GMR

Identification and characterization of the bZIP transcription factor involved in zinc homeostasis in cereals

A.R. Henriques, D. da R. Farias and A. Costa de Oliveira

Centro de Genômica e Fitomelhoramento, Universidade Federal de Pelotas, Capão do Leão, RS, Brasil

Corresponding author: A. Costa de Oliveira
E-mail: acostol@terra.com.br

Genet. Mol. Res. 16 (2): gmr16029558
Received November 29, 2016
Accepted May 15, 2017
Published June 29, 2017
DOI http://dx.doi.org/10.4238/gmr16029558

ABSTRACT. Members of the basic leucine zipper family, as bZIP19, are considered to be essential regulators of the adaptation to zinc deficiency. Knowing that this gene as well as its targets are conserved in the plant kingdom, we followed an *in silico* approach to identify and characterize the bzip19 gene in cereals. Through BLASTp in Phytozome database, 33 bzip19 genes were identified on the genomes of *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Glycine max*, *Triticum aestivum*, and *Brachypodium distachyon*. The analysis of conserved motifs and bZIP domains was performed using MEME and PFAM databases. In 25 of these genes, CysHis-motifs at the basic N-terminal region were found. This motif is conserved in group bZIP and suggested to play a role as a Zn-sensor. Regarding their phylogeny, it was possible to infer orthologous groups and explore the evolutionary relationship between these BZIP19 proteins. Data mining allowed us to select eight putative orthologous whose expression profile was analyzed under stress conditions in the Genevestigator platform. The comparison between the expression profiles of these eight putative orthologous and the original

Arabidopsis bzip19 also seems to indicate conserved transcriptional regulation. Thus, considering that modified expression of *bZIP19* genes has already been confirmed as an efficient tool to improve *Arabidopsis* tolerance to Zn deficiency, and that these new *bZIP* orthologous have a high level of conservation when compared to the original *bZIP19* sequence, they can be useful for the development of tolerant crops enabling plants to grow in areas of low zinc bioavailability.

Key words: Zinc deficiency; Plant breeding; bZIP; Cereals; Bioinformatics

INTRODUCTION

Zinc is an essential micronutrient responsible for the maintenance of vital processes in all living organisms. It plays a fundamental role in several critical cellular functions such as biomembrane integrity, carbohydrate metabolism, antioxidative defense, and stability of genetic materials (Henriques et al., 2012). However, zinc deficiency is a very common problem recognized as one of the major global challenges to public health and agriculture in the 21st century (Pfeiffer and McClafferty, 2007).

Cereal grains, staple food and principal source of calories for developing countries, show naturally low zinc concentration in grains, especially when grown in Zn-deficient soils (Cakmak et al., 2010). Currently, reports estimate that about half of the cultivated soil in the world contains low amount of soluble zinc (Cakmak, 2008). The low bioavailability of this nutrient limits its uptake by plants, leading to inadequate functioning of many essential physiological cell processes. This condition results in significant decreases in plant growth, yield and zinc content translating in human diet Zn deficiency (Assunção et al., 2013). Hence, techniques that help to unravel the complex network of homeostasis of zinc in plants may contribute to minimize severe yield losses worldwide and alleviate the malnutrition problem in humans.

The utilization of plant genetic resources through gene discovery to achieve cultivars with enhanced nutritive content and/or tolerance to abiotic stress has proven to be an effective and promising approach (Yoshioka and Shinozaki, 2009; Osorio et al., 2012). Some success has been achieved by modulating transcription factors to maximize the expression of candidate or key metabolic genes. The overexpression of *DREB1A* transcription factor has provided an increased tolerance to salt and drought stress in *Arabidopsis* (Kasuga et al., 2004). In rice, *OsIRO2* overexpression, also a transcription factor, involved in the regulation of iron homeostasis genes in *Oryza sativa* L. increases iron deficiency tolerance by improving growth and yield of rice plants (Ogo et al., 2011). Positive results under drought stress in *Camelina sativa* were achieved using *Arabidopsis MYB96* transcription factor (Lee et al., 2014). Therefore, plant improvement strategies aiming to increase Zn concentration and/or efficiency use seem to be effective solutions to enhance zinc accumulation and/or tolerance to zinc deficiency in plants. The first regulators of zinc homeostasis in plants, bZIP19 and bZIP23 transcription factors, have been identified (Assunção et al., 2010). In *Arabidopsis thaliana*, they are essential for proper Zn deficiency response and the modification of these transcription factors can improve plant Zn deficiency tolerance (Song et al., 2010) because stimulate the expression of a set of the target genes, called zinc deficiency response elements. This expression constitutes the primary response to zinc deficiency. bZIP19/23 belongs to the F group of bZIP transcription factors characterized by the presence of a His-rich motif at the

