da superfamília APETALA2 (AP2)/ERF, que também contém as famílias AP2 e RAV (Riechmann et al., 2000). A superfamília é definida pela presença do domínio AP2/ERF, uma região conservada de ligação ao DNA com aproximadamente 60 a 70 aminoácidos (Weigel, 1995). A família AP2 é constituída por proteínas com dois domínios AP2/ERF; enquanto que na família ERF, as proteínas apresentam um domínio AP2/ERF simples. A família RAV é constituída de proteínas que possuem um domínio B3 de ligação ao DNA, associado a um único domínio AP2/ERF. Alguns estudos propuseram que a família ERF fosse subdividida em duas subfamílias; a subfamília ERF e a subfamília CBF/DREB (Sakuma et al., 2002). Dados mais recentes sugerem a presença de 10 grupos, caracterizados por motivos distintos conservados, dentro da família e dentro das subfamílias (Nakano et al., 2006).

A associação da função das proteínas ERF às respostas ao estresse faz com que estes fatores de transcrição sejam considerados candidatos a reguladores de características agronômicas de interesse, como resistência ao estresse hídrico e baixas temperaturas. Igualmente, o acúmulo de dados genômicos e de sequências expressas permitiram que fosse realizada uma análise bioinformática detalhada da subfamília ERF em várias espécies vegetais, incluindo *Arabidopsis*, arroz, *Populus*, soja e videira (Nakano et al., 2006; Zhang et al., 2008; Zhuang et al., 2008, Zhuang et al., 2009). Na produção de frutas, a família ERF está associada à biossíntese de compostos de aromas e à modificação da textura durante o amadurecimento (Alba et al., 2005; Da Silva et al., 2005; Ziliotto et al., 2008). Em maçã (*Malus x domestica* Borkh. Cv. *Golden Delicious*) dois transcritos da família ERF (*MdERF1* e *MdERF2*) foram caracterizados, respectivamente, como sendo predominante e exclusivamente expressos no amadurecimento de frutos (Wang et al., 2007). Da mesma forma, Newcomb et al. (2006) identificaram sequências semelhantes a ERFs em processos de amadurecimento em uma coleção extensa de sequências expressas.

Atualmente, estão disponíveis aproximadamente 400.000 etiquetas de sequências expressas (EST, do inglês Expressed Sequence Tags) para espécies de Rosáceas no National Center for Biotechnology Information (NCBI) e no Genome Database for Rosaceae (GDR) (www.rosaceae.org). Adicionalmente, cerca de 150.000 regiões 5' de ESTs de maçã foram depositadas no banco de acesso público HortResearch na Nova Zelândia (Newcomb et al., 2006). O número de EST publicamente disponíveis soma mais de 260.000 sequências, resultando em um conjunto de 17.000 unigenes. Ferramentas de bioinformática e análises de genômica comparativa foram usadas para investigar a função das proteínas ERF na regulação da expressão de genes responsáveis pelas mudanças na qualidade, textura e desordens fisiológicas pós-colheita da fruta em espécies do gênero *Malus*.

6.2- RESULTADOS

6.2.1- A família ERF em maçã

Foram utilizadas ferramentas de bioinformática para identificar os genes da família *ERF* no gênero *Malus*, a partir de sequências disponíveis em bancos de dados de acesso livre. Foram identificadas 15 sequências de maçã com similaridade a sequências de proteínas que contêm dois domínios AP2/ERF, sendo, portanto, atribuídas à família AP2. Quatro destas sequências exibem um único domínio AP2/ERF completo, diferente do ERF típico, sendo mais próximo ao tipo AP2. Foram identificados sete ESTs cujas sequências deduzidas de aminoácidos exibem conservação significativa do domínio funcional com relação a membros da família RAV. Sessenta e seis sequências de *Malus* foram agrupadas na família ERF. Desse modo, a superfamília AP2/ERF em *Malus* é composta de 80 genes, agrupados em três famílias; a família ERF (66 genes), a família AP2 (15 genes) e a família RAV (7 genes).

Não foram identificados em *Malus* ortólogos para o gene At4g13040 de *Arabidopsis*. Entretanto, no momento, não se pode concluir que este gene não existe em maçã devido à ausência da sequência completa do genoma da espécie. A identidade das proteínas do ERF em maçã com relação a seus homólogos em *Arabidopsis* variou de 23 a 70%, sendo maior para membros da família de WIN/SHN1, onde variou de 53 a 70%. O menor grau de conservação de sequência em comparação a *Arabidopsis* foi observado em ERFs de *Malus* pertencentes aos grupos VI e X.

As sequências deduzidas de aminoácidos do domínio AP2/ERF dos genes identificados em *Malus* foram alinhadas para estabelecer as relações filogenéticas. Os resíduos conservados Gly-4, Arg-6, Glu-16, Trp-28, Leu-29, Gly-30, e Ala-38 estão presentes em todas as 66 proteínas ERF de maçã. As regiões C-terminais dos domínios de AP2/ERF de duas proteínas (MdERF#41, MdERF#42) apresentam divergência na região de consenso.

6.2.2- Motivos conservados fora do domínio AP2/ERF

A fim de associar os ERFs putativos de maçã a funções biológicas, outros motivos conservados, fora da região AP2/ERF, foram investigados nas sequências deduzidas de aminoácidos, empregando análises de sequências múltiplas e o conjunto de programas MEME v.4.0 (Muralha & Elkan 1994). A maioria das sequências identificadas como ERF em maçã compartilham um ou mais motivos fora do domínio AP2/ERF com os seus homólogos de *Arabidopsis*.

Um grande número de sequências identificadas em *Malus* (21.2%, 14 sequências) foi agrupado com as sequências de *Arabidopsis* do grupo VII, do grupo VIII e X (15.1%, 10 membros de cada grupo). Em *Malus*, uma quantidade menor de proteínas foi agrupada com os ERFs de *Arabidopsis* pertencendo aos grupos III e IX (13.6% e 12.1%, 9 e 8 membros, respectivamente).

Os grupos dos ERF com menor número de membros em maçã foram II, IV, V e VI (4.5%, 3.0%, 4.5% e 3.0%, correspondendo a 3, 2, 3 e 2 membros, respectivamente). No grupo VII, as sequências de maçã exibem um motivo conservado, CMVII-9 (KALFSLKVSTRGNIN), exclusivo de *Malus*. Embora ausente de ferramentas de predição funcional, o novo motivo é hidrofóbico e consiste de resíduos sem carga, não-alifáticos e polares. Os ERF de maçã do grupo VIII exibem o CMVIII-1, idêntico ao motivo de repressão anfifílico (EAR).

6.2.3- Características moleculares e celulares

As propriedades físicas das sequências ERF, como o peso molecular (PM), pH e ponto isoelétrico (pI) e índice de dobramento, foram determinadas por análises de bioinformática. Os cálculos prevêm que a grande maioria das sequências ERF em maçã tem baixo peso molecular, variando de 1.5 a 4.48 kDa e em média 233 aminoácidos,

variando de 111 a 413 aminoácidos. Em maçã, a maioria dos ERF (74.2%, 49 sequências) foi classificada como básica ou neutra, enquanto que apenas 25.8% (17 sequências) exibiram predominância de aminoácidos ácidos.

O estado de dobramento das proteínas da família ERF em maçã foi determinado pelo programa FoldIndex (Prilusky et al., 2005). A maioria das sequências (59.1%, 39 sequências) encontra-se desdobradas em condições fisiológicas. A localização sub-celular das proteínas foi investigada usando o software WoLF pSORT (Horton et al., 2007). A maioria das sequências ERF de maçã (83.3%, 55 sequências) mostrou-se exclusivamente nuclear ou com localização sub-celular ambígua, entre o núcleo e organela ou citoplasma.

6.2.4- Perfis *in silico* da expressão gênica

O perfil de expressão dos genes *ERF* em *Malus* foi preliminarmente investigado *in silico* por Northern Blot virtual. As sequências investigadas correspondem a membros da família AP2 (Q8VWW8) e ERF (A1LM1, A1LM2, Q2LMD4 e TC6242), incluindo membros dos grupos VII, IX, III e VIII.

Níveis elevados de expressão foram observados em bibliotecas derivadas de gemas, raízes, xilema, floema e tecidos de fruto. Em bibliotecas provenientes da análise diferencial de expressão em tecidos infectados com patógenos foi detectada uma baixa frequência de transcritos derivados dos genes *ERF* investigados. A análise de agrupamento hierárquico indicou que o perfil de expressão *in silico* de Q2LMD4 foi divergente com relação ao padrão observado para os membros da família ERF, embora a indução em bibliotecas derivadas de gemas, xilema e tecidos de floema de maçã tenha sido observada para membros de ambas as famílias. Transcritos semelhantes ao *A1LM1* foram frequentes em bibliotecas associadas ao fruto, enquanto que transcritos de *A1LM2* também foram encontrados em culturas de células e bibliotecas de gemas. Em *Malus*, o

gene correspondente ao fator de transcrição da família AP2 (Q8VWW8) foi mais expresso em bibliotecas obtidas de flores, frutas e tecidos vasculares.

7 REFERENCES

Aharoni A; Dixit S; Jetter R; Thoenes E; Van Arkel G; Pereira A (2004) The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. *Plant Cell* 16:2463–2480

Allen M.D; 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

Alonso J.M; Stepanova A.N; Leisse T.J; Kim C.J; Chen H; Shinn P; Stevenson D.K; Zimmerman J; Barajas P; Cheuk R; Gadrinab C; Heller C; Jeske A; Koesema E; Meyers C.C; Parker H; Prednis L; Ansari Y; Choy N; Deen H; Geralt M; Hazari N; Hom E; Karnes M; Mulholland C; Ndubaku R; Schmidt I; Guzman P; Aguilar-Henonin L; Schmid M; Weigel D; Carter D.E; Marchand T; Risseuw E; Brogden D; Zeko A; Crosby W.L; Berry C.C; Ecker JR (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653–657

Altschul S.F; Madden L; Schäffer A.A; Zhang J; Zhang Z; Miller W; Lipman D.J (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc Acids Res* 25:3389-3402

Bailey T.L; Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Altman R, Brutlog D, Karp P, Lathrop R, Searls D, editors. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*. Menlo Park, CA: American Association for Artificial Intelligence Press p 28–36.

Banno H; Ikeda Y; Niu Q.W; Chua N.H (2001) Overexpression of *Arabidopsis* ESR1 induces initiation of shoot regeneration. *Plant Cell* 13:2609–26018

Benedict C; Skinner J.S; Meng R; Chang Y; Bhalerao R; Huner N.P; Finn C.E; Chen T.H; Hurry V (2006) The CBF1-dependent low temperature signalling pathway, regulon and increase in freeze tolerance are conserved in *Populus* spp. *Plant Cell Environ* 29:1259-1272

Berrocal-Lobo M; Molina A; Solano R (2002) Constitutive expression of *ETHYLENE-RESPONSE-FACTOR1* in *Arabidopsis* confers resistance to several necrotrophic fungi. *Plant J* 29:23–32

Brocard I.M; Lynch T.J; Finkelstein R.R (2002) Regulation and role of the *Arabidopsis* *ABSCISIC ACID-INSENSITIVE 5* gene in abscisic acid, sugar, and stress response. *Plant Physiol* 129:1533-1543

Broglie K.E; Biddle P; Cressman R; Broglie R (1989) Functional analysis of DNA sequences responsible for ethylene regulation of a bean chitinase gene in transgenic tobacco. *Plant Cell* 1:599-607

Broun P; Poindexter P; Osborne E; Jiang C.Z; Riechmann JL (2004) WIN1, a transcriptional activator of epidermal wax accumulation in *Arabidopsis*. *Proc Natl Acad Sci U S A* 101: 4706–4711

Büttner M; Singh K.B (1997) *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA binding protein interacts with an OCS element binding protein. *Proc Natl Acad Sci U S A* 94:5961–5966

Chandler J.W; Cole M; Flier A; Grewe B; Werr W (2007) The AP2 transcription factors DORNROSCHEN and DORNROSCHEN-LIKE redundantly control *Arabidopsis* embryo patterning via interaction with PHAVOLUTA. *Develop* 134:1653-1662

Chuck G; Meeley R; Hake S (2008) Floral meristem initiation and meristem cell fate are regulated by the maize *AP2* genes *ids1* and *sid1*. *Develop* 135:3013-3019

Eisen M.B; Spellman P.T; Brown P.O; Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95:14863-14868

Finkelstein R.R; Wang M.L; Lynch T.J; Rao S; Goodman H.M (1998) The *Arabidopsis* abscisic acid response locus *ABI4* encodes an APETALA 2 domain protein. *Plant Cell* 10:1043–1054

Fujimoto S.Y; Ohta M; Usui A; Shinshi H, Ohme-Takagi M (2000) *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* 12:393-404

Gilmour S.J; Sebolt A.M; Salazar M.P; Everard J.D; Thomashow MF (2000) Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiol* 124:1854–1865

Gilmour S.J; Zarka D.G; Stockinger E.J; Salazar M.P; Houghton J.M; Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced gene expression. *Plant J* 16:433–442

Gu Y.Q; Wildermuth M.C; Chakravarthy S; Loh Y.T; Yang C; He X; Han Y; Martin G.B (2002) Tomato transcription factors *pti4*, *pti5*, and *pti6* activate defense responses when expressed in *Arabidopsis*. *Plant Cell* 14:817–831

Gu Y.Q; Yang C; Thara V.K; Zhou J; Martin G.B (2000) *Pti4* is induced by ethylene and salicylic acid, and its product is phosphorylated by the *Pto* kinase. *Plant Cell* 12: 771–786

Guo Z.J; Chen X.J; Wu X.L; Ling J.Q; Xu P (2004) Overexpression of the AP2/EREBP transcription factor *OPBP1* enhances disease resistance and salt tolerance in tobacco. *Plant Mol Biol* 55:607–618

Gutterson N; Reuber T.L (2004) Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr Opin Plant Biol* 7:465–471

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

Hao D; Ohme-Takagi M; Sarai A (1998) Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *J Biol Chem* 273:26857-26861

Horton P; Park K-J; Obayashi T; Fujita N; Harada H; Adams-Collier C.J; Nakai K (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35: W585-7.
doi:10.1093/nar/gkm259

Janssen B.J; Thodey K; Schaffer R.J; Alba R; Balakrishnan L; Bishop R; Bowen J.H; Crowhurst RN, Gleave AP, Ledger S, McArtney S, Pichler FB, Snowden KC, Ward S (2008) Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. *BMC Plant Biol* 8:16

Kannangara R; Branigan C; Liu Y; Penfield T; Rao V; Mouille G, Höfte H; Pauly M; Riechmann J.L; Broun P (2007) The transcription factor WIN1/SHN1 regulates cutin biosynthesis in *Arabidopsis thaliana*. *Plant Cell* 19:1278–1294

Kendrick M.D; Chang C (2008) Ethylene signaling: new levels of complexity and regulation. *Curr Opin Plant Biol* 11:479-485

Kirch T; Simon R; Grunewald M; Werr W (2003) The *DORNROSCHEN/ ENHANCER OF SHOOT REGENERATION1* gene of *Arabidopsis* acts in the control of meristem cell fate and lateral organ development. *Plant Cell* 15:694–705

Kunst L; Klenz J.E; Martinez-Zapater J; Haughn G.W (1989) *AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. *Plant Cell* 1:1195-1208

Lin R.C; Park R.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

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

McGrath K.C; Dombrecht B; Manners J.M; Schenk P.M; Edgar C.I; Maclean D.J; Scheible W.R; Udvardi M.K; Kazan K (2005) Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol* 139: 949–959

Miller M; Shuman J.D; Sebastian T; Dauter Z; Johnson P.F (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem* 278:15178-15184

