ISOLATION AND CHARACTERIZATION OF A C-REPEAT BINDING FACTOR GENE FROM *TEVANG-1* MAIZE CULTIVAR

Nguyen Thuy Linh[1], Luu Han Ly[1], Nguyen Thuy Duong[1,2], Huynh Thi Thu Hue[1,2,*]

[1]Institute of Genome Research, VAST, Vietnam
[2]Graduate University of Science and Technology, VAST, Vietnam

ABSTRACT

C-repeat binding factor (CBF) proteins are transcription factors involved in plant response to abiotic stresses, especially low-temperature condition. In this research, a CBF3-coding gene was isolated from a cold-acclimation maize variety, *Zea mays* var. *Tevang-1* and denoted as *ZmCBF3tv*. The isolated gene shared 96.49% homology with the B73-reference gene and had no intron in the coding sequence. By using bioinformatic tools, a number of variations in the nucleotide and amino acid sequences were identified. An alignment between ZmCBF3tv and other CBF/DREB1 proteins from various species revealed functional regions and typical features, such as nuclear localization signal (NLS), the AP2 DNA-binding domain, and acidic-amino-acid-rich segments. Additionally, a phylogenetic analysis based on the AP2 domain showed that the maize CBF3 transcription factor had the highest similarity with that from rice and closely related to other CBF/DREB1 protein of monocots. The function of the *ZmCBF3tv* product is suggested to be a CBF/DREB1 transcription factor.

Keywords: CBF/DREB1, transcription factor, *Tevang-1*, maize, cold tolerance.

INTRODUCTION

Plant responses to abiotic stresses encompass the harmonization of different genes which encode regulatory and functional proteins. Among them, transcription factors (TF) have a central role in plant defense mechanisms that being activated upon stress then subsequently induce the expression of other stress-responsive genes. Therefore, analyzing and clarifying these TF genes' function are critical for understanding the plant response at the molecular level. In turn, it may contribute to creating enhanced stress-tolerant crops by genetic engineering (Century et al., 2008). Hundreds of TF genes related to plant stress resistance have been isolated and characterized in many species. Most of TF genes are classified into several large families, such as AP2/ERF, NAC, MYC, bZIP, etc (Umezawa et al., 2006).

C-repeat binding factors (CBF), also referred to as dehydration responsive-element binding factor 1 (DREB1) proteins are members of ERF subfamily of AP2/ERF superfamily. The CBF/DREB1 belonging to the A-1 subgroup are mainly participated in plant response to low temperature (Akhtar et al., 2012). *CBF* genes have been found in a wide range of plants, such as *AtCBFs* from *Arabidopsis thaliana* (Gilmour et al., 2004), *BnCBFs* from *Brassica napus* (Jaglo et al., 2001), *CbCBF* from *Capsella bursa-pastoris* (Wang et al., 2004), *CfCBF3* from sweet pepper (Yang et al., 2011), *EglCBFs* from *Eucalyptus* (Nguyen et al., 2017; Navarrete-Campos et al., 2017), *HvCBF3* from barley (Choi et al., 2002), *LpCBF3* from perennial ryegrass (Xiong & Frei, 2006), *OsCBF1* from rice (Choi et al., 2002), *PpCBF3* from bluegrass (Zhuang et al., 2015), *PaCBF* from sweet cherry (Kitashiba et al., 2004), *ScCBF* from rye (Jaglo et al., 2001), *LeCBFs* from tomato (Zhang et al., 2004), *TaCBFs* from wheat (Babben et al., 2018), *TmCBFs* from *Triticum monococcum* (Miller et al., 2006), *VvCBFs* from grapevine (Londo & Garris, 2015), and *ZmCBF3* from maize (Wang et al., 2008).

CBF proteins contain several conserved domains including the DNA-binding domain - AP2, the N-terminus nuclear localization signal (NLS), and the acidic domains present in C-terminus. CBF/DREB1 proteins are capable to bind to DNA regulatory elements called CRT (C-repeat)/DREB (dehydration responsive element) located at the promoter region of target genes. It has been reported that CBF/DREB1 transcription factors enhance the cold, drought, and salinity tolerance of the transgenic plants. For instance, the transgenic *Arabidopsis* plants harboring *CBF1* gene showed a remarkable increase in abiotic-stress tolerance due to the high-level expression of *cor 6.6, cor15a, cor47*, and *cor78* (Kasuga et al., 1999). Meanwhile, CBF3 induced the transcription of *rd29A, rd17, cor6.6, cor15a, erd10*, and *kin1* and biochemical changes related to cold acclimation including total soluble sugars and proline accumulation (Gilmour et al., 2000). The same result was observed when introduce *CaCBF3* into tobacco plant (Yang et al., 2011). Recently, seven *CBFs* that significantly related to cold-tolerance were identified in wheat (Babben et al., 2018). The association analysis clarified the relationship between polymorphic sites located on the promoter and the coding sequence of *TaCBFs* and cold-acclimation phenotype (Babben et al., 2018).