N-terminal of basic region, hypothesized to act as a zinc-sensor playing a role in Zn regulation (Nijhawan et al., 2008; Assunção et al., 2013). Thus, to understand how bZIP19/23 genes as well their targets are conserved in the plant kingdom (Assunção et al., 2010) a comprehensive *in silico* approach to identify existing relationships between different species groups is potentially exciting.

This study provides the identification and functional characterization of bZIP19 genes, furthermore performs a comparative genomic analysis through the phylogenetic relationship of the bzip19 gene in field crops featured using *A. thaliana* as a reference. The use of the Genevestigator (Hruz et al., 2008) platform enabled a complementary analysis of expression profiles of a few bZIP19 genes during the development stage and under stress condition of candidate genes. These findings might shed additional light and contribute for the development of zinc deficiency-tolerant genotypes and high-Zn content plants via genetic engineering.

MATERIAL AND METHODS

Sequence identification and annotation

In order to identify bZIP19 protein orthologous, a search was made using the gene ID from The Arabidopsis Information Resource - TAIR (http://www.arabidopsis.org/) for the bZIP19 sequence. The genomes of *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Triticum aestivum*, *Triticum urartu*, *Glycine max*, and *Brachypodium distachyon* were investigated through BLASTp, with default parameters, at the Phytozome (http://www.phytozome.net/). Searches were also made in the Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn/), being the redundant sequences manually discarded. The putative bZIP19 protein sequences obtained were scanned with the Interpro program using the PFAM database (http://pfam.sanger.ac.uk/) and their bZIP domains deduced. The amino acid consensus sequence of the representative domain of the *AtbZIP19* gene retrieved from the Pfam database was used as query in BLAST2Seq (www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). The potential orthologous genes encoded in the sampled genomes were used as subject and inserted, one by one, in the database. Through BLASTp, identity, coverage, and e-value of sequences were analyzed.

Analysis of conserved motifs

The putative complete sets of bZIPs from brachypodium, sorghum, maize, soybean, wheat, arabidopsis, and rice served as input for a conserved motif analysis performed with MEME (http://meme.sdsc.edu/meme/mee.html) version 4.9. The deduced conserved protein domains were employed. A given motif was allowed to appear at any number of repetitions, the maximum width of a motif was set to 50, and the maximum number of motifs was set to 3. For the other parameters, the default settings were used.

Gene structure - orthologous of the *AtbZIP19*

To obtain exon-intron organization of putative *bZIP19* genes, the full-length genomic DNA sequences were aligned to coding sequences (cDNAs) in the Gene Structure Display Server program (gsds.cbi.pku.edu.cn/).

Sequence alignment and phylogenetic analysis

The phylogenetic analyses of the orthologous *bZIP19* genes predicted in *O. sativa*, *S. bicolor*, *Z. mays*, *B. distachyon*, *G. max*, *T. aestivum* L., *T. urartu*, and *A. thaliana* genomes were performed using conserved protein domain sequences. Multiple-sequence alignments were conducted with protein domain sequences using the CLUSTALW tool (Thompson et al., 1997) implemented in MEGA ver. 6.0 (Tamura et al., 2011). The phylogenetic analysis was performed using a PAM matrix and Neighbor Joining Method. Branch points were tested for significance by bootstrapping with 1000 replications. The unrooted phylogenetic trees of *bZIP19* orthologous were visualized and edited using the FancyGene software.

Gene expression data mining

The *in silico* expression profile of the selected putative orthologous was analyzed at development and anatomical levels under stress condition by retrieving in Genevestigator platform database (https://www.genevestigator.com/gv/plant.jsp), which contains gene expression data from both microarray and RNAseq experiment of genes in different biological contexts. For all plants, only the wild-type background was evaluated.

RESULTS AND DISCUSSION

In silico identification of *bZIP19* orthologous

The *in silico* sequence-similarity search for bZIP19 genes orthologous of *Arabidopsis* identified 33 non-redundant bZIP19 genes in Phytozome and Transcription Factor Database over the genome of *O. sativa*, *S. bicolor*, *Z. mays*, *B. distachyon*, *G. max*, *T. aestivum*, and *T. urartu*. However, as some sequences showed less than 70% similarity not being reliable to infer functional and evolutionary relationships, further analyzes were performed.

