

IDENTIFICAÇÃO E ANÁLISE FILOGENÉTICA INICIAL DE MEMBROS DA FAMÍLIA AP2/ERFS NO TRANSCRIPTOMA DE AVEIA (*Avena sativa* L.)

NATÃ DIENES MACHADO¹, RAILSON SCHREINERT DOS SANTOS², DANIEL DA ROSA FARIAS³, ARTUR TEIXEIRA DE ARAÚJO JUNIOR⁴, MARTINA BIANCA FUHRMANN⁵, ANTONIO COSTA DE OLIVEIRA⁶

¹ Universidade Federal de Pelotas - natamachado@live.com

² Universidade Federal de Pelotas - rschsan@hotmail.com

³ Universidade Federal de Pelotas - fariasdr@gmail.com

⁴ Universidade Federal de Pelotas - arturtaj@hotmail.com

⁵ Universidade Federal de Pelotas – martinabfuhmann@gmail.com

⁶ Universidade Federal de Pelotas - acostol@terra.com.br

1. INTRODUÇÃO

Apesar do enorme interesse nas características nutricionais e nas propriedades promotoras de saúde relacionado ao consumo de aveia (*Avena sativa* L.), o genoma desse cereal ainda é um dos menos explorado entre as culturas de cereais (GUTIERREZ-GONZALEZ et al., 2013). Diversos agentes causadores de estresse em plantas, tanto bióticos quanto abióticos, podem prejudicar severamente a produção de aveia, assim como das demais culturas.

Entre os genes relatados como os responsáveis por parte das respostas adaptativas de plantas à diferentes tipos de estresse estão diversos membros da superfamília gênica APETALA2/*ethylene-responsive element-binding protein* (AP2/EREBP). O domínio AP2/ERF, característico dos genes dessa família, consiste em aproximadamente 60 aminoácidos, os quais estão envolvidos na ligação das proteínas codificadas por estes genes ao DNA. Esta família foi inicialmente identificada no tabaco, e era constituída por quatro proteínas que aparentemente se ligavam especificamente ao GCC Box, uma sequência de DNA envolvida na transcrição de genes responsiva a etileno (OHME-TAKAGI & SHINSHI, 1995), no entanto, tal especificidade foi posteriormente contestada (LICAUSI et al., 2013). Os AP2/EREBPs codificam reguladores transcricionais com uma ampla variedade de funções na regulação de processos fisiológicos vegetais. Um total de 139 ERFs foram identificados em arroz (*Oryza sativa* L. ssp. japonica) e 122 em arabidopsis (*Arabidopsis thaliana* (L.) Heynh.), e classificados em 15 e 12 grupos, respectivamente (Nakano et al., 2006).

O objetivo desse trabalho foi portanto realizar tanto a identificação, como também uma análise filogenética inicial de membros da superfamília AP2/ERFs no transcriptoma de aveia (*Avena sativa* L.).

2. METODOLOGIA

As sequências utilizadas para busca de membros da família AP2/EREBP foram obtidas do banco de dados DDBJ/EMBL/GenBank acesso GAJE01000000, o qual, contém sequências de transcritos obtidas de diferentes estádios de desenvolvimento da semente enriquecida com ESTs (*Expressed Sequence Tags*) do GenBank (fevereiro de 2012), totalizando 71.050 sequências expressas (GUTIERREZ-GONZALEZ et al., 2013). A identificação de membros da família AP2/EREBP foi feita através da busca de domínios “AP2” pelo *Protein Families Database* (Pfam) (BATEMAN et al., 2002), com base nas sequências preditas pelo *Expert Protein Analysis System* (ExPASy) (GASTEIGER et al., 2003). Uma

curadoria manual foi realizada para selecionar somente as sequências proteicas que possuíam o domínio AP2 entre uma “metionina” e um “stop códon”.