In this study, we isolated and characterized a *CBF* homolog gene from *Zea mays* var. *Tevang-1*. The *Tevang-1* is local maize of Vietnam Northern mountainous area. This variety was reported to be able to tolerate drought and low-temperature condition, in other words, cold acclimation. Then, the molecular characteristics of *ZmCBF3tv* gene and its putative protein were analyzed using bioinformatic tools.

MATERIALS AND METHODS

Plant materials and stress treatment

Seeds of *Z. mays* var. *Tevang-1* were sown in soil in the greenhouse at $28^{o}C$ under a photoperiod of 14-hours light and 10-hours dark. Plants at the 4-leaf stage were subjected

to cold stress by placing in a 4°C refrigerator for 24, 48, and 72 hours. Leaf samples were collected and stored in the RNAlater™ Stabilization Solution (Invitrogen™, USA) for RNA extraction.

RNA extraction and isolation of the *ZmCBFtv* gene

Total RNA was extracted using GeneJET RNA Purification Kit (Thermo Scientific™, USA). Then, the cDNA was synthesized by USB® First-Strand cDNA Synthesis Kit for Real-Time PCR (Affymetrix Inc., USA). Based on the *ZmCBF3* sequence reported previously (Wang et al., 2008), specific primers for this gene were designed. The sequence containing CBF-coding region along with 5'UTR and 3'UTR segments were amplified from cDNA of *Tevang-1* cultivar using the Forward primer (5'- AACAGCAAGCTCAAGCAGTCAG -3') and the Reversed primer (5'- TTGGGGTTGGCTGTCAAGCT -3'). Each 25 µL RT-PCR reaction contained 12.5 µL 2X Thermo Scientific Dream*Taq* PCR Master Mix, 1 µL of 10 µM each primer, 1 µL of c DNA, and 9.5 µL ddH$_2$O. The amplification program was 4 minutes at 95$^{\text{o}}$C followed by 30 cycles of 30 seconds at 95$^{\text{o}}$C, 60 seconds at 60$^{\text{o}}$C, and 60 seconds at 72$^{\text{o}}$C; then 5 minutes of 72°C. The PCR product was purified by GeneJET PCR Purification Kit (Thermo Scientific™, USA), then cloned into a pJET1.2 cloning vector following the manufacturer's instruction of CloneJET PCR Cloning Kit (Thermo Scientific™, USA).

Molecular and bioinformatic analysis of *ZmCBF3tv* gene and putative protein

The nucleotide composition of the isolated gene was obtained by sequencing the recombinant vector with pJET1.2 forward and reverse primers (5'- CGACTCACTATAGG GAGAGCGGC -3' and 5'- AAGAACATCG ATTTTCCATGGCAG -3', respectively) using the Big Dye Terminator kit (ABI Foster City, USA) on an Applied Biosystems™ 3500 system. The sequencing results were edited and assembled by BioEdit program (http://www.mbio.ncsu.edu/BioEdit/-page2.ht ml). The obtained sequence was annotated by BLASTN and BLASTX tools against appropriate collection database on https://www.ncbi.nlm.nih.gov/. Then, the isolated gene was compared to the reference gene from B73 RefGen_v4 on http://maizegdb.org. The deduced amino acid sequence of *ZmCBF3tv* was translated using the ExPASy tool (https://web.expasy.org/translate/). Physical and chemical characteristics of the putative *ZmCBF3tv* protein were estimated by the ProtParam web tool (https://web.expasy.org/protparam/). The relationship of ZmCBF3tv and other transcription factors of the ERF/AP2 family was carried out by comparing the AP2 domain sequences. The protein alignment was proceeded by BioEdit and then manually edited to final consistency. The phylogenetic tree was constructed using the neighbor-joining method by MEGAX software (https://megasoftware.net/). All ERF/AP2 protein sequences were retrieved from GenBank and listed in table 1.