Transcription factors AtbZIP19/AtbZIP23 have a conserved bZIP (basic leucine zipper region) domain in all Magnoliophyta species. The domains, which are conserved regions of a protein sequence, can function and exist independently of the rest of the protein chain serving as the basis for the generation of new sequences by nature (Bailey and Elkan, 1994; Lesk, 2002). Based on this understanding, the protein sequences found were analyzed in PFAM (http://pfam.xfam.org/) and through it was possible to confirm the presence of bzip2 family, a domain identified by Pfam 7716, in all the previously selected sequences in database Phytozome and Transcription Factors. The presence of bzip2 family is substantial information that reinforces the idea that the sequences found are putative orthologous of AtbZIP19, since those protein domains are considered functional units of the proteins.

In order to increase the reliability of the selected sequences, an alignment of domain versus protein was performed in Blast2seq (www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). The *AtbZIP19* domain consensus sequence, identified by Pfam database, was used as "query" subject and the previously found protein sequences were inserted, one by one. Through Blast2seq, the identity, coverage, and e-value of sequences were analyzed. According to the results, it was possible to note that most of the sequences showed high similarity and identity (over >70%), good coverage (100-83%) and e-value (10^{-10}) indicating that the sequences were not paired at random, but potential orthologous sequences and thus exercising similar functions (Table 1).

Table 1. Comparison of the sequence identity of AtbZIP19 putative orthologous and their related information.

Species	Sequence ID	Cover	Identity	e-value
Arabidopsis thaliana	AT2G16770.1 (bZIP23)	100	85	9e-29
Glycine max	Glyma11g11540.4	96	80	3e-24
Oryza sativa	LOC_Os06g50310.1	96	77	2e-25
Sorghum bicolor	Sb10g030250.1	96	75	6e-25
Zea mays	GRMZM2G000171	96	75	3e-25
Glycine max	Glyma11g16050.7	96	73	7e-23
Triticum urartu	EMS67651	96	71	2e-22
Triticum aestivum	Tae024846	96	71	8e-22
Triticum aestivum	Tae031862	96	69	6e-21
Brachypodium distachyon	Bradi1g30140.1	96	63	8e-22
Arabidopsis thaliana	AT3G51960.1 bZIP24	96	62	1e-18
Brachypodium distachyon	Bradi2g21197.1	95	75	1e-23
Sorghum bicolor	Sb03g037300.1	94	94	3e-23
Oryza sativa	LOC_Os05g41540.1	94	78	5e-25
Zea mays	GRMZM2G033230	94	78	1e-24
Triticum urartu	EMS45371	94	78	7e-23
Zea mays	GRMZM2G175870	94	76	3e-23
Sorghum bicolor	Sb09g024290.1	94	75	2e-23
Triticum urartu	EMS45140	94	75	2e-23
Oryza sativa	LOC_Os01g58760.1	94	71	3 e-23
Triticum urartu	EMS54685	94	71	4e-21
Glycine max	Glyma12g03690.4	92	76	2e-23
Brachypodium distachyon	Bradi2g52590.1	92	76	5e-23
Zea mays	GRMZM2G055413	85	72	3 e-21
Triticum aestivum	Tae067773	83	76	5e-21
Brachypodium distachyon	Bradi4g04720.1	87	43	3e-07
Triticum aestivum	Tae35200	85	67	1e-20
Triticum urartu	EMS47152	85	61	3e-18
Triticum urartu	EMS52367	83	64	1e-18
Triticum aestivum	Tae034985	81	45	3e-08
Triticum urartu	EMS60905	81	45	1e-07
Triticum aestivum	Tae000909	77	45	3e-07

Analysis of conserved motifs

The bZIP19 and bZIP23 transcription factors from *A. thaliana* belong to the group F of bZIPs, having two characteristic histidine-rich motifs (CysHis-motifs) in the basic N-terminal region found between the position 55-68 and 71-81 (Guo et al., 2007). Among the bZIP putative protein sequences of brachypodium, sorghum, maize, soybean, wheat, arabidopsis, and rice only 25 of the 33 sequences showed histidine-rich motifs, highly stable and located on the motif 3 of sequence logo (Figure 1). The existence of a strong conservation of histidine-rich motifs in bZIP19/23 proteins in different plant species during evolution has been reported (Corrêa et al., 2008). These motifs have been claimed to be involved in zinc homeostasis and functioning as zinc cell sensors (Assunção et al., 2013).