Newcomb R.D; Crowhurst R.N; Gleave A.P; Rikkerink E.H.A; Allan A.C; Beuning L.L; Bowen J.H; Gera E; Jamieson K.R; Janssen B.J; Laing W.A; McArtney S; Nain B; Ross G.S; Snowden K. C; Souleyre E.J; Walton E.F; Yauk Y.K (2006) Analyses of expressed sequence tags from apple. *Plant Physiol* 141:147–166

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

Ohta M; Ohme-Takagi M; Shinshi H (2000) Three ethylene-responsive transcription factors in tobacco with distinct trans-activation functions. *Plant J* 22:29–38

Okamuro J.K; Caster B; Villarroel R; Van Montagu M; Jofuku K.D (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc Natl Acad Sci U S A* 94:7076-7081

Oñate-Sánchez L; Singh K.B (2002) Identification of *Arabidopsis* ethylene-responsive element binding factors with distinct induction kinetics after pathogen infection. *Plant Physiol* 128:1313–1322

Panne D; Maniatis T; Harrison S.C (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J* 23:4384-4393

Park J.M; Park C.J; Lee S.B; Ham B.K; Shin R; Paek K.H (2001) Overexpression of the tobacco *Tsi1* gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. *Plant Cell* 13:1035–1046

Park S; Sugimoto N; Larson M.D; Beaudry R; Van Nocker S (2006) Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags. *Plant Physiol* 141:811-824

Prilusky J; Felder C.E; Zeev-Ben-Mordehai T; Rydberg E; Man O; Beckmann J.S; Silman I; Sussman J.L (2005) FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21:3435-3438

Riechmann J.L; Heard J; Martin G; Reuber L; Jiang C; Keddie J; Adam L; Pineda O; Ratcliffe O.J; Samaha R.R; Creelman R; Pilgrim M; Broun P; Zhang J.Z; Ghandehari D; Sherman B.K; Yu G (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290:2105–2110

Sakuma Y; Liu Q; Dubouzet J.G; 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 Com* 290:998–1009

Schramm F; Larkindale J; Kiehlmann E; Ganguli A; Englich G; Vierling E; Von Koskull-Dör P (2008) A cascade of transcription factor DREB2A and heat stress transcription factor HsfA3 regulates the heat stress response of *Arabidopsis*. *Plant J* 53:264-274

Sémon M; Wolfe K.H (2007) Consequences of genome duplication. *Curr Opin Gen Dev* 17:505-512

Shulaev V; Korban S.S; Sosinski B; Abbott A.G; Aldwinckle H.S; Folta K.M; Iezzoni A; Main D; Arús P; Dandekar A.M; Lewers K; Brown S.K; Davis T.M; Gardiner S.E; Potter D; Veilleux R.E (2008) Multiple models for Rosaceae genomics. *Plant Physiol* 147:985–1003

Soltis D.E; Bell C.D; Kim S; Soltis P.S (2008) Origin and early evolution of angiosperms. *Ann New York Acad Sci* 1133:3-25

Song C.P; Agarwal M; Ohta M; Guo Y; Halfter U; Wang P; Zhu J.K (2005) Role of an *Arabidopsis* AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *Plant Cell* 17:2384–2396

Stockinger E.J; Gilmour S.J; Thomashow M.F (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. *Proc Natl Acad Sci U S A* 94:1035–1040

Sun S; Yu J.P; Chen F; Zhao T.J; Fang X.H; Li Y.Q; Sui S.F (2008) TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in *Arabidopsis*. *J Biol Chem* 283:6261-6271

Thompson J.D; Gibson T.J; Plewniak F; Jeanmougin F; Higgins D.G (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876-4882

Tiwari S.B; Hagen G; Guilfoyle T.J (2004) Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* 16:533–543

Van der Graaff E; Dulk-Ras A.D; Hooykaas P.J, Keller B (2000) Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*. *Develop* 127:4971–4980

Wang A; Tan D; Takahashi A; Li T.Z; Harada T (2007) *MdERFs*, two ethylene-response factors involved in apple fruit ripening. *J Exp Bot* 58:3743-3748

Wang Z; Triezenberg S.J; Thomashow M.F; Stockinger E.J (2005) Multiple hydrophobic motifs in *Arabidopsis* CBF1 COOH-terminus provide functional redundancy in trans-activation. *Plant Mol Biol* 58:543–559

Weigel D (1995) The APETALA2 domain is related to a novel type of DNA binding domain. *Plant Cell* 7:388-389

Yamaguchi-Shinozaki K; Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Ann Rev Plant Biol* 57:781-803

Yamamoto S; Suzuki K; Shinshi H (1999) Elicitor-responsive, ethyleneindependent activation of GCC box-mediated transcription that is regulated by both protein phosphorylation and dephosphorylation in cultured tobacco cells. *Plant J* 20:571–579

Yang Z; Tian L; Latoszek-Green M; Brown D; Wu K (2005) *Arabidopsis* ERF4 is a transcriptional repressor capable of modulating ethylene and abscisic acid responses. *Plant Mol Biol* 58:585–596

Zhang G; Chen M; Chen X; Xu Z; Guan S; Li L.C; Li A; Guo J; Mao L; Ma Y (2008) Phylogeny, gene structures, and expression patterns of the *ERF* gene family in soybean (*Glycine max* L.). *J Exp Bot* 9:4095–4107

Zhou J; Tang X; Martin G.B (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. *EMBO J* 16:3207–3218

Zhuang J; Cai B; Peng R.H; Zhu B; Jin X.F; Xue Y; Gao F; Fu X.Y; Tian Y.S; Zhao W; Qiao Y.S; Zhang Z; Xiong A.S; Yao Q.H (2008) Genome-wide analysis of the *AP2/ERF* gene family in *Populus trichocarpa*. *Biochem Biophys Res Commun* 371:468–474

Zhuang J; Peng R.H; Cheng Z.M; Zhang J; Cai B; Zhang Z; Gao F; Zhua B; Fua X.Y; Jin X.F; Chen J.M; Qiao Y.X (2009) Genome-wide analysis of the putative *AP2/ERF* family genes in *Vitis vinifera*. *Scientia Horticulturae*, v.123, p.73-81

CAPÍTULO II

Expressão dos genes *MdERF1e MdERF2* em plantas infectadas com fungo *Venturia inaequalis*

1 INTRODUÇÃO

Níveis de expressão gênica em plantas podem mudar em resposta a diferentes sinais ambientais. As plantas percebem esses sinais e iniciam uma cascata de eventos, alterando a expressão gênica e suas respostas fisiológicas (Sierotzki e Gessler, 1998). Alguns genes destas respostas são agrupados sob o termo de proteínas PR (Pathogenesis Related Proteins) (Kitajima et al, 1999). Esta grande família é dividida em proteínas PR básicas, que são reguladas a nível transcricional por hormônios etileno e metil jasmonato e proteínas PR ácidas que são reguladas por ácido salicílico e radicais de oxigênio. Análise dos promotores de proteínas PR básicas em tabaco permitiu identificar um elemento regulador de resposta a etileno, chamado Ethylene Responsive Element (ERE), o qual leva a ativação de tais proteínas (Ohme-Takagi e Shinshi, 1995). Esta sequência TAAGAGCCGCC, também chamada de GCC, foi posteriormente identificada em muitas outras espécies (Ohme-Takagi e Shinshi, 1995; Sato et al. 1996, Solano et al. 1998; Jia e Martin, 1999). Os fatores de transcrição de ligação a este elemento foram isolados em tabaco (Ohme-Takagi e Shinshi, 1995) e nomeado Ethylene Responsive Element Binding Proteins (EREBPs), sendo renomeados ETHYLENE RESPONSE FACTOR (ERF) (Solano et al. 1998). Atkinson et al. (1996) citam que em maçã, vários homólogos do gene PR-10 foram identificados e verificados que se

acumulam nos frutos durante o amadurecimento. A expressão da subclasse do gene *APa* foi fortemente induzida em folhas, localmente e sistematicamente pelo acibenzolar-S-metil (ASM), um análogo funcional do ácido salicílico. Pascal et al. (2003) verificaram que os genes *APa* mostraram-se altamente induzidos em folhas por estímulos abióticos (ethephon, ferimento), bem como pela inoculação de cepas virulentas ou avirulentas de *Venturia inaequalis*. Esses mesmos autores citam que a superexpressão do gene *APa* após infecção com cepas avirulentas foi concomitante com a ocorrência de sintomas de resistência.

A sarna, causada pelo fungo *Venturia inaequalis*, é a principal doença da macieira nas regiões de clima temperado e úmido, podendo causar perdas de até 100% na produção de maçãs. As perdas causadas se manifestam por meio da queda das flores, danos aos frutos, desfolhamento e consequentemente diminuição do vigor da planta. Nas folhas novas, tanto na face superior quanto na inferior, surgem pequenas lesões de coloração verde-oliva. Essas desenvolvem-se radialmente e adquirem aspecto aveludado, de coloração cinza escura, e de contorno aproximadamente circular (Valdebenito-Sanhueza et al. 2005). Aplicações de fungicidas e pesticidas ajudam no controle das doenças, mas o controle químico é oneroso, não só economicamente, mas também ambientalmente indesejáveis. Portanto, a utilização mais eficaz dos recursos genéticos naturais como fonte de resistência para o melhoramento genético tem sido foco de interesse para a investigação. Resistência em maçã pode ser alcançada através da combinação de genes em uma simples variedade, reduzindo assim o risco de um patógeno superar a resistência associada a um gene único. Resistência a essa doença é derivada de uma

espécie selvagem de maçã, *Malus floribunda* 821, e a região cromossômica contendo um loco de resistência, *Vf*, tem sido amplamente introduzido em cultivares comerciais suscetíveis (Korban 1998). De fato, o gene *Vf* de resistência a sarna foi superada nos últimos anos em maçã na Europa (Bénaouf Parisi 2000). Outros genes de resistência à sarna estão disponíveis em parentes selvagens de maçã (Williams e Kuc 1969). A combinação de diferentes genes de resistência com o gene *Vf* seria muito vantajosa, seja ela via melhoramento clássico como em futuras plantas transgênicas. Portanto, o objetivo desse trabalho foi verificar o nível de expressão de dois genes *ERFs* em resposta a infecção de um dos principais patógenos que afetam a cultura da maçã. Esse trabalho poderá futuramente propor a utilização de um gene-candidato associado a resistência da planta a doenças fungicas em maçãs.

2 MATERIAL E MÉTODOS

2.1- Material Vegetal

2.1.1- Cultura de tecidos

Brotações laterais de aproximadamente 1 cm de comprimento foram removidas de plantas de *Malus x domestica*, cultivar Royal Gala, mantidas *in vitro*, em meio MS (Murashige; Skoog, 1962) a pH 5,8, suplementado com 2 mg.L⁻¹ de BAP. As brotações foram transferidas para frascos contendo aproximadamente 30 mL de meio de micropropagação, consistindo de sais e vitaminas MS (Murashige; Skoog, 1962) a pH 5,8, suplementado com 2 mg/L de BAP e 30 g.L⁻¹ de sacarose, solidificado com 6 g.L⁻¹ de agar. As plantas foram mantidas em câmara de crescimento com temperatura de 22°C ± 2°C e fotoperíodo de 16:8 h. Após quatro semanas, foram removidas brotações jovens com aproximadamente 2 cm de comprimento e transferidas para meio

MS (Murashige; Skoog, 1962) a pH 5,8, suplementado com 30 g.L⁻¹ de sacarose, para enraizamento. As plantas exibindo raízes bem desenvolvidas após seis semanas foram gradativamente aclimatadas em copos plásticos com capacidade para 250 mL, contendo uma mistura 1:1 de vermiculita média e substrato vegetal comercial (Plantmax[®] Horticultura, Eucatex), sendo mantidas em câmara úmida por aproximadamente quatro semanas sob temperatura de 25°C ± 3°C e iluminação fluorescente constante. Após este período, a câmara úmida foi removida e as plantas submetidas à inoculação.

2.1.2- Inoculação

O inóculo de *Venturia inaequalis* (sarna da macieira) foi preparado no laboratório de fitopatologia da Embrapa Uva e Vinho a partir de isolados monoconidiais cultivados em placas de Petri contendo meio de Batata-Dextrose-Ágar-Extrato de Levedura (40; 5; 17; 3 g.L⁻¹). O fungo foi incubado a 16°C e luz contínua, por 15 dias. As suspensões conidiais foram ajustadas para 2,5 x 10⁵ conídios. mL⁻¹, a qual foi aspergida nas plantas até o ponto de escorrimento. Para cada planta inoculada, uma planta foi igualmente aspergida com água e mantida como testemunha, não inoculada. Depois da inoculação as macieiras foram colocadas em incubadoras BOD (Eletrolab) durante 30 dias utilizando um fotoperíodo de 12:12 (luz:escuro), umidade relativa de 90-100% e temperatura de 16 a 20°C. Após esse período as plantas (inoculadas e controle) foram congeladas em nitrogênio líquido e armazenadas a -80°C.

2.2- Estudos da expressão do *MdERF1* e *MdERF2*

Os estudos de expressão dos genes *MdERF1* e *MdERF2* realizados a partir do material coletado de plantas infectadas com o fungo da sarna ou controle foram realizados por RT-PCR de acordo com Zegzouti *et al* (1999). O procedimento envolveu a extração de RNAs, a síntese de cDNAs e amplificação por PCR dos fragmentos correspondentes e ao controle interno *MdACT*(actina).

2.3- Extração de RNA total

Para a extração dos RNA total foi usado o método descrito por Boss, *et al* (1996). O material foi macerado em nitrogênio líquido, colocado em tubos o equivalente a 15ml de amostra e posteriormente colocado o mesmo volume de tampão de extração (SDS 1%, PVP 2%, solução de fenol 5 %, EDTA 25mM), fenol: clorofórmio (5:1) e em seguida homogeneizado em vortex por 5 minutos. Foi centrifugado á 3600 rpm durante 50 minutos a 4°C, recuperado a fase superior e a essa adicionou-se 1/10 do volume de 3M NaAc pH 4,5, um volume de isopropanol e deixou-se durante 1 noite a -20°C. Após a centrifugação a 3600 rpm por 30 min a 4°C, descartou-se o sobrenadante, adicionou-se 5ml de etanol 70% gelado, foi centrifugado mais 5 minutos nas mesmas condições, eliminado novamente a fase aquosa e o pellet foi submetido à secagem. Foi colocado 1ml de água milliQ autoclavada para dissolver o pellet e posteriormente adicionou-se mais 3 ml de água para dissolução total, imediatamente após foi colocado o mesmo volume (4ml) do tampão CTAB 2X (CTAB 2%, 100mM Tris-HCL pH 8,0, 20mM EDTA pH 8,0 e 1,4M NaCl), 8ml do CTAB 1x (CTAB 1%, 50mM Tris-HCl pH 8,0, 10mM EDTA pH 8,0) e centrifugou-se a 3600 rpm por 50 min a 22°C. Depois de eliminado o

sobrenadante, colocou-se 1ml de água milliQ autoclavada para dissolver o pellet o mais rápido possível, colocou-se mais 1ml de água, 2ml de CTAB 2x e 4ml CTAB 1x e agitou-se o tubo e após centrifugou-se nas mesmas condições anteriores mas por 30 minutos. Eliminada a fase aquosa adicionou-se 1ml de NaCl 1,4M (+ 4°C) para dissolução total, mais 5ml de etanol 100% e acondicionado a -20°C por uma noite. Após centrifugação a 3600 rpm por 30 minutos a 4°C, eliminou-se o sobrenadante, foi acrescentado 1ml etanol 70% gelado, transferiu-se o pellet com ajuda de pipeta para um micro tubo no gelo e centrifugou-se a 13600 rpm por 10 minutos a 4°C, descartado a fase aquosa, recentrifugou-se para eliminação total do sobrenadante, o pellet foi submetido a secagem. Os RNAs foram ressuspensos em 400 µL de água milliQ autoclavada.