Table 1: ERF/AP2 transcription factors and their accession numbers

Species	ID	Abbreviation
CBF/DREB1 transcription factors		
Dicots		
Brassica napus	AAL38243.1	Bn
Brassica rapa	AHL44965.1	Br
Brassica oleracea	AAQ02702.1	Bo
Ageratina adenophora	AIT39763.1	Aa
Nicotiana tabacum	ACE73693.1	Nt
Arabidopsis thaliana	ABV27151.1	Ath

Malcomia scorpioides	AFZ94953.1	Msc
Arabis pumila	ABA42927.1	Ap
Iris lactea	AAZ57434.1	Il
Capsella bursa-pastoris	AAR26658.1	Cb
Capsicum annuum	AAQ88400.1	Ca
Capsicum frutescens	ADM73296.1	Cf
Solanum melongena	AWV55520.1	Sm
Solanum tuberosum	NP 001274894.1	St
Solanum lycopersicum	AAS77819.1	Sl
Solanum pimpinellifolium	ASK51844.1	Sp
Hevea brasiliensis	AAY43213.1	Hb
Eucalyptus globulus	AQX36212.1	Egl
Santalum album	ARM20323.1	Sa
Morus alba	AFQ59977.1	Ma
Medicago truncatula	AGR40678.1	Mt
Glycine max	NP 001235507.1	Gm
Glycine soja	AAV66464.1	Gs
Eucalyptus gunnii	APG53558.1	Egu
Populus trichocarpa	ABO48365.1	Pt
Malus domestica	ART85560.1	Md
Malus sieversii	ARO50175.1	Msi
Prunus avium	BAD27123.1	Pa
Vitis labrusca	AIL00737.1	Vl
Vitis vinifera	AIL00738.1	Vv
Monocots		
Bambusa emeiensis	AFH68054.1	Be
Aegilops tauschii	AVO65154.1	Ata
Secale cereale	AAL35760.1	Sc
Festuca arundinacea	AAQ98965.1	Fa
Aegilops biuncialis	CBX87017.1	Ab
Triticum monococcum	AAY32553.1	Tm
Triticum aestivum	AWT24536.1	Ta
Hordeum vulgare	AEP32120.1	Hv
Avena sativa	CAJ21277.1	As
Brachypodium distachyon	AFD96409.1	Bd
Lolium perenne	AAX57275.1	Lp
Poa pratensis	ALG05365.1	Pp
Sorghum bicolor	AAX28960.1	Sb
Oryza sativa	AAG59619.1	Os
Zea mays	XP 008670333.1	Zm
DREB2 transcription factors		
Glycine max	AAQ57226.1	Out.Gm
Sorghum bicolor	ABD66654.1	Out.Sb
Zea mays	NP 001292873.1	Out.Zm

RESULTS AND DISCUSSION

Isolation of *ZmCBF3tv* gene from *Tevang-1* maize

The total RNA samples extracted from leaves of one, two, and three-day cold treatment plant were checked by electrophoresis on 0.8% agarose gel and then used as templates for cDNA synthesis experiment (Fig. 1a). The RT-PCR analysis was carried out using the specific primer pair and condition as described above. As the result, DNA bands which were nearly 0.9 kb in length were observed in the 24 and 48 hours cold-treated samples (Fig. 1b). The PCR products were purified and then inserted into the cloning vector - pJET1.2.

Figure 1. Isolation of *ZmCBF3tv* gene from *Tevang-1* maize: a. Result of total RNA extraction. 1–3: Total RNA from maize plants after 24, 48, and 72 hours under low-temperature condition, respectively; b. Result of RT-PCR reaction. M: GeneRuler 1 kb DNA Ladder (Thermo Scientific, USA). 1–3: PCR products using cDNA templates of 24, 48, and 72-hours cold-treated samples, respectively