The amino acid alignment of the putative bZIP19 orthologous from different plant species showed highly conserved regions (Figure 2) evidencing a strong conservation of these proteins during the evolution. Thus, the seven protein sequences lacking Cys-His-conserved motifs and also lacking similar alignment were discarded.

Phylogenetic analysis

Through phylogeny and structure of putative *bZIP* genes, it was possible to infer orthology and to explore the evolutionary relationships between predicted bZIP19 proteins.

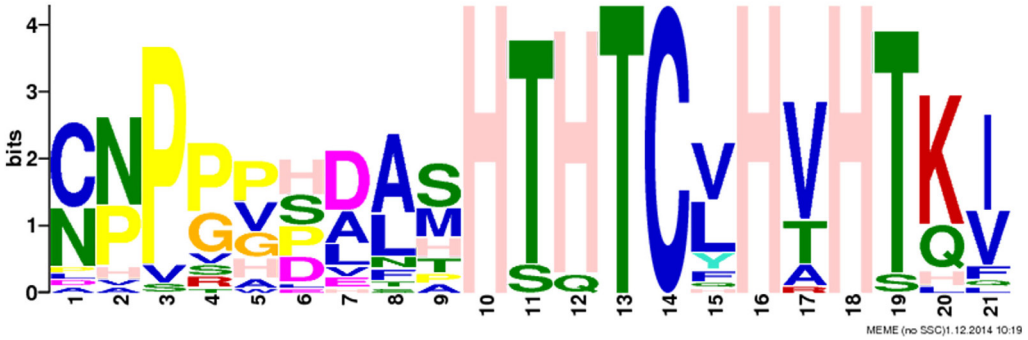

Figure 1. Conserved domain analysis of putative orthologous AtbZIP19 proteins.

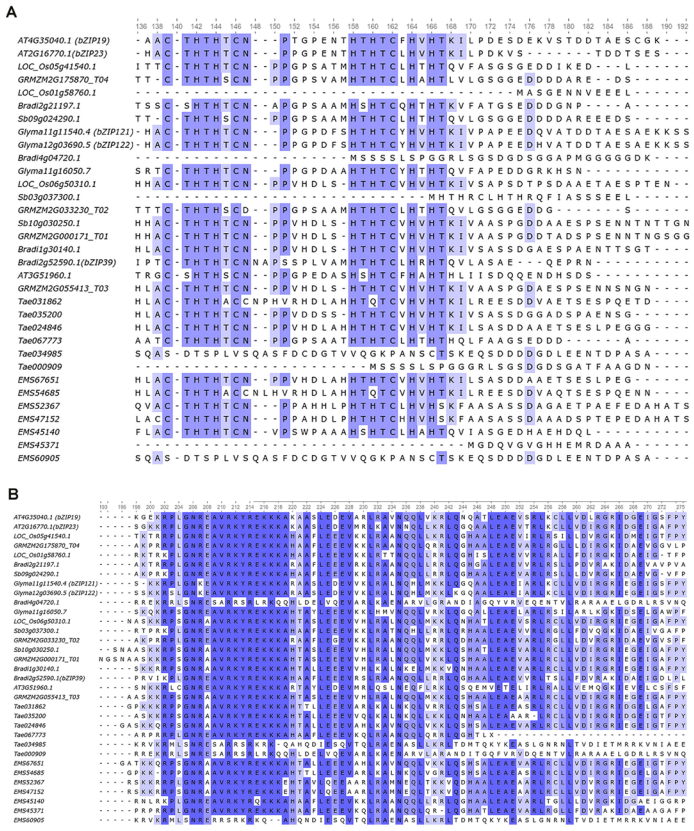

Figure 2. A. Two conserved motifs rich in histidine residues represented by aa 138-147 and 158-167. **B.** Amino acid alignment of bZIP domain of plants with sequence similarity to the *Arabidopsis*. The conserved bZIP domain is represented by aa 200-235. The alignment was generated by ClustalW and displayed with the Fancy software. Identical amino acid residues concerning the AtbZIP19 are shaded in dark blue, amino acids that are conserved among two or three proteins are indicated in blue, and blocks of similar amino acids are indicated in light blue. Species are indicated are as follows: LOC_Os, *Oryza sativa*; GRMZM, *Zea mays*; Bradi, *Brachypodium distachyon*; Sb, *Sorghum bicolor*; Glima, *Glycine max*; Tae, *Triticum aestivum*; EMS, *Triticum urartu*.