2.3.1-Tratamento com DNase e RNase-free

Para eliminação do DNA foi realizado o tratamento com 1µL de inibidor de RNase (Invitrogen), 45 µL de tampão de DNase 10x e 8µL de DNase I (BioLabs) e incubado por 1h a 37°C em banho-maria. Os RNAs foram reconcentrados através de uma série de precipitações.

2.3.2- Precipitação dos RNA

Ao RNA obtido na etapa anterior (454 µL) foi adicionado o mesmo volume de fenol: clorofórmio (5:1) agitou-se por 30 segundos e centrifugou-se a 12000 rpm por 5 minutos a 4°C. A fase superior foi transferida para outro micro tubo, ajustado o mesmo volume de fenol: clorofórmio (5:1), homogeneizado no “vortex” por 30 segundos e foi centrifugado nas mesmas condições anteriores, mas com a rotação de 13600rpm. Novamente a fase superior foi transferida

para outro micro tubo, ajustado o mesmo volume de LiCl 6M (4°C) e foi incubado -20°C por uma noite. Após a centrifugação a 13600 rpm por 15 minutos a 4°C, foi descartado o sobrenadante, adicionou-se 1ml de LiCl 3M (4°C), homogeneizou-se e foi centrifugado a 13600 rpm por 10 minutos a 4°C. Depois de eliminado o sobrenadante, recentrifugou-se nas mesmas condições por 30 segundos e ressuspendeu-se em 400 µl KoAc 2% (+4°C), em seguida foi ajustado o volume de 1ml de etanol 100%, homogeneizado por 10 segundos e deixado por uma noite a -20°C. Posteriormente, centrifugou-se a 13.600rpm por 30 minutos a 4°C, foi descartado o sobrenadante, ajustado o volume de 1ml de etanol 70% (-20°C) e centrifugou-se a 13600rpm por 10 minutos a 4°C. O sobrenadante foi descartado, recentrifugou-se por mais 30 segundos e o pellet foi submetido à secagem. Os RNAs foram então, ressuspensos em água milliQ autoclavada (50 a 150 µL) e procedeu-se a análise qualitativa e quantitativa dos mesmos.

2.3.3- Análise quantitativa e qualitativa dos RNAs totais extraídos

Para a quantificação do ácido nucléico, a amostra foi diluída na proporção de 1:50 em água milliQ autoclavada. A determinação da concentração foi realizada em espectrofotômetro (*Amersham Biosciences®*, modelo *GeneQuant Pro*) com leituras nos comprimentos de onda 230 (açúcares), 260 (ácidos nucléicos) e 280 (proteínas) nm. A quantidade de RNA da amostra foi calculada a partir da leitura 260 nm, uma vez que 1 unidade de DO_{260nm} equivale a 40µg. mL⁻¹ de RNA em solução. A análise qualitativa foi realizada através de três características: pureza, integridade e a presença de DNA residual. A pureza(P) foi determinada pela razão DO_{260}/DO_{280} (é considerada pura se $1,8 < P < 2,0$). A integridade foi verificada através de uma eletroforese em gel de agarose 0,8%

à 80 V.cm^{-1} , em tampão TAE 1X. Pode-se verificar pela migração também se o DNA residual foi eliminado.

2.4- Síntese de cDNA

Para a síntese de cDNAs foi utilizado o kit comercial M-MLV Reverse Transcriptase RT-PCR (Promega). Essa técnica consiste na síntese de cDNA (uso da enzima transcriptase reversa) com oligonucleotídeos selecionados a partir de seqüências já conhecidas e de interesse.

Em tubos tipo Eppendorf®, para um volume final de $125\mu\text{L}$, foi adicionado $1\mu\text{L}$ de RNA ($10\mu\text{g}/\mu\text{L}$), $10\mu\text{L}$ de Oligo(dt) 12-18($0,5\mu\text{g}/\mu\text{L}$). As amostras foram incubadas 70°C por 10 minutos e após mantidas em gelo por 10 minutos. Posteriormente foi adicionado $25\mu\text{L}$ de Tampão 5x RT, $6,25\mu\text{L}$ de solução de dNTPs mix 10 mM, $5\mu\text{L}$ de M-MLV RT, $1\mu\text{L}$ de RNaseOUT Recombinant RNase Inibitor, incubado a 37°C por 90 minutos em termociclador (Bio-Rad®), e depois 95°C por 5 minutos.

Para a purificação dos cDNAs obtidos foi completado o volume para $500\mu\text{L}$ com água milliQ autoclavada, acrescentado $500\mu\text{L}$ de clorofórmio, homogeneizou-se por inversão e foi centrifugado a 13.600 rpm por 5 minutos a 20°C . A fase aquosa foi transferida para outro tubo, adicionou-se 1/10 do volume de acetato de amônio 3M e completou-se o volume com isopropanol. Foi mantido a -20°C por 2 horas e posteriormente centrifugado a 13600rpm por 10 minutos. O isopropanol foi descartado e deixou-se secar o pellet. O cDNA limpo foi ressuspendido em $50\mu\text{L}$ de água milliQ autoclavada.

2.5- PCR semi-quantitativa

As amplificações por PCR (*Polimerase Chain Reaction*), foram realizadas num volume final de 25µL, contendo 2,5µL tampão Taq[®] DNA polimerase 10x, 2,5µL MgCl₂ (25mM), 4,0µL dNTPs (2,5mM), 1µL oligonucleotídeos(10mM), 1µL cDNA e 2,0µL Taq[®] DNA polimerase(Invitrogen). Os oligonucleotídeos utilizados para amplificação foram *MdERF1F* (5'ATGACCTGGTGGCATATCAG3'), *MdERF1R*(5'CACCGTAGCAAACAACACAC3') e *MdERF2F* (5'TATGCTGGCAATTGGCGAGC3'), *MdERF2R* (5'ATGACCAATCCCGCACTCAC3') que amplificam um fragmento de 196 pb e 198 respectivamente (Wang et al., 2007). Para normalizar a concentração de RNA de cada amostra, como controle interno utilizou-se os oligonucleotídeo *MdACTF* (5'GGCTGGATTGCTGGTGATG3') e *MdACTR* (5'TGCTCACTATGCCGTGCTCA3') para uma amplificação de 179 pb (Wakasa, 2006). As amplificações foram realizadas em termociclador nas seguintes condições: 1 ciclo de 4 minutos a 95°C; 27 ciclos de 45 segundos a 95°C, 2 minutos a 52°C, 2 minutos a 72°C e 1 ciclo de 10 minutos a 72°C. O produto da PCR foi submetido à eletroforese em gel de agarose 2%. A migração foi realizada em cubas de eletroforese (Bio-Rad[®]) a 80 V.cm⁻¹ e tampão TAE 1X por aproximadamente 40 minutos.

3 RESULTADO E DISCUSSÃO

3.1- Avaliação das plantas inoculadas com *Venturia inaequalis*

Pode-se observar na figura 1B que após o período de 30 dias de inoculação com o fungo de *Venturia inaequalis*, as plantas apresentavam os sintomas típicos de manchas de sarna da macieira. As plantas controle por outro lado

não demonstraram os mesmos sintomas (Figura 1A). É importante comentar que o estabelecimento de infecções fúngicas sob condições controladas permite estudar respostas genéticas específicas que estejam associadas ao efeito da doença causada. Isso muitas vezes pode ser difícil em condição de campo, visto que outros fatores e agentes patogênicos podem interferir nos resultados obtidos.



Figura 1: Plantas controle (A) e com sintomas de sarna (B)

3.2- Estudo de expressão dos ERF

Pode-se observar nos resultados obtidos pela técnica de RT-PCR (Figura 2), que o gene *MdACT* apresentou um mesmo nível de expressão constitutiva basal normalizada em ambas as amostras avaliadas (controle(1) X inoculada(2)). Esse resultado foi obtido a partir de avaliações preliminares realizadas onde se estabeleceu uma otimização de 27 ciclos para amplificar os fragmentos desejados dentro da fase exponencial de amplificação. Portanto, nessa mesma figura observa-se que as amostras tratadas com o fungo *Venturia inaequalis* induziu a expressão do gene *MdERF1*. Verificou-se que na amostra controle (1) o mesmo gene apresentou um nível de acumulação de transcritos bastante inferior, praticamente não aparecendo na imagem obtida no gel. Esse mesmo resultado não foi observado para o gene *MdERF2*, sendo que o mesmo mostrou um mesmo nível de expressão em ambas as amostras (controle e fungo).

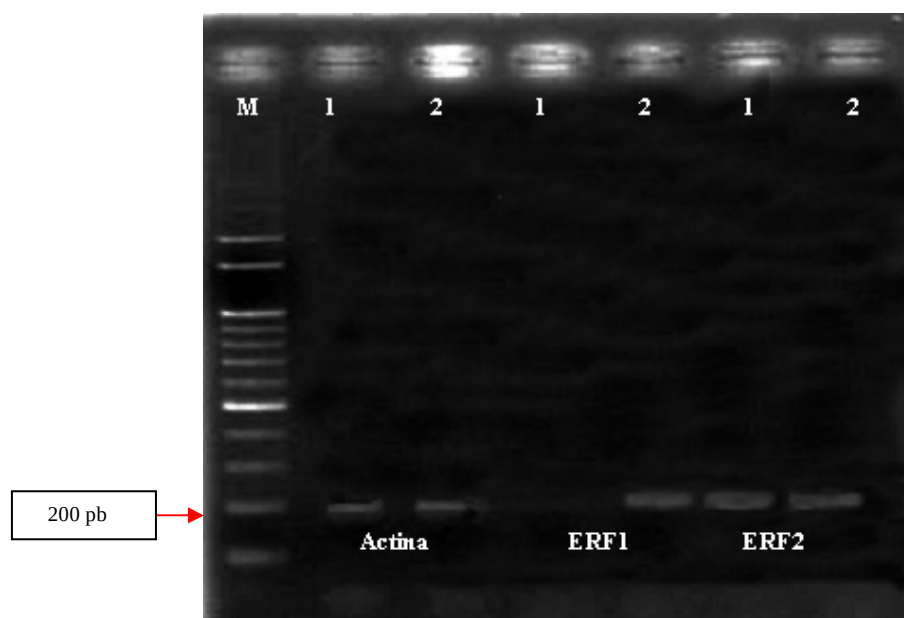


Figura 2: Acúmulo de transcritos dos genes *MdActina*, *MdERF1* e *MdERF2* em plantas de maçã, cv Royal Gala não inoculadas (1) e inoculadas (2) com fungo *Venturia inaequalis*.

Pode-se dizer, portanto, que a expressão do gene *MdERF1* é induzida por patógenos, sendo que essa resposta pode estar associada de uma maneira indireta a mecanismos de resistência da planta moduladas por mudanças metabólicas associadas, tais como a indução de proteínas PR. Tournier et al. (2003) verificaram que a expressão do gene *LeERF1* de tomate é fortemente induzido pelo etileno, enquanto que *LeERF2* não foi afetado. Segundo Métraux, (2001), o ácido jasmônico (jasmonato), o ácido abscísico e o etileno podem desempenhar um papel fundamental no fenômeno de transdução de sinais intercelulares resultantes da indução por patógenos. Entretanto, esses mecanismos mostraram-se específicos em apenas um dos genes estudados, devendo os mesmos serem posteriormente confirmados por outras técnicas moleculares.

4 CONCLUSÃO

Pode-se concluir que a expressão gene *ERF1* é induzida pelo fungo *Venturia inaequalis*. Esses resultados são preliminares para traçar afirmativas mais contundentes, necessitando novos estudos serem realizados. Em um primeiro momento é necessário confirmar esse resultado por Real Time PCR, para posteriores caracterizações moleculares mais específicas.

5 REFERÊNCIAS BIBLIOGRÁFICAS

Atkinson R.G; Perry J; Matsui T; Ross G.S; Macrae E.A (1996) A stress, pathogenesis, and allergen-related cDNA in apple fruit is also ripening-related. New Zealand Journal of Crop and Horticultural Science; 24:103–7.

Be'naouf G; Parisi L. (2000) Genetics of host-pathogen relationships between *Venturia inaequalis* races 6 and 7 and *Malus* species. *Phytopathology* 2000; 90:236–42

Boss P.K; Davies C; Robinson S.P. (1996) Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv. Shiraz grape berries and the implications for pathway regulation. *Plant Physiology*, vol. 111, no. 4, p. 1059-1066.

Jia Y; Martin G.B (1999) Rapid transcript accumulation of pathogenesis-related genes during an incompatible interaction in bacterial speck disease-resistant tomato plants. *Plant Mol Biol* 40: 455-465

Kitajima S; Sato F (1999) Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. *J. Biochem* 125: 1-8

Korban S. S; (1998) What's new with disease-resistant apple cultivars. *Proc. Trans. Ill. Hort. Soc.* 131: 74–76.

Murashige T; Skoog F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiologia Plantarum*, v.15, p.473–497,.

Métraux J. P.(2001). Systemic acquired resistance and salicylic acid: current of knowledge. *Eur. J. Plant Pathol.* 107, p. 13–18.

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

Poupard P; Parisi L; Campion C; Ziadi S; Simoneau P; (2003). A wound- and ethephon-inducible PR-10 gene subclass from apple is differentially expressed

during infection with a compatible and an incompatible race of *Venturia inaequalis*. *Physiological and Molecular Plant Pathology* 62 3–12).

Parisi L; Lespinasse Y; Guillaumes J; Kruger J (1993) A new race of *Venturia inaequalis* virulent to apples with resistance due to the *Vf* gene. *Phytopathology* 83:533-537

Sato F; Kitajima S; Koyama T; Yamada Y (1996) Ethylene-induced gene expression of osmotin-like protein, a neutral isoform of tobacco PR-5, is mediated by the AGCCGCC cis-sequence. *Plant Cell Physiol* 37: 249-255

Sierotzki H; Gessler C.(1998). Inheritance of virulence of *Venturia inaequalis* toward *Malus × domestica* cultivars. *J. Phytopathol.* 146:509-514.

Solano R; Stepanova A; Chao Q; Ecker J.R (1998) Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ethylene-response-factor1. *Genes Dev* 12: 3703-3714

Tournier B; Sanchez-Ballesta M.T; Jones B; Pesquet E; Regad F; Latche´ A; Pech J.C; Bouzayen M. (2003). New members of the tomato ERF family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. *FEBS Letters* 550, 149–154.

Valdebenito-Sanhueza R. M; Betti J.A. (2005) Doenças da macieira. In: Kimati, H.; Amorim L.; Bergamin Filho A.; Camargo, L.E.A.; Rezende, J.A.M. (Eds.) *Manual de fitopatologia: doenças das plantas cultivadas*. 4. ed. São Paulo: Agronômica Ceres. v.2, p. 422-433.

Wakasa Y; Kudo H; Ishikawa R; Akada S; Senda M; Niizeki M; Harada T. (2006). Low expression of an endopolygalacturonase gene in apple fruit with long-term storage potential. *Postharvest Biology and Technology* 39, 193–198.

Wang A; Tan D; Takahashi A; Li T.Z; Harada T. (2007). MdERFs, two ethylene response factors involved in apple fruit ripening. *J. Exp. Bot.* 58, 3743– 3748.

Williams E.B; Kuc J. (1969) Resistance in *Malus* to *Venturia inaequalis*. *Ann. Rev. Phytopathol.* 7:223-246.

Zegzouti H; Jones B; Frasse P; Marty C; Maitre B; Latch A; Pech J.C; Bouzayen M (1999) Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. *Plant J* 18: 589-600

CONSIDERAÇÕES FINAIS

As análises de bioinformática nos forneceram informações preliminares das seqüências ERF em *Malus* para estudos posteriores mais aprofundados em relação aos mecanismos de sinalização na cultura da maçã.

A expressão do gene *MdERF1* é induzida por patógenos, sendo que essa resposta pode estar associada indiretamente aos mecanismos de resistência da planta moduladas por mudanças metabólicas, tais como a indução de proteínas PR. Com relação ao *MdERF2* não foi observado a mesma indução tanto no controle como nas plantas infectadas com *Venturia inaequalis*. Esses resultados são muito preliminares, necessitando novos estudos serem realizados para confirmação do resultado.