A busca de regiões microssatélites (SSRs) nos transcritos identificados com domínios AP2 foi feita utilizando-se o programa SSRLocator (DA MAIA et al., 2008). A análise filogenética foi realizada com o software *Molecular Evolutionary Genetics Analysis 6* – MEGA 6 (TAMURA et al., 2013), com ClustalW (THOMPSON et al., 1994) e *Neighbor-Joining* (SAITOU & NEI, 1987) com 10.000 replicatas de *bootstrap*. A sequência ABG53516.1 da cianobactéria *Trichodesmium erythraeum* IMS101 foi utilizada como *outgroup* e as proteínas “Soloist” de arroz e arabidopsis foram incluídas na análise para conferir se elas se distanciavam das demais, aumentando a credibilidade da análise.

A identificação de motivos conservados foi feita no *Multiple Motif In Elicitation* - MEME (http://meme.sdsc.edu/meme4_4_0/intro.html) versão 4.4.0 (BAILEY et al., 2009), utilizando-se como parâmetro de buscas, um motivos com mínimo 6 e máximo 65 aminoácidos e um número máximo de 10 motivos.

Para identificação das sequências de arroz mais similares, foram utilizados os dados obtidos por Nakano et al., (2006), com o uso do BLASTp (ALTSCHUL et al., 1990). Foram selecionados *hits* de menor e-value para realizar o agrupamento das sequências de aveia e nos grupos obtidos por Nakano et al. (2006).

3. RESULTADOS E DISCUSSÃO

A procura através de *Hidden Markov Models* (HMMs) pelo Pfam conseguiu identificar 44 membros da família AP2/ERF, porém foram selecionados somente 23 destes através de curadoria manual, aliando ExPASy translate com o Pfam.

Nos 44 membros iniciais foram encontradas 18 sequências SSRs em 17 diferentes genes (Tabela 1). Para estes foram construídos *primers* (dados não exibidos) para 16, objetivando avaliar a variabilidade disponível nestes *loci*.

Através da análise filogenética dos 23 membros que passaram por curadoria manual, foi possível formar três diferentes grupos, com base na sua estrutura geral, e considerando motivos conservados obtidos no MEME (Figura 1).

A comparação das sequências obtidas em aveia com as sequências obtidas por Nakano et al. (2006), feita por meio do BLAST é exibida na tabela 2. Levando em consideração os grupos classificados em arroz, considerou-se um e-value máximo de $1e^{-20}$ para uma possível conservação dos mesmos grupos em aveia, e talvez até mesmo uma conservação funcional.

Tabela 1. *Loci* contendo putativos genes contendo domínios AP2/ERF e regiões SSR encontradas nestas sequências.

Locus	(Motivo)repetições
Locus_12831_Transcript_1/2_Confidence_1.000_Length_448	(CGG)5
Locus_1284_Transcript_1/6_Confidence_0.667_Length_1874	(CTA)5
Locus_13624_Transcript_1/1_Confidence_1.000_Length_1026	(CTC)5
Locus_13979_Transcript_1/1_Confidence_1.000_Length_813	(CGAT)6
Locus_14346_Transcript_1/2_Confidence_1.000_Length_1692	(AGTG)4
Locus_14594_Transcript_1/3_Confidence_0.400_Length_847	(TCG)4-(GCC)4
Locus_14594_Transcript_1/3_Confidence_0.400_Length_847	(GGC)4
Locus_14782_Transcript_3/6_Confidence_0.643_Length_1043	(GGC)5
Locus_15383_Transcript_1/1_Confidence_1.000_Length_1275	(GCG)4
Locus_15383_Transcript_1/1_Confidence_1.000_Length_1275	(GCG)4
Locus_18213_Transcript_1/1_Confidence_1.000_Length_1388	(CGC)4
Locus_1973_Transcript_5/7_Confidence_0.588_Length_1331	(GGA)4
Locus_2518_Transcript_1/1_Confidence_1.000_Length_376	(GAG)6

Locus_2518_Transcript_1/1_Confidence_1.000_Length_376	(GAC)4
Locus_4454_Transcript_2/2_Confidence_1.000_Length_1035	(ACC)5
Locus_6154_Transcript_2/4_Confidence_0.400_Length_1271	(GGC)4
Locus_6154_Transcript_2/4_Confidence_0.400_Length_1271	(GAG)5

Tabela 2. Possíveis homólogos dos genes ERF encontrados em aveia no arroz.