The phylogenetic tree formed two stable clusters (Figure 3). In the first cluster, AtbZIP19, AtbZIP23, and similar sequences were grouped, while in the second cluster, AtbZIP24 and similar protein sequences were grouped. AtbZIP24 is also a transcription factor belonging to the bZIP group F, nonetheless it does not seem to be involved in zinc homeostasis but instead with salt stress (Assunção et al., 2010). The phylogenetic tree showed a consistent formation of clades supported by bootstrap analysis. Furthermore, proteins related phylogenetically among monocots and dicots, such as *Arabidopsis* and soybean, showed a close relationship with the genetic structure, which carries the imprint of the evolution of a gene family, validating the results of clustering and evidencing the large scale of expansion of bZIP genes. Based on the premise that similar sequences tend to display similar functions, and therefore, similar structures through mining data, eight putative orthologous genes were selected to profile expression analysis in Genevestigator platform: one from rice (LOC_Os6g50310), three from wheat (Tae 031862, Tae 024846, EMS 67651), one from brachypodium (Bradi1g30140), one from soybean (Glyma11g115404), one from maize (GRZM2g000171), and one from sorghum (Sb10g032501).

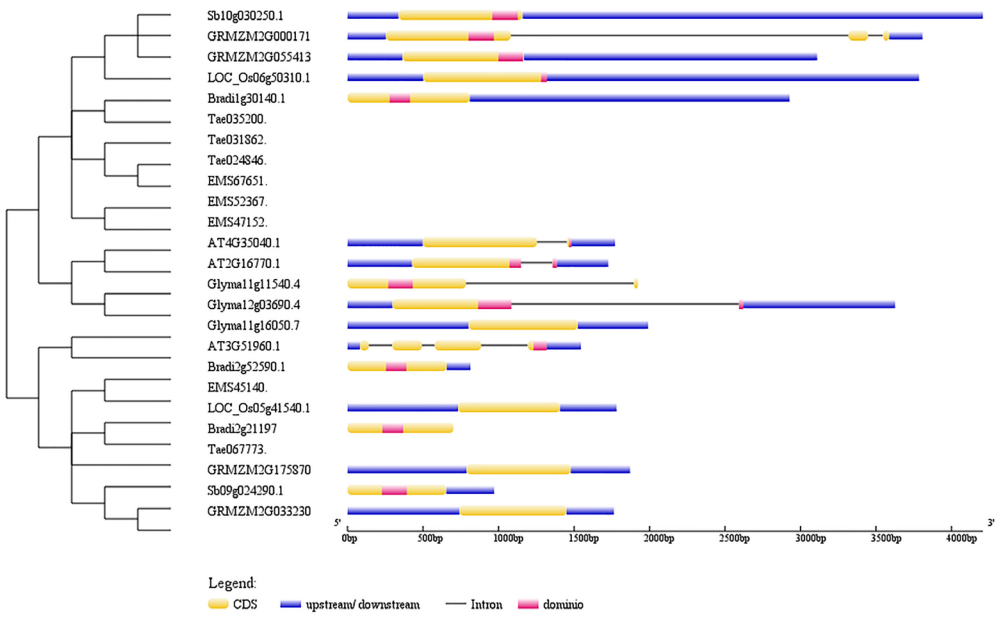

Figure 3. Neighbor joining tree, including distance values, showing the phylogenetic relationship between the *AtbZIP19* and putative *bZIP* orthologous. The gene intron-exon structure was arranged corresponding to the phylogenetic tree.

Functional analysis of the gene expression profile predicted by *bZIP19* Genevestigator

From data mining, it was possible to identify eight *AtbZIP19* orthologous genes. In order to determine whether they are similar to their *Arabidopsis* counterpart and whether they somehow are connected to stress response, an *in silico* analysis of these genes was performed using Genevestigator response viewer. However, the expression profile data in Genevestigator

platform could be retrieved only for rice (LOC_Os6g50310) and maize (GRZM2g000171) (Figure 4), while the probe id for other orthologous genes was unavailable. The differential expression of the soybean Glyma 12g03690 gene was also included in the analysis, since Genevestigator indicates the same as possible orthologous of AtbZIP19.