ANEXOS

Supplemental Table I. Unigene list of AP2/ERF superfamily in Malus

Family	EST	Remarks	Database	website
AP2	Q8VWW8	two complete AP2 domain	Prosit ExPaSy Swissprot	http://br.expasy.org/prosit/
	A6XA55	two complete AP2 domain	Prosit ExPaSy Swissprot	
	PTMx00034.1	one complete AP2 domain and one defective AP2 domain	Plant Transcription Factor Databases	http://plantfdb.cbi.pku.edu.cn/
	PTMx00035.1	one complete AP2 domain and one defective AP2 domain	Plant Transcription Factor Databases	
	PTMx00040.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00047.1	one complete AP2 domain and one defective AP2 domain	Plant Transcription Factor Databases	
	PTMx00071.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00072.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00080.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00081.1	one complete AP2 domain and one defective AP2 domain	Plant Transcription Factor Databases	
	PTMx00082.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00085.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00091.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00097.1	two complete AP2 domain	Plant Transcription Factor Databases	
	PTMx00099.1	two complete AP2 domain	Plant Transcription Factor Databases	
RAV	CV631542	incomplete AP2 domain, partial B3 domain	GenBank	http://www.ncbi.nlm.nih.gov/Genbank/index.html
	CV658804	incomplete AP2 domain	GenBank	
	CN489076	incomplete AP2 domain, divergent B3 domain	GenBank	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=apple
	TC9932	incomplete AP2 domain	TIGR	
	PTMx00073.1	incomplete AP2 domain, complete B3 domain	Plant Transcription Factor Databases	

	PTMx00077.1	incomplete AP2 domain, incomplete B3 domain	Plant Transcription Factor Databases
	PTMx00079.1	incomplete AP2 domain, incomplete B3 domain	Plant Transcription Factor Databases
ERF	A1LM1	a single complete AP2 domain	Prosite ExPaSy Swissprot
	A1LM2	a single complete AP2 domain	Prosite ExPaSy Swissprot
	A5YRQ8	a single complete AP2 domain	Prosite ExPaSy Swissprot
	Q2LMD4	a single complete AP2 domain	Prosite ExPaSy Swissprot
	CO576318	a single complete AP2 domain	GenBank
	CO865301	a single complete AP2 domain	GenBank
	CV997565	a single complete AP2 domain	GenBank
	EB114677	a single complete AP2 domain	GenBank
	EB120281	a single complete AP2 domain	GenBank
	EB135117	a single complete AP2 domain	GenBank
	EB176565	a single complete AP2 domain	GenBank
	TC14438	a single complete AP2 domain	TIGR
	TC4497	a single complete AP2 domain	TIGR
	TC6242	a single complete AP2 domain	TIGR
	PTMx00020	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00021	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00022	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00023	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00024	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00026	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00027	a single complete AP2 domain	Plant Transcription Factor Databases
	PTMx00028	a single complete AP2 domain	Plant Transcription Factor

PTMx00029	a single complete AP2 domain	Databases
PTMx00030	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00031	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00032	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00033	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00036	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00037	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00038	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00041	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00044	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00045	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00048	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00049	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00050	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00054	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00056	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00057	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00058	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00059	a single complete AP2 domain	Plant Transcription Factor Databases

PTMx00060	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00061	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00062	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00063	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00064	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00065	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00066	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00067	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00068	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00069	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00070	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00073	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00074	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00076	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00077	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00078	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00083	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00084	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00088	a single complete AP2 domain	Plant Transcription Factor Databases
PTMx00089	a single complete AP2 domain	Plant Transcription Factor

PTMx00092	a single complete AP2 domain
PTMx00095	a single complete AP2 domain
PTMx00096	a single complete AP2 domain
PTMx00098	a single complete AP2 domain
PTMx00100	a single complete AP2 domain

Databases

Plant Transcription Factor
Databases
Plant Transcription Factor
Databases
Plant Transcription Factor
Databases
Plant Transcription Factor
Databases
Plant Transcription Factor
Databases

Supplemental Table II. Conserved motifs in the ERF family in *Malus*

Group Name	Generic Name	Locus Identifier ^a	Conserved motifs ^b						
group I	MdERF#01	PTMx00041.1	CMI-1	CMI-2	CMI-3	CMI-4			
	MdERF#02	PTMx00050.1	CMI-1	CMI-2		CMI-4			
	MdERF#03	PTMx00057.1	CMI-1	CMI-2	CMI-3				
	MdERF#04	PTMx00088.1				CMI-4	CMI-5	CMI-6	CMI-7
	MdERF#05	PTMx00089.1				CMI-4	CMI-5	CMI-6	CMI-7
group II	MdERF#06	PTMx00026.1	CMII-1	CMVII-2					
	MdERF#07	PTMx00032.1	CMII-1						
	MdERF#08	PTMx00054.1	CMII-1	CMVII-2					
group III	MdERF#09	A5YRQ8	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#10	Q2LMD4	CMIII-1	CMIII-2	CMIII-3				
	MdERF#11	PTMx00030.1	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#12	PTMx00036.1	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#13	PTMx00045.1	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#14	PTMx00059.1	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#15	PTMx00060.1	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#16	PTMx00084.1	CMIII-1	CMIII-2	CMIII-3		CMIII-5*		
	MdERF#17	PTMx00100.1	CMIII-1		CMIII-3				

		1			
	MdERF#41	PTMx00023. CMVIII-1	CMVIII-2	CMVIII-3	
		1			
	MdERF#42	PTMx00024. CMVIII-1	CMVIII-2	CMVIII-3	
		1			
	MdERF#43	PTMx00061. CMVIII-1		CMVIII-3	
		1			
	MdERF#44	PTMx00062.			
		1			
	MdERF#45	PTMx00063. CMVIII-1			
		1			
	MdERF#46	PTMx00064. CMVIII-1	CMVIII-2	CMVIII-3	
		1			
	MdERF#47	PTMx00070. CMVIII-1	CMVIII-2	CMVIII-3	
		1			
	MdERF#48	PTMx00076. CMVIII-1		CMVIII-3 divergent	
		1			
group IX	MdERF#49	A1LM2	CMIX-2		
	MdERF#50	CO865301			CMIX-6
	MdERF#51	TC14438	CMIX-2		
	MdERF#52	PTMx00021. CMIX-1			
		1			
	MdERF#53	PTMx00044.	CMIX-2		
		1			
	MdERF#54	PTMx00073.	CMIX-2	CMIX-3	
		1			
	MdERF#55	PTMx00095. CMIX-1			CMIX-6
		1			
	MdERF#56	PTMx00096.	CMIX-2	CMIX-3	
		1			
group X	MdERF#57	PTMx00027. CMX-1	CMX-2*		
		1			
	MdERF#58	PTMx00028. CMX-1	CMX-2*		
		1			
	MdERF#59	PTMx00038. CMX-1	CMX-2*		
		1			
	MdERF#60	PTMx00048. CMX-1	CMX-2*		
		1			

MdERF#61	PTMx00049. CMX-1	CMX-2*
	1	
MdERF#62	PTMx00056.	CMX-2*
	1	
MdERF#63	PTMx00066. CMX-1	CMX-2*
	1	
MdERF#64	PTMx00074. CMX-1	
	1	
MdERF#65	PTMx00092. CMX-1	CMX-2*
	1	
MdERF#66	PTMx00098. CMX-1	CMX-2*
	1	

^a The locus identifier is defined by database annotation.

^b Sequence of these conserved motifs are summarized in Supplemental Table 3.

* Malus specific motif, divergent from *A. thaliana*

Supplementary Table III. Summary of conserved motifs(CMs) within Malus ERF family in comparison to A. thaliana orthologs

Group I

CMI-1

Sequence name	Start	P-value	Site
PTMx00041.1	210	5.11e-60	FLDFSDSGFD EIGSFGLKFPSVEIDWDAI SKLSES
PTMx00050.1	223	1.32e-59	FLDFSDFE EIGSFGLRKFPSVEIDWDAI SKLANS
PTMx00057.1	102	3.42e-48	TFADPEEPPQ WNETFSLEKYPSEIDWDSI LA
	Multilevel	EIESFGLKYPSEIDWDAI	
	consensus	WNG	F V S
	sequence		

CMI-2

Sequence name	Start	P-value	Site
PTMx00041.1	83	2.22e-29	NGSHIGGDFG EYKPLHSSVDAKLEAICK SMAETQKQDK
PTMx00050.1	98	2.23e-28	LNFPDLRHND EYQPLQSSVDAKLEAICQ NLAETTQKQV
PTMx00057.1	68	6.19e-21	HEDGYYGGGS CFNPLHSSVDAKLQEICQ SLRKTEDIDL
	Multilevel	EYNPLHSSVDAKLEAICQ	
	consensus	Q Q A	I Q S K
	sequence		

CMI-3

Sequence name	Start	P-value	Site
PTMx00041.1	22	1.04e-17	TSFGVNKSMP LGLNQLTPYQIHQIQNQLNHR RSTISNLSPN
PTMx00057.1	22	2.11e-16	TVFGFNQTTS LGLNQLTPYQIHQIQNQLNQR RNIISPNLAP

Multilevel IGLNQLTPYQIHQIQNQLHHR
consensus L N SSS L SEINLP
sequence T T Q

CMI-4

Sequence name	Start	P-value	Site
PTMx00041.1	5	3.21e-16	MADLFGGG HGGELMEALQPFYK SASTSASNPA
PTMx00050.1	9	3.35e-17	SNKTFQQSDP FGGELMEALLPFIK SPSNDSSAFA
PTMx00088.1	11	4.21e-15	LYTCSRSFQD SGGELMDALVPFIK SVSDSPSSSS
PTMx00089.1	5	5.21e-14	MALNMNA YVDEFMEALEPFMK VTSSSSTSNS
		Multilevel	FGGELMEALQPFIK
		consensus	HQD F D E YMT
		sequence	SVE K L T
			Y V Y

CMI-5

Sequence name	Start	P-value	Site
PTMx00088.1	102	9.16e-19	SSLYPQPLLQ QQQSPLQNQQMISF GSSQQQQQQQ
PTMx00089.1	100	1.74e-17	MSQPASPQQ QQQQPLQHQQMISF GASPPCSTTQ
		Multilevel	QDPSPLQQQQMISF
		consensus	NQQQGQGHK
		sequence	TM RN

CMI-6

Sequence name	Start	P-value	Site
PTMx00088.1	97	1.53e-27	MISFSPQQQQ QQQQYMAQYWSDTLNLSPRGRMMMM MSQEAVQPYI
PTMx00089.1	81	4.25e-24	TTQFVVPENA QQQQMLLRYWSEALNLSPRGGPGGV PPWLYQQLLR
		Multilevel	QQQQHLLQYWSDALNLSPRGRMGGM

consensus E RLMAR REI K P GPMMA
sequence M T S VR V
Y

CMI-7

Sequence name	Start	P-value	Site
PTMx00088.1	134	2.72e-19	GGMWVAPDES WFSTWGPSSFWDDY DMDSARGLFL
PTMx00089.1	112	9.31e-17	TGWAEAWFNA IPEEWGPGSPLWDDY HFPISNHKDD
	Multilevel	IFSAWGP GSPLWDDY	
	consensus	WPEE E SF YD	
	sequence	T G V L	
		T	

Group II

CMII-1

Sequence name	Start	P-value	Site
PTMx00026.1	102	1.57e-19	GGGGVNGGGD MSAAYIRRKAAEVGAQVDA LEAAGAGG NR
PTMx00032.1	131	2.70e-14	ELLPCTSAED MSAATIRKKATEVGAQVDA IGATVVQNNK
PTMx00054.1	102	1.57e-19	LPGEKFSD ED MSAATIRKKATEVGAQVDA LGTAVQNNRH
	Multilevel	LSPASIQK KAAEVGAQVDA	
	consensus	M AK RRA SRAANS D	
	sequence	EV T F A	

CMII-2(EAR-motif)

Sequence name	Start	P-value	Site
PTMx00026.1	83	4.44e-12	YEKIENCRKS GNGSLERVDLNKLPDPENS DDDDECVKRR
PTMx00032.1	111	1.43e-08	ECSSKEGFKR CNGSLERVDLNKLPDPETSD DD
PTMx00054.1	83	4.44e-12	RRVFSQKRDF GGGELLEVDLNKLPDPENLD DDLVGK
Conserved amino acids		D[L/M]N[K/Q]P[D/E]	

CMII-3(LWSY motif) - absent

Group III

CMIII-1

Sequence name	Start	P-value	Site
A5YRQ8	118	7.94e-22	LRGRSACLNF ADSAWRLRIPESTCAKDIQAAAAEALAF QDEMCDATTD
Q2LMD4	115	4.01e-21	LRGRSACLNF ADSAWRLRIPESTCAKDIQAAAAEALAF QDETCDTTTT
PTMx00030.1	90	7.50e-21	IKGSSAILNF PELADFLPRPVSLSQQDIQAAAAEALMD FKTVPFHLQD
PTMx00036.1	142	7.35e-21	IKGSSAHLNF PELAYHLRPASADPKDIQAAAAAAAV AIDMDVETSS
PTMx00045.1	123	4.95e-21	IKGTAAILNF PELADSFPRPVSLSPRDIQTAALKAAHME PTTSFSSSTS
PTMx00059.1	112	4.74e-21	IKGSSAVLNF PELASLFPRPASSSPHDIQTAAAEAAAMV VEEKLLEKDE
PTMx00060.1	131	1.23e-21	IKGNSGFLNF PELSGLLPRPVSCSPKDIQAAATKAAEAT TWHKPVIDKK
PTMx00084.1	89	3.40e-20	IKGASAILNF PDLAGSFPRPSSLSPRDIQVAALKAAHME TSQSFSSSSS
PTMx000100.1	72	1.99e-12	IKGSSAHLNF PELAYHLRPASADPKDIQAAAAAAVD WKAPESPSST
Multilevel		PELAWSLPRPESTSPKDIQAAAAEAAAMF	
consensus		ADS GR I A S AR R K LA	
sequence		V	

CMIII-2

Sequence name	Start	P-value	Site
A5YRQ8	84	7.51e-38	TMVEAIYTPE QSEGA FYMDEETMFGMPTLLDNMAEGMLLPPP SVQWNHNYDG
Q2LMD4	81	1.75e-36	TLVEAIYTPE QSQDA FYMDEEAMLGMSLLDNMAEGMLLPSP SVQWNYNFDV
PTMx00030.1	56	7.23e-35	VREGERRAEE QNGGV FYMDDEALLGMPNFFENMAEGMLLPPP EVGWNHNDFD
PTMx00036.1	108	3.63e-32	TLVEAIYTAE QSENA FYMHDEAMFEMPSLLANMAEGMLLPLP SVQWNHNHEV
PTMx00045.1	89	1.42e-29	CHVDNDVGSS SWCGTEFMD EEEVLNMPNLLANMAEGMMVAPP SWMGSRPSDD
PTMx00059.1	110	8.20e-27	KEEELEEVS C NSCSLEFMDEEAMLNMP TLLTEMAEGMLMSPP RMMIHPTMED
PTMx00084.1	55	7.23e-25	GVACMNSSYD NTNMG FMD EDLVFDMPNVL MNMAEGMLLSPP RPTVFDAAYD
PTMx000100.1	38	2.55e-23	KKEDKVYDTA ESSTM GFVDEEELLNMPGLLADMAKGMMVAPP WMGSPPSDDS
	Multilevel	QSCGAFFMDEEAMLGMPNLLANMAEGMLLPPP	
	consensus	N E EY D ELFN S D MVA	
	sequence	G V T S	