Locus em Aveia	Nome genérico	Provável homólogo no arroz
Locus_1284_Transcript_1/6_Confidence_0.667_Length_1874	AsERF#050	OsERF#050_LOC_Os09g20350.1
Locus_14594_Transcript_1/3_Confidence_0.400_Length_847	AsERF#007	OsERF#007_LOC_Os06g07030.1
Locus_13979_Transcript_1/1_Confidence_1.000_Length_813	AsERF#015	OsERF#015_LOC_Os06g09790.1
Locus_1214_Transcript_2/5_Confidence_0.538_Length_1316	AsERF#029	OsERF#029_LOC_Os08g43210.1
Locus_22024_Transcript_1/1_Confidence_1.000_Length_996	AsERF#040	OsERF#040_LOC_Os01g07120.1
Locus_13569_Transcript_1/1_Confidence_1.000_Length_1258	AsERF#042a	OsERF#042_LOC_Os05g27930.1
Locus_5770_Transcript_2/3_Confidence_0.700_Length_1801	AsERF#042b	OsERF#042_LOC_Os05g27930.1
gi 226986458 gb GO593063.1 GO593063	AsERF#003	OsERF#003_LOC_Os02g10760.1
Locus_6154_Transcript_2/4_Confidence_0.400_Length_1271	AsERF#058	OsERF#058_LOC_Os03g60120.1
Locus_14346_Transcript_1/2_Confidence_1.000_Length_1692	AsERF#108	OsERF#108_LOC_Os01g04020.2
Locus_13624_Transcript_1/1_Confidence_1.000_Length_1026	AsERF#060	OsERF#060_LOC_Os03g08460.1
Locus_2518_Transcript_1/1_Confidence_1.000_Length_376	AsERF#061	OsERF#061_LOC_Os05g29810.1
Locus_1340_Transcript_3/4_Confidence_0.571_Length_933	AsERF#067a	OsERF#067_LOC_Os07g47790.1
Locus_12831_Transcript_1/2_Confidence_1.000_Length_448	AsERF#067b	OsERF#067_LOC_Os07g47790.1
Locus_5054_Transcript_1/1_Confidence_1.000_Length_1137	AsERF#068a	OsERF#068_LOC_Os01g21120.1
gi 226983970 gb GO590822.1 GO590822	AsERF#068b	OsERF#068_LOC_Os01g21120.1
Locus_23771_Transcript_1/1_Confidence_1.000_Length_646	AsERF#069	OsERF#069_LOC_Os03g08490.1
Locus_1973_Transcript_5/7_Confidence_0.588_Length_1331	AsERF#070	OsERF#070_LOC_Os02g54160.2
Locus_18651_Transcript_1/1_Confidence_1.000_Length_424	AsERF#071	OsERF#071_LOC_Os06g09390.1
Locus_26934_Transcript_1/1_Confidence_1.000_Length_271	AsERF#075	OsERF#075_LOC_Os01g58420.1
Locus_14366_Transcript_1/1_Confidence_1.000_Length_1465	AsERF#101	OsERF#101_LOC_Os04g32620.1
Locus_13356_Transcript_1/1_Confidence_1.000_Length_422	AsERF#103	OsERF#103_LOC_Os02g52670.1
Locus_3055_Transcript_1/1_Confidence_1.000_Length_1794	AsERF#106	OsERF#106_LOC_Os08g42550.1

*em vermelho: genes que possuíram e-value maior que $1e^{-20}$ quando alinhado ao putativo homólogo.