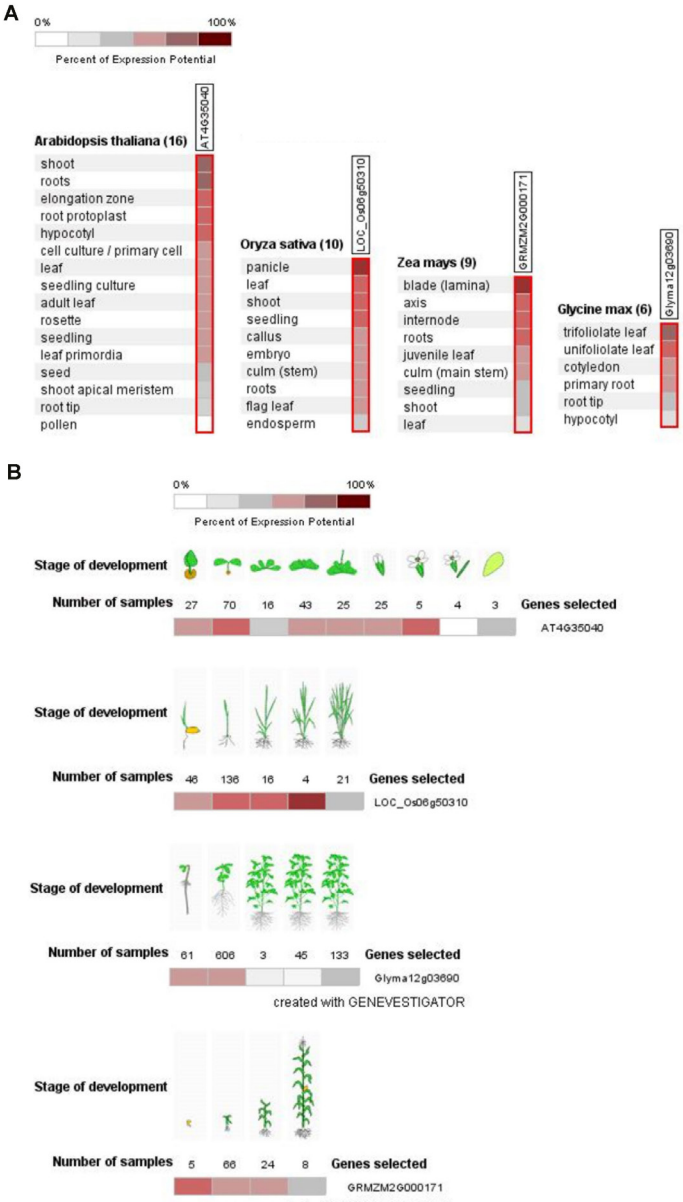

Figure 4. Differential expression patterns of the bZIP19 *Arabidopsis* gene and their orthologous in rice, maize, and soybean. Heat map showing percent of expression of the *bZIP19* gene in different anatomical features (**a**) and developmental stages (**b**) under stress condition.

It is expected that highly conserved genes during evolution do perform similar functions (Corrêa et al., 2008). The analysis showed that predicted proteins have a similar expression profile along the tissues and developmental stages under abiotic stress condition. The selected orthologous had high expression in leaf tissues and in both initial development and vegetative stages. Although these findings include the expression profile of genes undergoing a wide range of abiotic stresses, without depicting a specific response to zinc deficiency, the results of the analysis are valid since the AtbZIP19 transcription factor mediates many stress responses besides those involved in regulating primary gene responses to zinc deficiency, and these responses to stress are overlaid on various aspects.

The similar expression profiles, similar structures, and location of domains in the protein identical reinforce the idea that the selected proteins have been highly conserved and can operate the same activities. The conservation of putative orthologous when compared with the *Arabidopsis* bZIP19 gene may be useful in genetic engineering applied in the development of zinc deficiency tolerant cultivars.

CONCLUSION

Bioinformatics can provide valuable information about a gene of interest. In this study, the bZIP19 gene was investigated with the objective of finding the relationship pattern among the *Arabidopsis* bZIP19 and the bZIP proteins of cereals. Generally, we assumed that as more similar are the features, more closely related are the species, although this assertion needs to be performed cautiously. In this study, through multiple strategies of bioinformatics, it was possible to predict eight orthologous bZIP19 genes in field crops. The *Arabidopsis* transcription factor bZIP19, together with bZIP23, were recently identified as the first regulators of zinc homeostasis in plants found. Thus, with prediction of orthologous species of economic interest, it is expected to deliver results that can be used for functional genomics and biotechnological applications in cereals, especially assisting the development of cultivars showing tolerance to zinc-deficient environments.