CMIII-3

Sequence name	Start	P-value	Site
PTMx00036.1	127	6.13e-20	DHRSPVSDSS ECSPKLASSCPKKRAGRKKFR ETRHP[AP2/ERF domain]
PTMx00045.1	102	2.61e-20	MN NDDIILAEMRPKKRAGR RVFK ETRHP[AP2/ERF domain]
PTMx00030.1	99	2.15e-19	DYESPVSSGG DYSPKLATSCPKKPAGRKKFR ETRHP[AP2/ERF domain]
PTMx00084.1	98	2.12e-19	HPVYRGVRKR RWGKWVSEIREPRKK SRIWLGSFPV[AP2/ERF domain]
PTMx000100.1	103	3.31e-18	ME NDDITVAEMKPKKRAGR RIFK ETRHP[AP2/ERF domain]
Q2LMD4	74	2.58e-18	DYESSVSSGG DYIPTLASSCPKKPAGRKKFR ETRHP[AP2/ERF domain]
A5YRQ8	76	7.21e-16	FGSDYEPQGG DYCPTLATSCPKKPAGRKKFR ETRHP[AP2/ERF domain]
PTMx00059.1	92	2.72e-15	HPVYRGVRRR NNNKWVCEMREPNKK KSRIWLGTYP
	Multilevel	DYDPTLAESCPKKPAGRKKFR	
	consensus	NDSIK SM R R K	
	sequence	T	

CMIII-4(LWSY motif) - absent

CMIII-5

Sequence name	Start	P-value	Site
A5YRQ8	151	3.91e-11	LVMDLSRTES EELGEIVELPSLG ASYDVDSANL
PTMx00030.1	164	2.29e-11	SLVSSSATGS EELGEIVELPSLG SSYDGLTQLG
PTMx00036.1	156	1.93e-10	PSSEEAATA EELSEIVELPSLE TSYDESLSEF
PTMx00045.1	100	1.23e-10	SSSSSSSMLS GELGDIVELPSLE NNVKYDCALY
PTMx00059.1	123	5.45e-10	VTSIEESTVS DDLDEIVKLPSLG TSLNESNEFV
PTMx00084.1	121	1.65e-10	TTSIEESTVS DDLDEIVKLPSLG TSLNESNEFV
PTMx000100.1	105	9.88e-12	SYVAAESED EERLEKIVELPNIE EGSYDESVTS
	Multilevel	EELGEIVELPSLE	
	consensus	G	
	sequence		

Group IV

CMIV-1

Sequence name	Start	P-value	Site
PTMx00037.1	21	4.90e-12	LFQHSKRFVSEILVV ETECESSAGT ATCTFRGVRQ
PTMx00083.1	34	3.21e-11	KVPAKGSKKG CMKGKGGPDN SHCSFRGVRQ
	Conserved amino acids	[K/R]GKGGPxN	

CMIV-2 - absent

CMIV-3

Sequence name	Start	P-value	Site
PTMx00037.1	193	6.04e-08	NDQTSWDPNE CFDINELLGDLN EPGPHQSQDQ
PTMx00083.1	187	2.21e-07	QEPLEFGVDE TFDINELLGILN DNNVSGQETM
	Multilevel	CFDINELLGDLN	
	consensus	M VD RI	
	sequence	T	

Group V

CMV-1

Sequence name	Start	P-value	Site
PTMx00031.1	84	5.12e-48	EINSALRSPK SLSELLNAKLRKNCKDQTPYLTCLRLDNDSSHIGVWQKRAGSKTSPNWV
KLVELGDKVN			
PTMx00058.1	86	4.64e-48	DVNSPLMSPK SLSELLNAKLRKCKDLTPSLTCLRLDTDSSHIGVWQKRAGSKTSPTWV
MRLELGNVVN			
PTMx00065.1	85	2.10e-44	STMSSSTSSS SLSSILSAKLRKCKSPSPSLTCLRLDTASSHIGVWQKRAGSKSDSSWV
MTVELGPASS			
	Multilevel	SLSELLNAKLRKCKDLTPSLTCLRLDTDSSHIGVWQKRAGSKTSPNWV	
	consensus	SI S N SPS Y NA SDSS	
	sequence	S Q	

CMV-2

Sequence name	Start	P-value	Site
PTMx00031.1	163	5.23e-20	MKVRNEDVQE DDQMAMQMIEELLNW TCPGSGSIAQ
PTMx00065.1	150	8.28e-19	EEEEEAISD EDQLAMEMIEELLNW S
	Multilevel	EDQHAMQMIEELLNW	
	consensus	DEDLIEE HT	
	sequence	KM L Y	
		V	

CMV-3

Sequence name	Start	P-value	Site
PTMx00065.1	131	1.05e-09	ARTNFAPRIS PDCPLAVRIRGILHHKKLKK ARSAAAATAG
	Multilevel	PDSPLALRIRGILHHKKLKK	
	consensus	NNCQ SSK NL TQ QSM	
	sequence	V T	

Group VI

CMVI-1

Sequence name	Start	P-value	Site
PTMx00077.1	12	8.64e-18	METEEKVSLP RILRISVTDPTYATDSSSDDEEEE VDFDALSTKR
PTMx00078.1	13	3.91e-15	VSELTNTHQT RILRISVTDPDATDSSSDDEEEE EHQR FVSKRR
	Multilevel	RIVRISVTDPDATDSSSDDEEEE	
	consensus	KKL IFN Y SDD	
	sequence		

CMVI-2

Sequence name	Start	P-value	Site
PTMx00077.1	21	4.90e-12	SDEDEEDFLF PRRRVKRFVNEITVEP SCNNVVTGVS
PTMx00078.1	25	3.17e-12	SDDDDNNNVT VVPRVKRYVKEIRFCQ GESSSTAAR
	Multilevel	KRRRVKRYVKEIRFDP	
	consensus	P KF N VLPS	
	sequence	L V	

CMVI-3 - absent

CMVI-4 - absent

CMVI-5

Sequence name	Start	P-value	Site
PTMx00078.1	116	6.46e-14	QFDDFGLLDD IQGFEDNGPSALPDFDFADVE DLQLADSSFG
	Multilevel	IIGFECDBGSELPCDFADVE	
	consensus	FQRL DNL A F SEM	
	sequence	S NS Y T	

Group VII

CMVII-1

Sequence name	Start	P-value	Site
TC4497	0	1.47e-34	. MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
EB151494	15	1.47e-34	LKVSTRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
PTMx00068.1	16	1.47e-34	LKVSTRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
EB176565	17	1.47e-34	LKVSTRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
EB120281	17	1.47e-34	LKVSTRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
CO576318	6	1.47e-34	TRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI

EB135117	16	1.47e-34	LKVSTRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
CV997565	16	1.47e-34	LKVSTRGNIN MCGGAIISDFIAAKRGRKLTEKDLWSE LDTISDLLGI
A1ILM1	0	6.91e-29	MCGGAIISDFIAVKRALKLTAEDLWSD LDTISDLLGI
PTMx00067.1	0	5.89e-28	MCGGAIISDFIAPVRSRRLTADYLWPD LKKPSSGKRL
PTMx00029.1	36	4.13e-27	YLTQVCSTSR MCGGAIISDFIPRDRRRGV TASDIWPD SPFAKSNPDT
EB114677	0	2.45e-26	MCGGAIISDFIPPPRSLRV TNEFIWPD LKNKVKASKK
PTMx00033.1	40	9.75e-26	VNSTSFSISR MCGGAIIANFIPRDRRCVTASDIWPD SPFAKFD PDT

Multilevel MCGGAIISDFIPP
consensus V YAAL
sequence WS

CMVII-2

Sequence name	Start	P-value	Site
TC4497	27	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
PTMx00068.1	43	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
EB176565	44	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
EB120281	44	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
CO576318	33	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
EB135117	43	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
CV997565	43	3.90e-51	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
EB151494	42	5.60e-45	KLTEKDLWSE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKVATSDA KPKKATGAAA
A1ILM1	27	2.36e-41	KLTAEDLWSD LDTISDLLGIDYSNSINKQPENHKVVQKPKPSITKVVT SDE KPKQASGSAA
PTMx00020.1	33	1.90e-40	KLTEKDEESE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
PTMx00029.1	32	1.90e-40	LTEKDEEEE LDTISDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
PTMx00033.1	35	1.90e-40	KLTEKDEESS LDTISDLLRRIDHSNSINKQPEDYKVVQKPKPSVTKATSDK PKKATGAAAA
PTMx00067.1	41	1.90e-40	KLTEKKKESE DDDERDLLGIDHSNSINKQPEDYKVVQKPKPSVTKATSDAK PKKATGAAAA
PTMx00068.1	45	3.55e-39	KLTEKDEESE LDTISDLLGIDHSNSINKQPEDYAAAAKPKPSVTKATSDAK PKKATGGRRR

Conserved amino acids [E/A]xxxx[R/K]xx[K/S]

CMVII-3

Sequence name	Start	P-value	Site
TC4497	125	8.97e-33	WLGTYNTAEE AARAYDEAAVRIRGDKAKLNFSQPPSSSL LPPLAPATPP
EB151494	141	8.97e-33	WLGTYNTAEE AARAYDEAAVRIRGDKAKLNFSQPPSSSL LPPLAPATPP
PTMx00068.1	141	8.97e-33	WLGTYNTAEE AARAYDEAAVRIRGDKAKLNFSQPPSSSL LPPLAPATPP
CV997565	141	8.97e-33	WLGTYNTAEE AARAYDEAAVRIRGDKAKLNFSQPPSSSL LPPLAPATPP
EB114677	85	3.28e-30	WLGTYDTAEE AARAYDEAAVRIRGDKAKLNFAQPPSSSP LPSLAPDTPP
A1ILM1	124	3.28e-30	WLGTYDTAEE AARAYDEAAVRIRGDKAKLNFAQPPSSSP LPSLAPDTPP
EB135117	141	2.54e-28	WLGTYNTAEE AARAYDEAAVRIRGDKAKLNXSQPPSSSL LPPLA
PTMx00067.1	148	2.54e-28	WLGTFNTAEE AARAYDAEARRIRGKKAKVNFPEETPCAS AKRSIKENPQ
PTMx00020.1	149	1.66e-27	WLGTFNTAEE AARAYDAEARRIRGKKAKVNFPEETPRAS AKRSVKANSQ
PTMx00033.1	152	1.72e-26	WLGTFNTAEE AARAYDREARKIRGKKAKVNFNEDDHLP AKTYLRNPNT
PTMx00029.1	148	4.51e-25	WLGTFNTAEE AARAYDREARKIRGNKAKVNFNEDDDLS AQTYLKKNP
CO576318	137	4.50e-25	WLGTFNTAAE AARAYDREARKIRGNKAKVNNPNEDDDLS AQTYLKKNP
EB120281	124	2.11e-24	WLGTFNTAEE AARAYDREARKGRGNKAKVNFNEDDDLS AQTYLKKNP
EB176565	131	2.11e-24	TFNTAEE AARAYDREARKIRGGKAKENFPNEDDDLS AQTYLKKNP

Conserved amino acids

TP[D/E][I/Y][T/S]S

CMVII-4

Sequence name	Start	P-value	Site
PTMx00029.1	197	nd	EFGWSDQAPI TPDISS AVINNNNSAL
PTMx00033.1	194	nd	EFGWSDHGPK TPEISS MLVNNNEASF

Conserved amino acids

TP[D/E][I/Y][T/S]S

CMVII-5(LWSY motif)

Sequence name	Start	P-value	Site
PTMx00029.1	154	nd	AADDQSATDI GNLWSYQDSN
PTMx00033.1	231	nd	LSSLFDGGND NGLWS
Conserved amino acids		LWS(I/L/Y)	

CMVII-7

Sequence name	Start	P-value	Site
PTMx00033.1	286	5.00e-55	TKEGKRSLEA IARMEEDDEVQKLSEELMAYESMMKFYQIPYLDGQSTAAPQNPP PQESTLGDDL
PTMx00029.1	282	9.26e-54	TQEGELSLEA IARMEDDEVQKLSEELMAYENMMKFYQIPYLDGQFTATPQNPP PQESTVSDDL
	Multilevel	YQNFSSDQGSNSFGCSDFGWEDD	
	consensus	DMYIL TS DS E SQSNH	
	sequence	EPILN Q	
	FV M		

CMVII-8

Sequence name	Start	P-value	Site
PTMx00033.1	329	3.40e-18	QSTAAPQNPP PQESTLGDDLWSFDDHVSVA S
PTMx00029.1	325	3.09e-15	QFTATPQNPP PQESTVSDDLWSFVDDSVPA S
	Multilevel	LIEGGSDESIDSLL	
	consensus	Y AYD L F	
	sequence	M	

Group VIII**CMVIII-1(EAR motif)**

Sequence name	Start	P-value	Site
A1LM2	209	1.02e-12	EGGMEKRSQL LDLDLNLPPP SEQA
TC1438	194	2.10e-12	EGGAGKISPP LDLDLNLAPP AE
PTMx00024.1	187	7.29e-12	VVRYEGRRVW LDLDLNFPPP PEN
PTMx00022.1	187	3.01e-11	ETVKVEPRRE LNLDLNLAPP VVDV
PTMx00023.1	157	1.36e-09	SPSPVTVRRG LAIDLNEPPP LWL
PTMx00070.1	139	2.16e-08	SAVDCKLRME PDLDLNASP
PTMx00076.1	123	2.16e-08	ASSSRRRNPP FQFDLNFPL DCVDLFNGAD
PTMx00064.1	205	2.16e-08	ASSSRRKTP FQFDLNFPL DGVDLFAGGI
Conserved amino acids			DL [N/S]xxP

CMVIII-2

Sequence name	Start	P-value	Site
PTMx00022.1	175	3.47e-17	PRSPVHVAT KMACGAQSDSDSSSVVDF EGGGEKRSQLL
PTMx00023.1	176	1.17e-14	PFPKIQVKAG MMVFDERSESDSSSVVMD VVRYEGRRVWL
PTMx00024.1	174	2.56e-12	TTKRYPRTPP VVPEDCHSDCDSSSVVID DDDDIASSRR
PTMx00064.1	166	2.56e-12	YYYNMATYPK MTTGCVQSESETSSVDF EGGAGKISPPL
PTMx00070.1	166	2.56e-12	YYYNMATYPK MTTGCVQSESETSSVDF DGDDIASSSR
PTMx00076.1	101	8.43e-09	YYNMATTTKR MMGCGVQSESETSSVDF RDDDIAIKSSR
Conserved amino acids			S[D/E][S/C][D/S][S/T]SS[V/S]Vx[D/F]

CMVIII-3

Sequence name	Start	P-value	Site
TC6242	209	2.58e-17	NIFEPEPASS GLLQEIVQGFLPK PISQHASIPP
PTMx00022.1	221	6.96e-16	SFFPEESSDT GLLQEVVQEFLKK NRGVPPSPPT
PTMx00023.1	173	8.83e-15	SSAAAAAKQV LYLDVFRAGAMK NHQFHPRMRFDHPNFH X
PTMx00024.1	178	8.83e-15	SSAAAAAKQV LYLDVFRAGAMKNHQFHPRMRFDHPNFH QPDFHAAGGA
PTMx00061.1	190	1.96e-10	ASSAAAAANHM LYFDSVFRAGMLKNHQIQRLRFDQANFH QHDIQAAGGG
PTMx00064.1	195	1.96e-10	NSSAAAAANAM LYFDSVFRAGMLKNHQIQRLRFDQANFH DDDIQAAGGG
PTMx00070.1	206	1.96e-10	NIFAAAAANHM GYFGSVLRAGMLKNHQIQRLRFDQANFH
PTMx00076.1	233	3.17e-09	EEHNKSNNS GYLDCIVPNHCLK NNPEATRLTQN
	Multilevel	GLLQEIVQEFLKK	
	consensus	Y DCV PGHCL	
	sequence	N P	