Analysis of nucleotide sequence of *ZmCBF3t*v gene from *Tevang-1* maize

Sequencing of the inserted fragment in the recombinant vector pJET1.2 revealed that the isolated cDNA had 908 nucleotides (Fig. 2). A BLASTN analysis on GenBank showed that this sequence shared similarities to many genes encoding CRT/DRE binding factor protein from maize, sorghum, barley, and panicgrass. The isolated sequence had the highest homology (96.49%) to PREDICTED: *Zea mays* dehydration-responsive element-binding protein 1H (LOC103647602), mRNA (ID: XM_008672111.3) from *Z. mays* var. *B73* followed by *ZmCBF3* mRNA (95.86%) (ID: AY964718.1) of *Z. mays* var. *Qi319* (Wang et al., 2008). Besides, by using BLASTX tool against the UniProtKB-TrEMBL database and non-redundant proteins, it is confirmed that the query carried the genetic information of a TF belonging to the ERF/AP2 family. Therefore, the isolated cDNA from *Tevang-1* maize cultivar was denoted as *ZmCBF3tv*. Additionally, the genomic DNA sequence that corresponded to the *ZmCBF3tv* cDNA was also identified through a BLAST search on http://www.maizegdb.org/. According to the result, *CBF3* gene was located on chromosome 2 and contained no intron inside the coding sequence as other *CBF/DREB1* genes (Akhtar et al., 2012).

```
                              □  ■                                                           ●
ZmCBF3 reference  CTCACTCCTCAC··TCCACTTCCACTCCCAACAGCAAGCTCAAGCAGTCAGTCACCGGCAGGGGTCAGGGT  70
ZmCBF3tv          ................T...................................................T... 
                                                                                            M  V   2

                                                        ●
ZmCBF3 reference  CACAGTCACAGCAGCAGCCATGGACACGGCCGGCCTCGTCCAGCACGCGACCTCCTCGTCTTCCACCTCC 140
                                     M  D  T  A  G  L  V  Q  H  A  T  S  S  S  S  T  S  25
ZmCBF3tv          ...............................C..................................... 
                             T  V  T  A  A  A  .  .  .  .  R  .  .  .  .  .  .  .

                                       ▲□                                  ●        ●           ●
ZmCBF3 reference  ACCTCGGCGTCGTCCTCCTCGTCCGAGCAGCAGAGCAAGGCGGCGTGGCCGCCGTCGCCCGCTT 210
                   T  S  A  S  S  S  S  S  S  E  Q  Q  S  K  A  A  W  P  P  S  P  A  48
ZmCBF3tv          ...................G................T...C.........T...... 
                   .  .  .  .  .  .  .  .  .  .  .  .  V  C  .  .  .  S  .

ZmCBF3 reference  CCTCCCCGCAGCAGCCGCCCAAGAAGCGCCCCGCGGGGCGCACGAAGTTCCGGGAGACGCGGCACCCGGT 280
                   S  S  P  Q  Q  P  P  K  K  R  P  A  G  R  T  K  F  R  E  T  R  H  P  V  72
ZmCBF3tv          ..................................................................... 
                   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

ZmCBF3 reference  GTTCCGCGGCGTGCGGCGGCGGGGCGCCGCGGGCCGGTGGGTGTGCGAGGTGCGCGTCCCGGGGGAGGCGC 350
                   F  R  G  V  R  R  R  G  A  A  G  R  W  V  C  E  V  R  V  P  G  R  R  95
ZmCBF3tv          ..................................................................... 
                   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

ZmCBF3 reference  GGCGCGCGGCTGTGGCTCGGCACCTACCTCGCCGCCGAGGCGGCGGCGCGCGCACGACGCCGCGATGC 420
                   G  A  R  L  W  L  G  T  Y  L  A  A  E  A  A  A  R  A  H  D  A  A  M  118
ZmCBF3tv          ..................................................................... 
                   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

ZmCBF3 reference  TCGCCCTGCAGGGCCGCGGCGCGGGGGCGCCTCAACTTCCCGGACTCCGCGCGGCTGCTCGCCGTGCCGCC 490
                   L  A  L  Q  G  R  G  A  G  R  L  N  F  P  D  S  A  R  L  L  A  V  P  P  142