ACKNOWLEDGMENTS

We thank financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), oordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS).

REFERENCES

Assunção AGL, Herrero E, Lin YF, Huettel B, et al. (2010). *Arabidopsis thaliana* transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc. Natl. Acad. Sci. USA* 107: 10296-10301. https://doi.org/10.1073/pnas.1004788107

Assunção AGL, Persson DP, Husted S, Schjørring JK, et al. (2013). Model of how plants sense zinc deficiency. *Metallomics* 5: 1110-1116. https://doi.org/10.1039/c3mt00070b

Bailey TL and Elkan C (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2: 28-36.

Cakmak I (2008). Enrichment of cereal grains with zinc: agronomic or genetic biofortification? *Plant Soil* 302: 1-17. https://doi.org/10.1007/s11104-007-9466-3

Cakmak I, Pfeiffer W and Mcclafferty B (2010). Biofortification of durum wheat with zinc and iron. *Cereal Chem.* 87: 10-20. https://doi.org/10.1094/CCHEM-87-1-0010

Corrêa LGG, Riaño-Pachón DM, Schrago CG, dos Santos RV, et al. (2008). The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS One* 3: e2944. https://doi.org/10.1371/journal.pone.0002944

Guo AY, Zhu QH, Chen X and Luo JC (2007). GSDS: a gene structure display server. *Yi Chuan* 29: 1023-1026. https://doi.org/10.1360/yc-007-1023

Henriques AR, Chalfun-Junior A and Aarts M (2012). Strategies to increase zinc deficiency tolerance and homeostasis in plants. *Braz. J. Plant Physiol.* 24: 3-8. https://doi.org/10.1590/S1677-04202012000100002

Hruz T, Laule O, Szabo G, Wessendorp F, et al. (2008). Genevestigator v3: a reference expression database for meta-analysis of transcriptomes. *Adv. Bioform.* 420747.

Kasuga M, Miura S, Shinozaki K and Yamaguchi-Shinozaki K (2004). A combination of the *Arabidopsis DREB1A* gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. *Plant Cell Physiol.* 45: 346-350. https://doi.org/10.1093/pcp/pch037

Lee SB, Kim H, Kim RJ and Suh MC (2014). Overexpression of *Arabidopsis* MYB96 confers drought resistance in *Camelina sativa* via cuticular wax accumulation. *Plant Cell Rep.* 33: 1535-1546. https://doi.org/10.1007/s00299-014-1636-1

Lesk A (2002). Introduction to bioinformatics. 1st edn. Oxford University Press, New York.

Nijhawan A, Jain M, Tyagi AK and Khurana JP (2008). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.* 146: 333-350. https://doi.org/10.1104/pp.107.112821

Ogo Y, Itai RN, Kobayashi T, Aung MS, et al. (2011). OsIRO2 is responsible for iron utilization in rice and improves growth and yield in calcareous soil. *Plant Mol. Biol.* 75: 593-605. https://doi.org/10.1007/s11103-011-9752-6

Osorio MB, Bücker-Neto L, Castilhos G, Turchetto-Zolet AC, et al. (2012). Identification and in silico characterization of soybean trihelix-GT and bHLH transcription factors involved in stress responses. *Genet. Mol. Biol.* 35 (1 (suppl)): 233-246. https://doi.org/10.1590/S1415-47572012000200005

Pfeiffer WH and McClafferty B (2007). In "breeding major food staples for the 21st century" Eds MS Kang, PM Priyadarshan) pp. 61-91. Blackwell Scientific: Oxford.

Song WY, Choi KS, Kim DY, Geisler M, et al. (2010). *Arabidopsis* PCR2 is a zinc exporter involved in both zinc extrusion and long-distance zinc transport. *Plant Cell* 22: 2237-2252. https://doi.org/10.1105/tpc.109.070185

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739. https://doi.org/10.1093/molbev/msr121

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882. https://doi.org/10.1093/nar/25.24.4876

Yoshioka K and Shinozaki K (2009). 1 st edn. Signal Crosstalk in Plant Stress Responses. Wiley-Blackwell, Ames.