Group IX

CMIX-□

Sequence name	Start	P-value	Site
PTMx00021.1	141	1.86e-13	ASASASSSSR QVFEFEYLDDSVLEELL EGEKPNKGK
PTMx00095.1	192	7.18e-13	STNVSQRQNR EVFEFEYLDDKVLEELL DSEERKR
	Multilevel	EVFEFEDLDDSYLEELL	
	consensus	QLVV Y GAEV	
	sequence	T	

CMIX-2

Sequence name	Start	P-value	Site
A1LM2	1	1.52e-20	. MATPNEVSALWFIEKHLLDE ASPVATDPWM
TC14438	17	1.55e-19	SPSAXHFDMS AHQEQDDSTLELIRQHLLGD FTFTDSFISH
PTMx00044.1	14	1.55e-19	FPFCSHFDMS AHQEQDDSTLELIRQHLLGD FTFTDSFISH

PTMx00073.1	3	1.23e-16	FEF LNMASEASALELIRRHLLNE FTTMDSFITS
PTMx00096.1	26	7.02e-16	SQRLYTKKLL SMALDQVSALELIKHHLLGE LSPVKSFAIE

Multilevel	MATEEESSALDFIQKHLDD
consensus	SFNDVAD EL TQ EE
sequence	PQ Y S R G
	S T

CMIX-□

Sequence name	Start	P-value	Site
PTMx00073.1	53	3.22e-39	RSSSFSSLYP CLTENWGDLPKEDDSEDMVLYGVLRDAVSVGWVPS GSSDTCSYGF
PTMx00096.1	26	4.21e-35	SNISSPHVMP CLTEQWGNLFPQVDDSEDMIVYNSLRDAFSLGWSPD HTFPFQVDDW

Conserved amino acids [N/S][W/N]xxLP[L/F]x[E/V][N/D]DS[E/Q][D/E]MxxY[G/N]x[L/I]x[D/Q][A/S]

CMIX-□- absent

CMIX-5

Sequence name	Start	P-value	Site
PTMx00095.1	208	1.97e-13	FPLRVNSGEP DPVRVTSKRSSSEPSSSSSENGSPKR RKKADGSSPV

Conserved amino acids P[L/P]LSPx[P/S][P/S]

CMIX-6

Sequence name	Start	P-value	Site
PTMx00095.1	15	1.07e-22	QTAPKPANPT RKPALTISLPNKTEWQIFA SSDQAAEKPA

Multilevel	CPLTPSSWKGFWD
consensus	V M
sequence	

Group X**CMX-1**

Sequence name	Start	P-value	Site
PTMx00098.1	72	6.08e-18	NYESSWTEAD MSAMVSALTQVIGTT DDHAAVQPNP
PTMx00066.1	33	6.08e-18	NLTRDQEHA I MVSTLRQVISN TGGDTSSSHY
PTMx00092.1	83	1.44e-17	FLGYRSR SATE MSAMVSALTQVVSGQ RGSDSRGHIG
PTMx00028.1	36	1.85e-14	SARSQHDMRV MVSALTQVIGN QQSKSHDNIS
PTMx00027.1	44	1.85e-14	SARSQYDMRA MVSALTQVIGN QSSSHDNNQH
PTMx00038.1	29	7.88e-14	NLTRDQEHVI MVSALQQVISN VGGDTNSNAW
PTMx00048.1	37	9.34e-14	YLTRDQEHEI MVSALRQVISN SGADDASSSN
PTMx00049.1	4	8.71e-13	MSA MVSALTQVWSA RSQTEAEGAH
PTMx00074.1	33	2.54e-11	QLSPEQE LSV IVSALQHVISG ENETAPCQGF
	Multilevel		MVSALTQVISN
	consensus		I QH VGG
	sequence		R

CMX-2

Sequence name	Start	P-value	Site
PTMx00027.1	63	2.20e-36	ASSSNLIITS VPPPDAGPCPLCGVAGCYGC LLQRNHREVK
PTMx00028.1	54	1.24e-26	TTSSSLIISS VPPPDAGPCPLCGVAGCYGC NNQRPHREVK
PTMx00038.1	23	1.19e-22	WEADNNLTIS VPPPDAGPCPLCGVAGCYGC TLQRPHRKKK
PTMx00048.1	49	5.94e-21	TTDNSTLSTS VPPPDAGPCPLCGVAGCYGC YYSAPHREVK
PTMx00049.1	56	3.59e-21	TSSSTTIITS VPPPDAGPCPLCGVAGCYGC TWAEDHREKK
PTMx00056.1	61	8.39e-20	NSSSLIITL VPPPDAGPCPLCGVAGCYGC YSSRPHREVK
PTMx00066.1	57	4.92e-15	SEA LPPSDAGPCPLCGVTSCYGC AFPQHEEIKQ
PTMx00074.1	72	5.19e-15	TSSSNLLTTS VPPPDAGPCPLCGVAGCYGC TLQRPHREVK
PTMx00092.1	53	7.36e-12	DTNSNAWEAD LPPLNAGPCPLCSVTGCGYGC NFPRHEAIK

PTMx00092.1 54 5.29e-10 SSDSTVISAG VPRLDSDTCQVCRIEGCLGC NYFFAPNQRI

Multilevel	LPPLDAGPCPLCGVTGCGYGC
consensus	M RPNSDT QV RIAS L
sequence	V S S E

CX-3 - absent

Supplementary Table IV. Prediction of the molecular and cellular properties of *Malus* ERF-like sequences.

<i>Malus</i> Sequences	AP2/ERF domain position ^a		Total aa	pI ^b	MW ^c (kDa)	Unfoldability Prediction ^d	Sub-cellular localization Prediction ^e
	beginning	end					
A1LM1	87	148	252	8.30	2.77	0.069 (charge: 0.008, phobic: 0.441)	nuclear
A1LM2	190	251	334	4.98	3.72	-0.040 (charge: 0.054, phobic: 0.418)	nuclear
A5YRQ8	92	119	251	4.91	2.77	-0.036 (charge: 0.036, phobic: 0.413)	nuclear
Q2LMD4	56	116	219	5.14	2.39	0.077 (charge: 0.027, phobic: 0.451)	nuclear
CO576318	95	127	131	9.92	1.44	-0.060 (charge: 0.076, phobic: 0.419)	nuclear
C0865301	89	151	224	4.94	2.54	0.032 (charge: 0.036, phobic: 0.437)	nuclear
CV997565	105	166	271	9.84	2.94	-0.015 (charge: 0.075, phobic: 0.435)	cytoplasm
EB114677	49	109	177	10.07	1.95	-0.118 (charge: 0.080, phobic: 0.400)	nuclear
EB1202281	106	170	179	11.18	1.98	-0.136 (charge: 0.134, phobic: 0.413)	nuclear
EB135117	105	166	175	10.05	1.91	-0.035 (charge: 0.080, phobic: 0.430)	nuclear, cytoplasm
EB151494	105	166	219	9.79	2.41	-0.021 (charge: 0.064, phobic: 0.429)	nuclear
EB17565	106	167	168	10.05	1.85	-0.061 (charge: 0.083, phobic: 0.421)	cytoplasm
TC14438	133	194	267	6.43	2.99	0.018 (charge: 0.015, phobic: 0.425)	nuclear
TC4497	89	149	221	9.34	2.43	-0.030 (charge: 0.036, phobic: 0.416)	nuclear
TC6242	24	81	203	10.25	2.21	0.002 (charge: 0.050, phobic: 0.432)	nuclear
PTMx00020.1	112	171	387	4.70	4.27	-0.017 (Charge: 0.031, Phobic: 0.418)	nuclear
PTMx00021.1	87	146	209	6.54	2.27	-0.038 (Charge: 0.014,	nuclear

PTMx00022.1	22	80	252	9.46	2.70	Phobic: 0.405) 0.038 (Charge: 0.016, Phobic: 0.433)	nuclear
PTMx00023.1	87	145	256	10.43	2.82	0.062 (Charge: 0.062, Phobic: 0.458)	chloroplast
PTMx00024.1	47	105	182	10.42	2.01	-0.039 (Charge: 0.066, Phobic: 0.423)	chloroplast
PTMx00026.1	30	88	160	9.01	1.78	-0.094 (Charge: 0.025, Phobic: 0.388)	nuclear
PTMx00027.1	133	191	274	8.05	3.09	-0.010 (Charge: 0.007, Phobic: 0.412)	nuclear
PTMx00028.1	124	183	207	9.54	2.32	-0.055 (Charge: 0.048, Phobic: 0.411)	nuclear
PTMx00029.1	111	170	346	4.87	3.86	0.005 (Charge: 0.032, Phobic: 0.426)	chloroplast
PTMx00030.1	30	88	186	6.33	2.06	-0.019 (Charge: 0.016, Phobic: 0.412)	nuclear
PTMx00031.1	4	62	183	8.27	2.05	-0.032 (Charge: 0.011, Phobic: 0.406)	nuclear
PTMx00032.1	58	116	254	7.20	2.73	0.119 (Charge: 0.000, Phobic: 0.456)	nuclear
PTMx00033.1	115	174	349	5.83	3.91	0.010 (Charge: 0.011, Phobic: 0.421)	chloroplast
PTMx00035.1	82	140	202	9.20	2.23	0.024 (Charge: 0.015, Phobic: 0.427)	mitochondria
PTMx00037.1	10	69	118	9.13	1.33	-0.049 (Charge: 0.025, Phobic: 0.405)	mitochondria
PTMx00038.1	37	95	187	9.79	2.06	-0.004 (Charge: 0.037, Phobic: 0.425)	nuclear
PTMx00041.1	158	217	342	6.73	3.71	0.049 (Charge: 0.009, Phobic: 0.434)	nuclear
PTMx00044.1	128	188	233	6.46	2.62	-0.007 (Charge: 0.013, Phobic: 0.415)	nuclear
PTMx00045.1	63	121	252	5.07	2.77	0.018 (Charge: 0.032, Phobic: 0.431)	nuclear
PTMx00048.1	132	192	197	5.85	2.21	-0.002 (Charge: 0.025, Phobic: 0.422)	nuclear
PTMx00049.1	148	209	319	7.16	3.52	0.047 (Charge: 0.003, Phobic: 0.431)	nuclear
PTMx00050.1	171	230	354	6.82	3.88	0.070 (Charge: 0.006, Phobic: 0.440)	nuclear

PTMx00054.1	30	89	160	9.24	1.80	-0.086 (Charge: 0.031, Phobic: 0.393)	nuclear
PTMx00056.1	23	82	154	8.94	1.76	-0.099 (Charge: 0.013, Phobic: 0.383)	nuclear
PTMx00057.1	50	109	180	9.94	2.03	-0.091 (Charge: 0.089, Phobic: 0.413)	nuclear
PTMx00058.1	31	89	156	10.07	1.74	-0.027 (Charge: 0.077, Phobic: 0.431)	chloroplast
PTMx00059.1	83	142	225	6.63	2.57	0.014 (Charge: 0.004, Phobic: 0.420)	nuclear
PTMx00060.1	61	129	184	10.24	1.99	-0.031 (Charge: 0.060, Phobic: 0.424)	nuclear
PTMx00061.1	1	45	182	4.53	1.96	0.032 (Charge: 0.044, Phobic: 0.440)	nuclear
PTMx00062.1	34	92	210	8.95	2.26	0.005 (Charge: 0.010, Phobic: 0.419)	nuclear
PTMx00063.1	41	99	168	10.02	1.82	-0.020 (Charge: 0.054, Phobic: 0.426)	nuclear
PTMx00064.1	2	53	139	5.33	1.47	0.039 (Charge: 0.029, Phobic: 0.437)	nuclear
PTMx00065.1	5	63	179	8.71	2.03	0.035 (Charge: 0.017, Phobic: 0.432)	mitochondria, nuclear
PTMx00066.1	86	144	265	5.29	2.95	-0.082 (Charge: 0.034, Phobic: 0.396)	nuclear
PTMx00067.1	111	170	386	4.54	4.30	-0.059 (Charge: 0.047, Phobic: 0.409)	nuclear
PTMx00068.1	104	162	221	9.77	2.43	-0.033 (Charge: 0.063, Phobic: 0.424)	nuclear
PTMx00070.1	21	79	272	9.29	2.90	0.063 (Charge: 0.011, Phobic: 0.440)	nuclear
PTMx00073.1	150	209	211	6.50	2.37	0.051 (Charge: 0.009, Phobic: 0.435)	nuclear
PTMx00074.1	119	175	384	9.10	4.22	0.059 (Charge: 0.016, Phobic: 0.440)	chloroplast
PTMx00076.1	7	65	163	4.44	1.83	0.108 (Charge: 0.049, Phobic: 0.470)	cytoplasm
PTMx00077.1	69	115	185	5.03	2.06	-0.097 (Charge: 0.059, Phobic: 0.400)	nuclear
PTMx00078.1	44	103	228	4.48	2.56	-0.058 (Charge: 0.066, Phobic: 0.416)	nuclear
PTMx00083.1	43	102	176	10.07	2.01	-0.057 (Charge: 0.074, Phobic: 0.416)	nuclear

PTMx00084.1	29	87	188	6.01	2.10	Phobic: 0.419) -0.062 (Charge: 0.027, Phobic: 0.400)	nuclear
PTMx00088.1	132	191	323	6.69	3.58	0.063 (Charge: 0.006, Phobic: 0.438)	nuclear
PTMx00089.1	130	189	315	6.45	3.51	0.059 (Charge: 0.010, Phobic: 0.438)	nuclear
PTMx00092.1	236	296	413	6.46	4.48	0.040 (Charge: 0.007, Phobic: 0.430)	nuclear
PTMx00095.1	138	199	276	8.37	2.96	0.093 (Charge: 0.007, Phobic: 0.449)	nuclear
PTMx00096.1	214	275	303	8.71	3.46	0.014 (Charge: 0.010, Phobic: 0.422)	nuclear
PTMx00098.1	125	183	303	7.67	3.39	-0.054 (Charge: 0.000, Phobic: 0.394)	nuclear
PTMx000100.1	14	71	111	5.54	1.23	0.004 (Charge: 0.045, Phobic: 0.431)	nuclear

^a ERF domain location was determined by multiple sequence alignments using ClustalX.