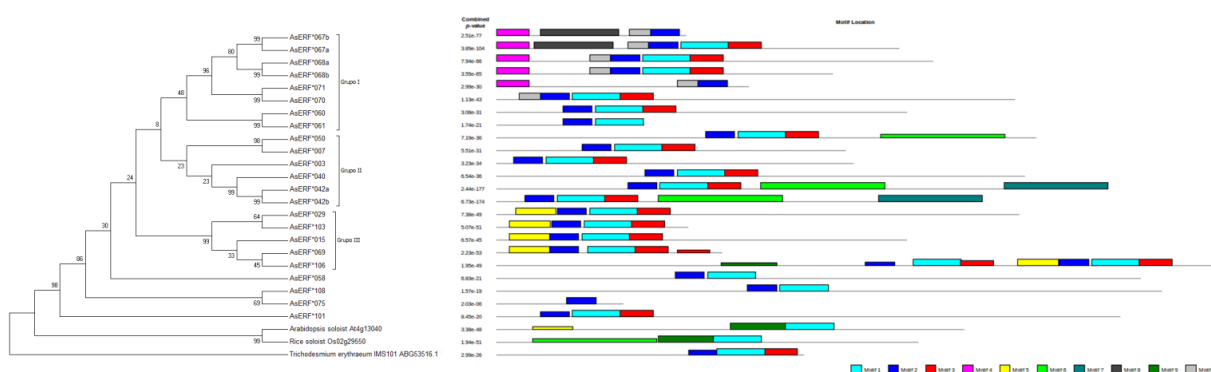


Figura 1. Árvore filogenética obtida e motivos conservados encontrados com base nas proteínas preditas a partir dos transcritos do genoma da aveia disponíveis até o presente momento.

4. CONCLUSÕES

O presente trabalho iniciou a identificação em aveia de genes relacionados à respostas moleculares de vegetais quando submetidas à diferentes tipos de estresse, mas considerando que genoma de arroz apresenta 139 membros (genes AP2/EREBP) classificados em 15 grupos de acordo com Nakano et al. (2006) e que uma classificação mais recente, proposta por Rashid et al. (2012), indica a presença de 170 membros, estima-se que muitos outros genes desta família ainda venham a ser descobertos em aveia.

Mas para que isso ocorra, maiores esforços são necessários para que se incluam mais dados genômicos e de transcritomas. A disponibilização do genoma completo deste cereal deverá contribuir para a realização de outros trabalhos deste gênero.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL S.F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.
- BAILEY T.L. et al. MEME SUITE: tools for motif discovery and searching. **Nucleic Acids Research**, Oxford, W202-208, 2009.
- BATEMAN A. et al. The Pfam protein families database. **Nucleic Acids Research**, Oxford, v. 30, p. 276-80, 2002.
- DA MAIA L.C. et al. SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. **International Journal of Plant Genomics**, 2008, 412696, 2008.
- GASTEIGER E. et al. ExpASY: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, Oxford, v. 31, p. 3784-3788, 2003.
- GUTIERREZ-GONZALEZ, J.J.; TU, Z.J.; GARVIN, D.F. Analysis and annotation of the hexaploid oat seed transcriptome. **BMC Genomics**, v. 14, p. 471, 2013.
- LICAUSI F.; OHME-TAKAGI M.; PERATA P. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. **New Phytologist**, v. 199, p. 639-49, 2013.
- NAKANO, T. et al. Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice **Plant Physiology**, v. 140, p. 411–432, 2006.
- OHME-TAKAGI M.; SHINSHI H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. **Plant Cell**, v. 7, p. 173-82, 1995.
- RASHID M. et al. AP2/ERF Transcription Factor in Rice: Genome-Wide Canvas and Syntenic Relationships between Monocots and Eudicots. **Evolutionary Bioinformatics**, Online. v. 8, p. 321-355, 2012.
- SAITOU N.; NEI M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, p. 406-425, 1987.
- TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution**, v. 30, p. 2725–2729, 2013.
- THOMPSON J.D.; HIGGINS D.G.; GIBSON T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, Oxford, v. 22, p. 4673-80, 1994.