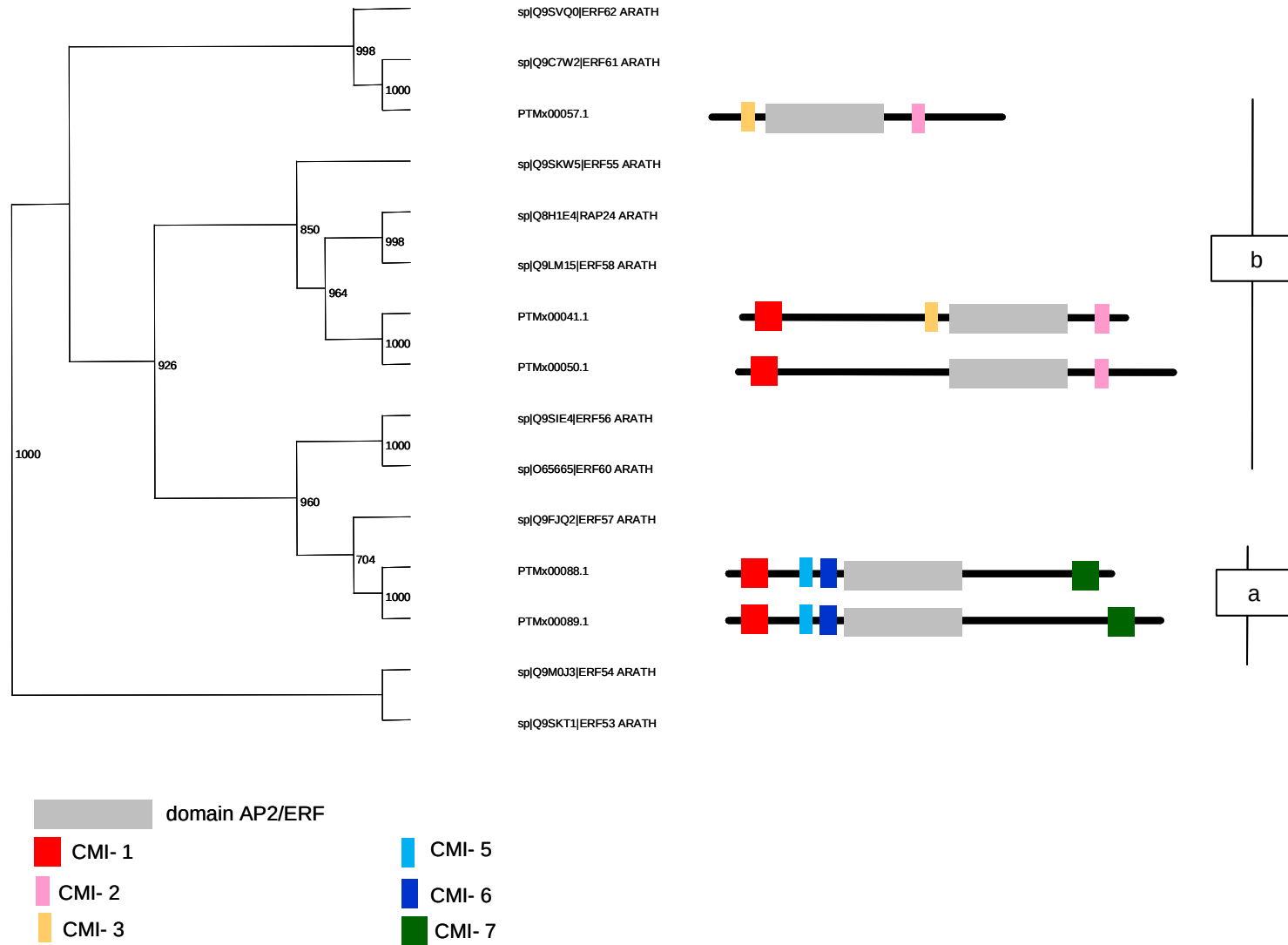
^b Isoelectric point (pI) prediction is represented as pH values and was calculated using Lasergene Editseq software, employing the plant algorithm built in the program (DNASTAR Madison, WI, USA, <http://www.dnastar.com/web/index.php>).

^c Molecular weight (MW) prediction was calculated using Lasergene Editseq software, employing the plant algorithm built in the program (DNASTAR, Madison, WI, USA).

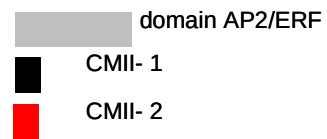
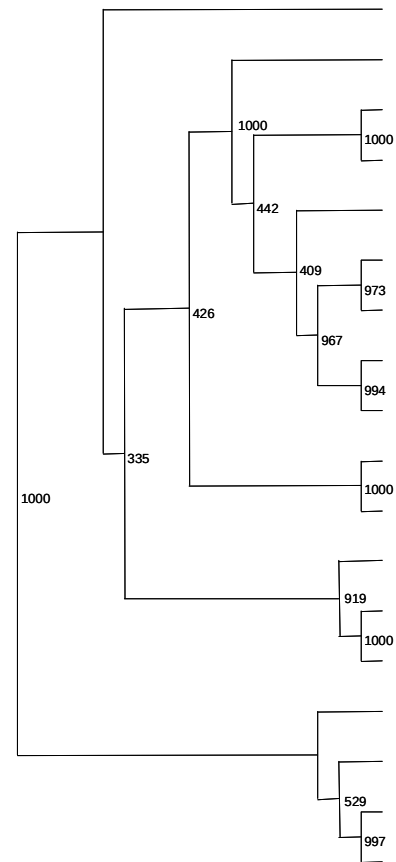
^d The folding status of the deduced amino acid sequences of the apple ERFs was calculated using the algorithm FoldIndex (Prilusky et al., 2005)

^e The prediction the sub-cellular compartment of the deduced amino acid sequences was obtained using WoLF pSORT, using the plant parameters (Horton et al., 2007)

A Group I (A-6)



B Group II (A-5)



sp|Q6NLD5|ERF15 ARATH

sp|Q8LC30|RAP21 ARATH

PTMx00026.1

PTMx00054.1

sp|Q8W3M3|RAP29 ARATH

sp|Q9SNE1|ERF11 ARATH

sp|Q9FH94|ERF10 ARATH

sp|Q9SW63|RA210 ARATH

sp|Q22174|ERF08 ARATH

sp|O80542|ERF19 ARATH

sp|Q9C9I8|ERF20 ARATH

sp|Q9C591|ERF16 ARATH

sp|Q84QC2|ERF17 ARATH

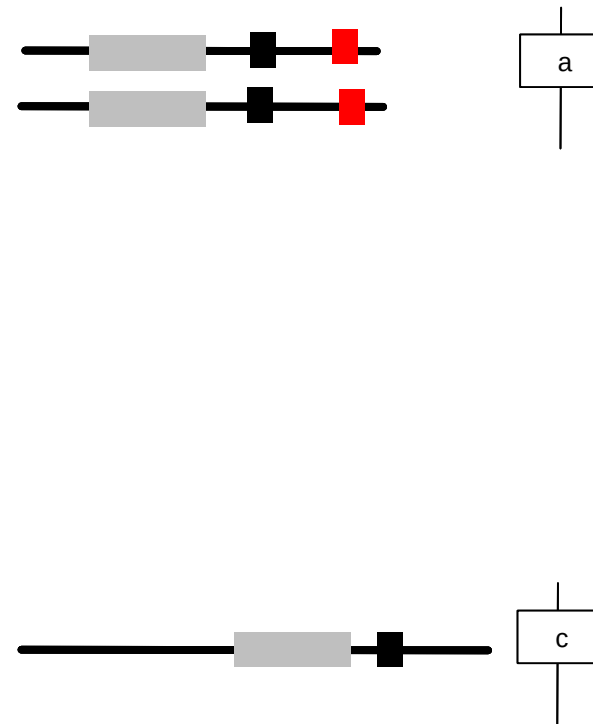
sp|Q9S7L5|ERF18 ARATH

PTMx00032.1

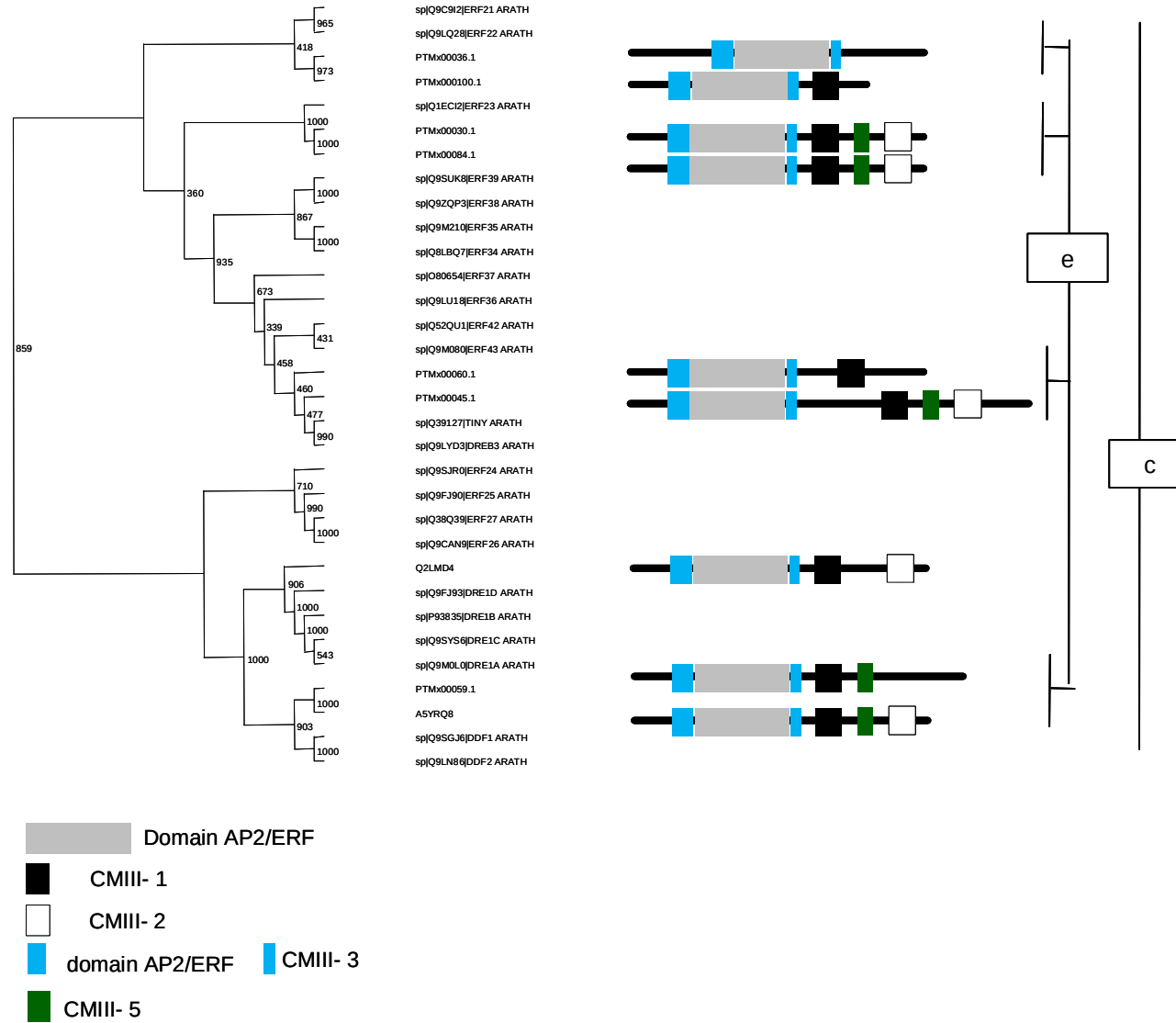
sp|Q9LPE8|ERF14 ARATH

sp|Q9CAP4|ERF13 ARATH

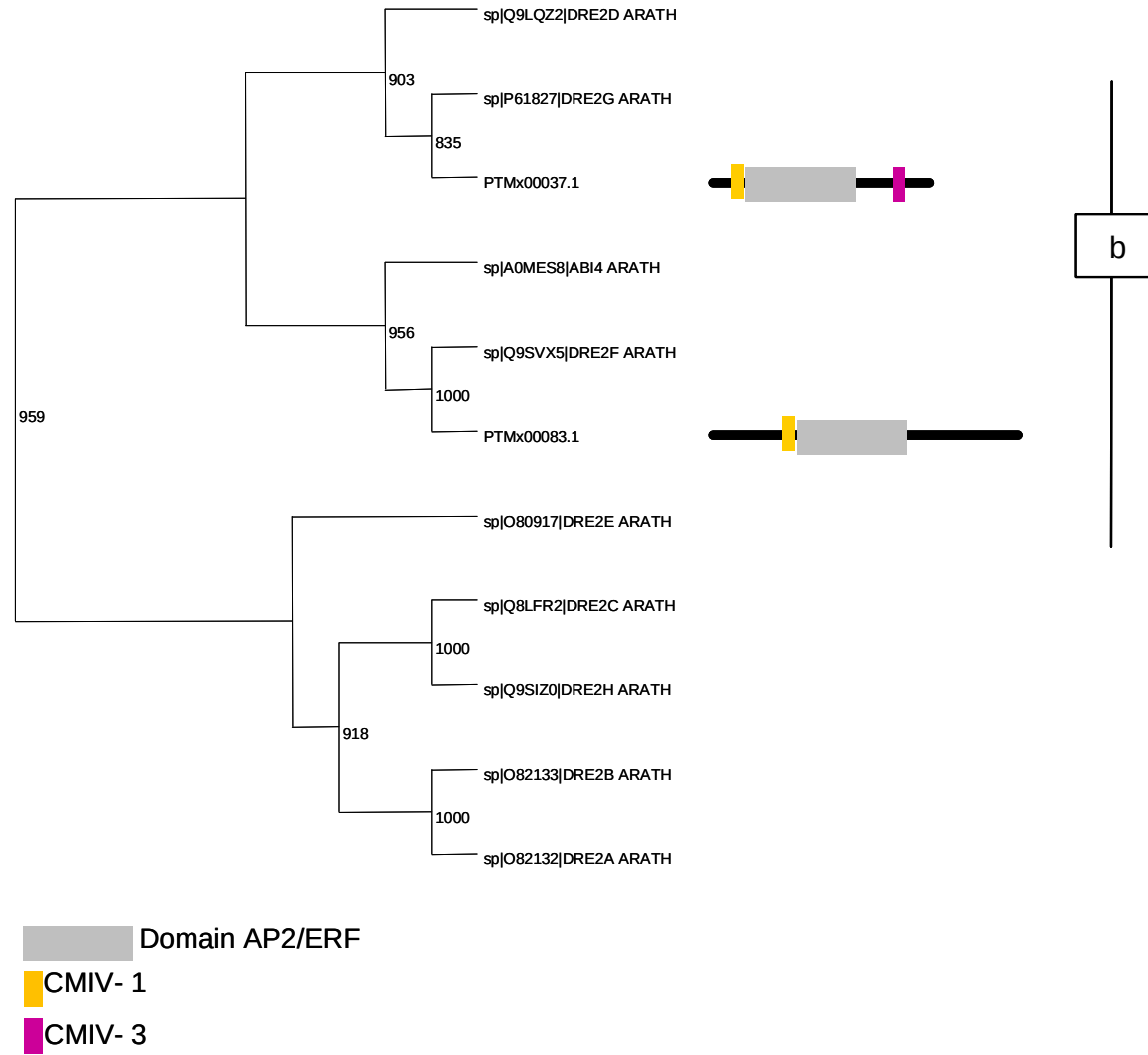
sp|Q9SFE4|ERF12 ARATH



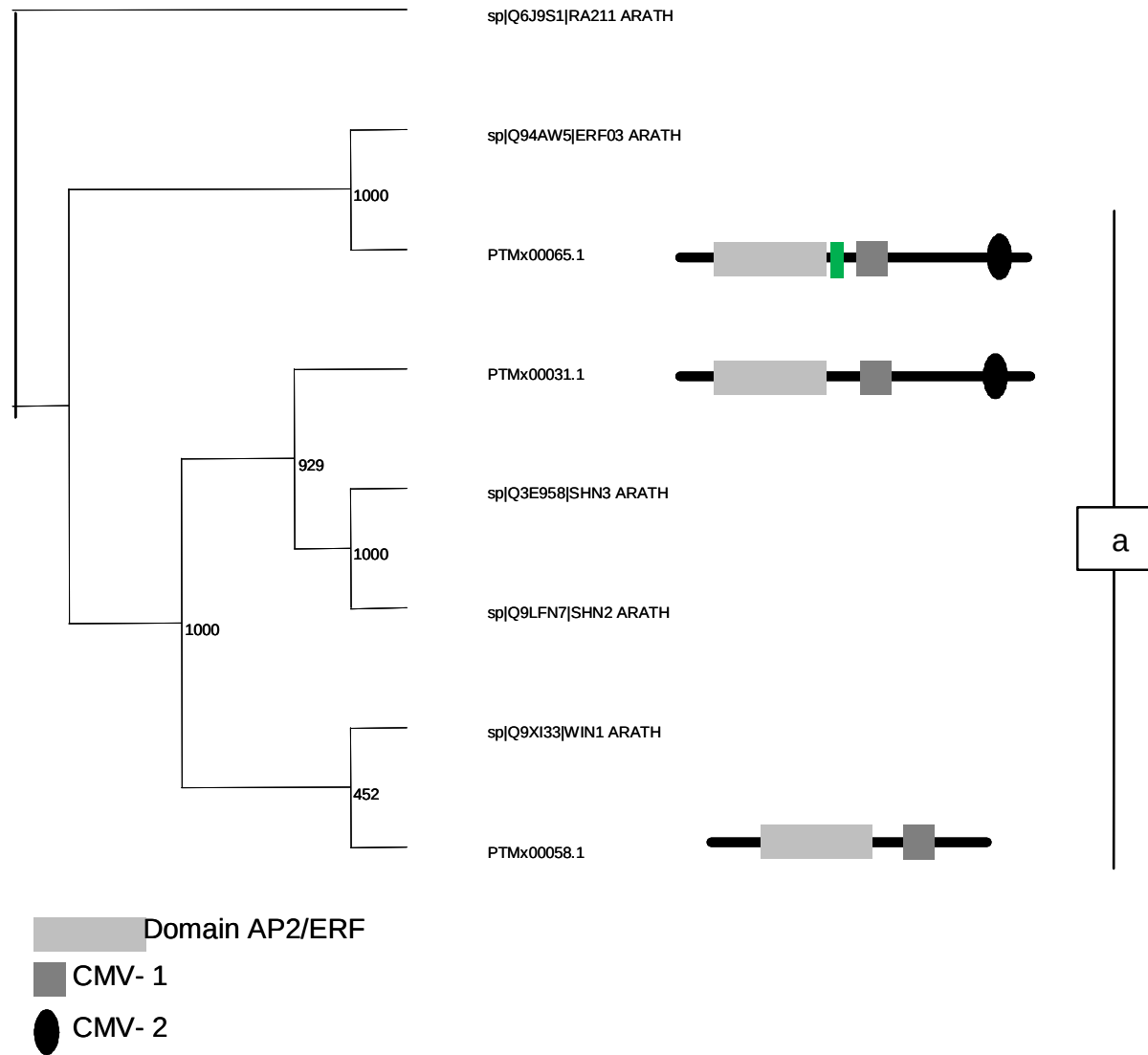
C Group III



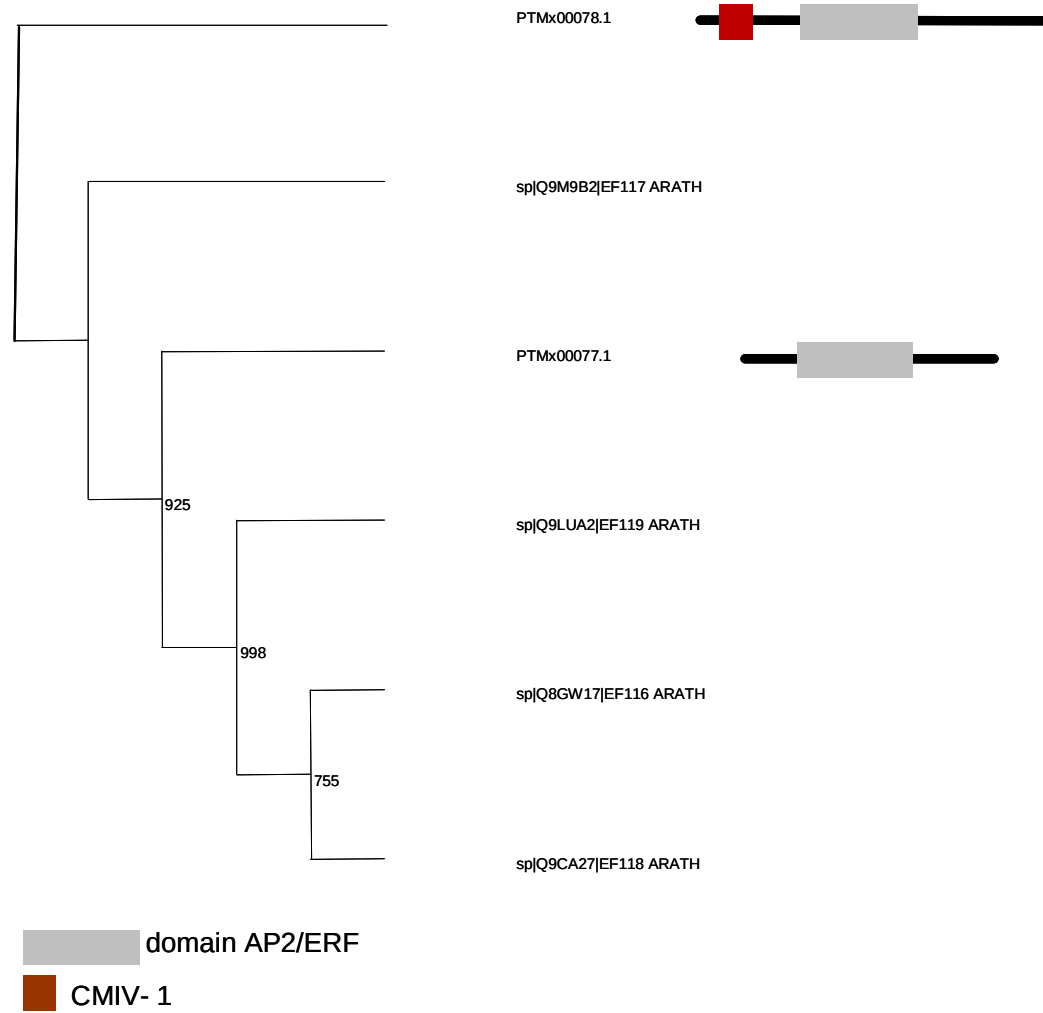
D Group IV (A-2)



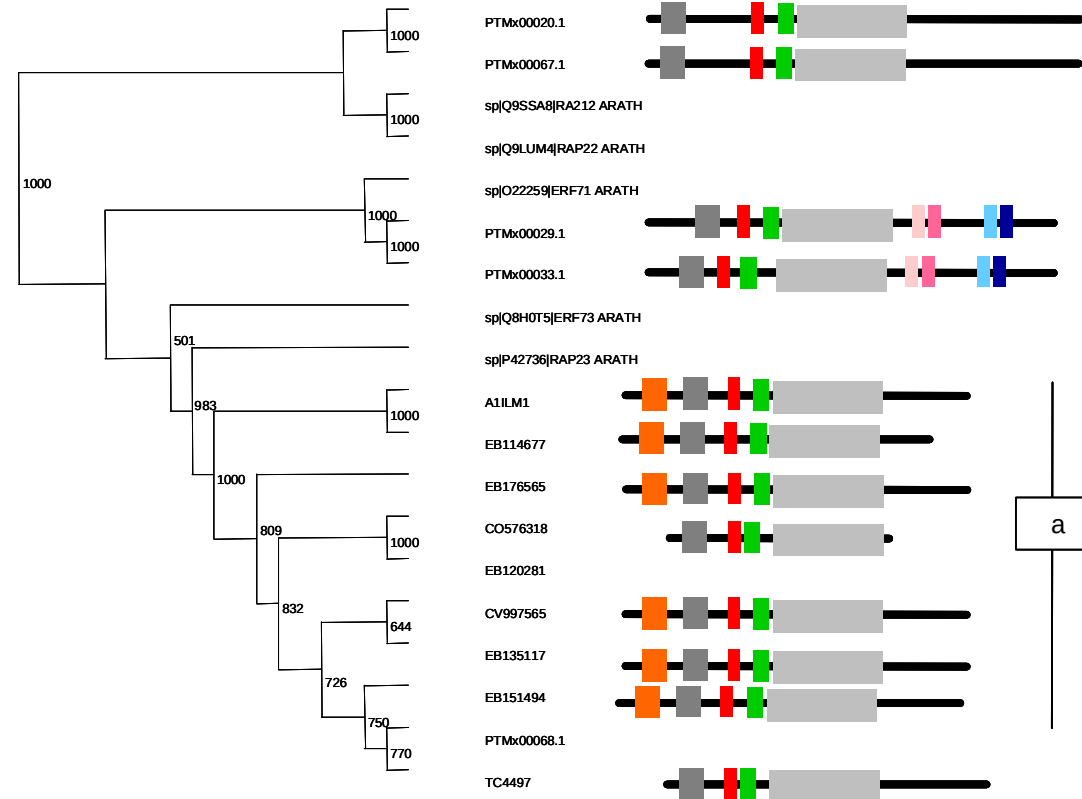
E Group V (B-6)



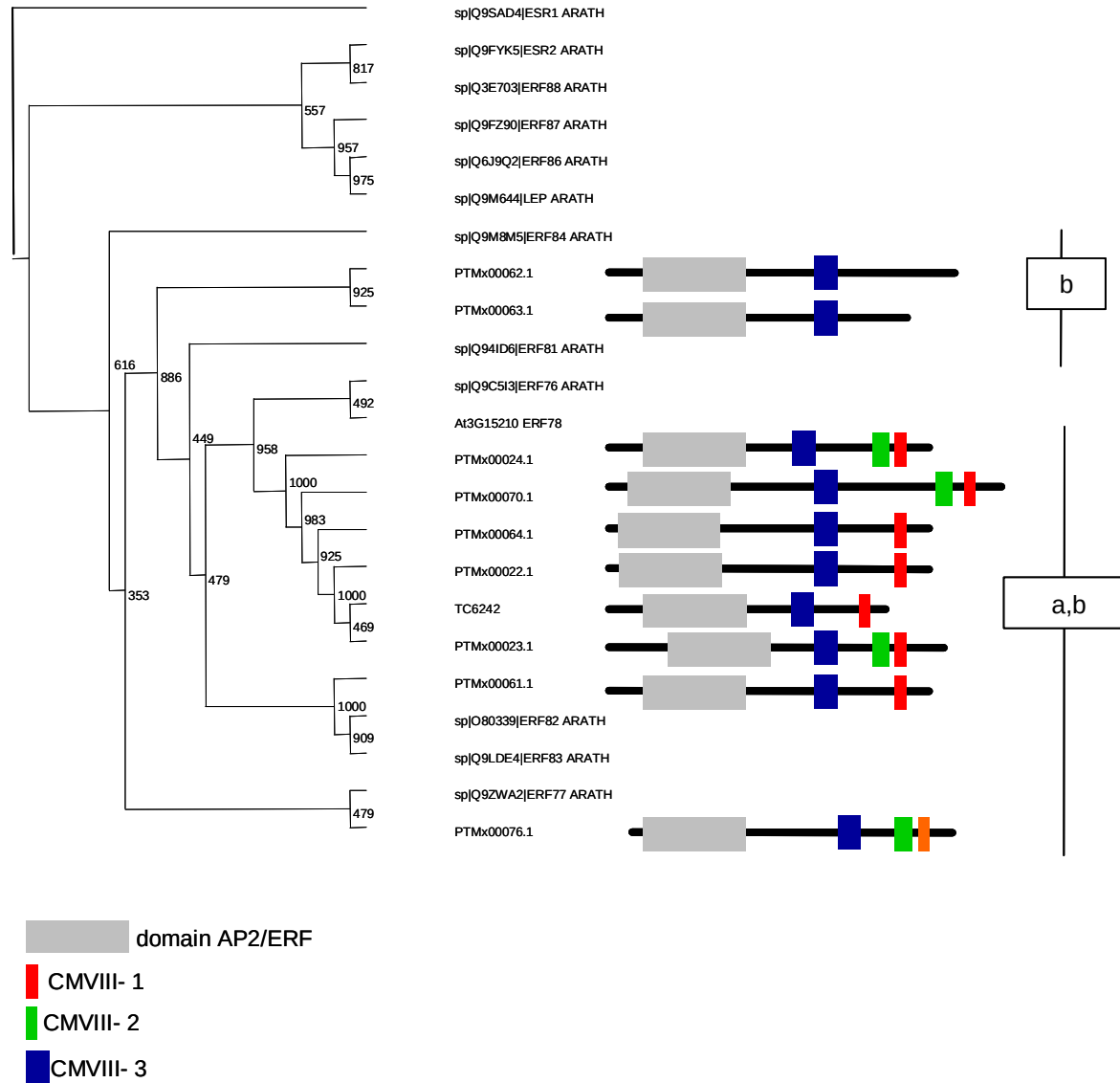
F Group VI-L (B-5)



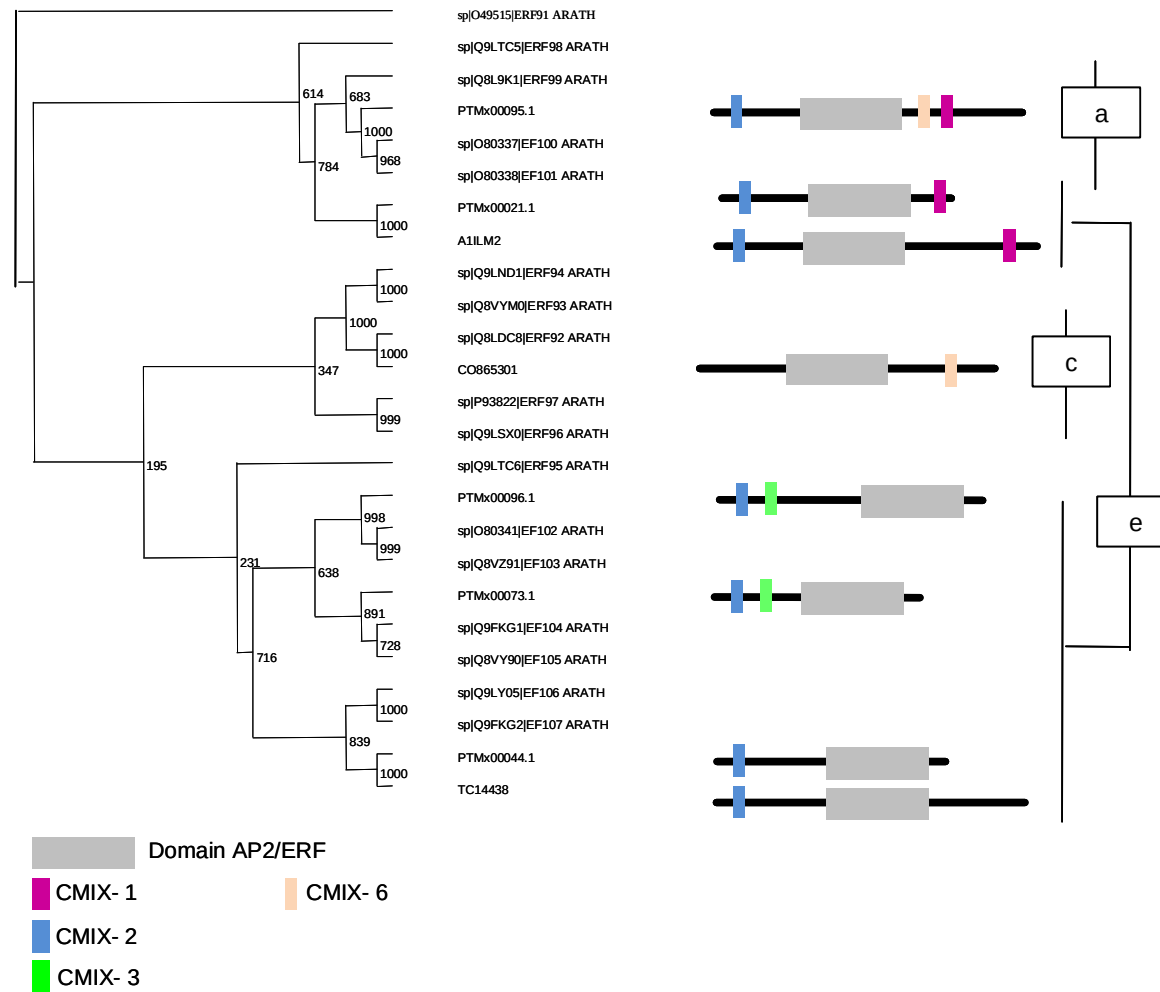
G Group VII (B-2)



H Group VIII (B-1)



I Group IX (B-3)



J Group X (B-4)