ZmCBF3tv          ..................................................................... 
                   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

ZmCBF3 reference  CCCGTCCGCGCTCCCGGGCCTGGACGACGCCCGCCGGGCGGCGCTCGAGGCCGTCGCGGAGTTCCAGCGC 560
                   P  S  A  L  P  G  L  D  D  A  R  R  A  A  L  E  A  V  A  E  F  Q  R  165
ZmCBF3tv          ..T.................................................................. 
                   L  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

                              ●    ▬▬▬▬                 ●                            ▬▬▬
ZmCBF3 reference  CGCTCCGGGGCC··GCCGACGAAGCGACCTCGGGCGCGTCTCCTCCCTCCTCGTCGCCGTCG··C 630
                   R  S  G  A        A  D  E  A  T  S  G  A  S  P  P  S  S  S  P  S  188
ZmCBF3tv          .........T..GGGCCC.................T...........................TCG. 
                   .  .  S  G  P  .  .  .  .  .  V  .  .  .  .  .  .  .  .  S

                              ●                  ●     ●                 ●        ▲
ZmCBF3 reference  TGCCGGACGTTTCTGCTGCTGGCTCGCCGGCGGCGGCGCTTGAGCACGTGCCTGTGAAGGCCGACGAAGC 700
                   L  P  D  V  S  A  A  G  S  P  A  A  A  L  E  H  V  P  V  K  A  D  E  A  212
ZmCBF3tv          .........A...................C....A..............A....T....... 
                   .  .  .  I  .  .  .  .  .  .  P  .  I  .  .  .  E  .  .  .

                                                  ▲   ▲●
ZmCBF3 reference  AGTGGCGTTGGACTTGGACGGCGATGTGTTTGAGCCCGACTGGTTCGGGGACATGGACCTGGAGTTGGAT 770
                   V  A  L  D  L  D  G  D  V  F  E  P  D  W  F  G  D  M  D  L  E  L  D  235
ZmCBF3tv          ...........................C.....C.G................................ 
                   .  .  .  .  .  .  .  .  G  .  .  .  .  .  .  .  .  .  .  .  .  .

                                                                       ▲●
ZmCBF3 reference  GCGTACTACGCCAGCCTCGCGGAAGGGTTGCTCGTGGAGCCGCCGCCGCCAGCGGCCGCCTGGGATCATG 840
                   A  Y  Y  A  S  L  A  E  G  L  L  V  E  P  P  P  P  A  A  A  W  D  H  258
ZmCBF3tv          .............................................GC.................... 
                   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  P  .  .  .  .  .  .

                   ●                       ▲
ZmCBF3 reference  AAGACTGCTGTGACTCCGGAGCCGCCGCTGACGTCGCGCTCTGGAGCTACTACTAGCAAAGTTAACAATAATA 910
                   E  D  C  C  D  S  G  A  A  D  V  A  L  W  S  Y  Y  *  275
ZmCBF3tv          G..................................T................................ 
                   G  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  *

                   ...
ZmCBF3 reference  AGC
ZmCBF3tv          ...
```
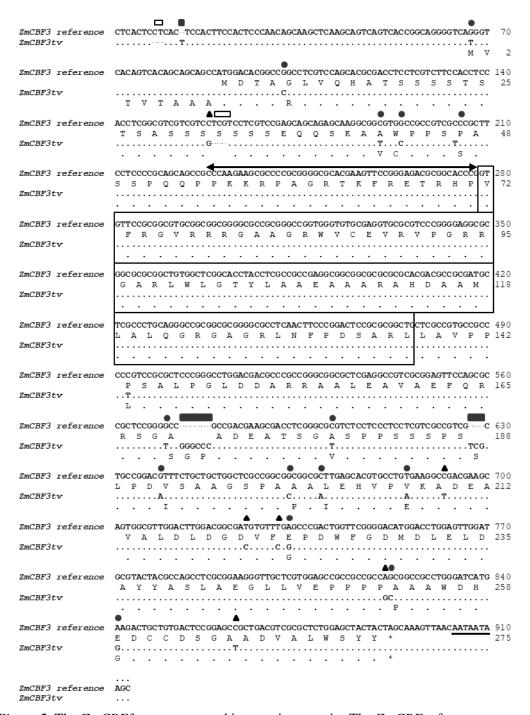
Figure 2. The *ZmCBF3tv* sequence and its putative protein: The *ZmCBF* reference sequence originated from B73 maize that downloaded from http://www.maizegdb.org/. The putative protein sequences were translated by ExPASy tool. The AP2 DNA-binding domain was enclosed inside boxes. A double-headed arrow demonstrated the nuclear localization signal (NLS) and the polyadenylated signal was underlined. Unfilled and filled rectangles represented deletion and insertion mutation, respectively. Triangles and dots showed nucleotide substitutions including synonymous mutation and missense mutation, respectively

The 908 bp *ZmCBF3tv* isolated from *Z. mays* var. *Tevang-1* contained an entire coding region and part of 5' and 3'UTR in which polyadenylated signal was underlined (Fig. 2). A nucleotide alignment confirmed that the *ZmCBF3tv* shared 96.49% homology with the *ZmCBF3* sequence from B73 reference maize (2:200398600-200397688). In the meantime, a number of variations in the coding region and 5'UTR region were identified. As shown in Fig. 2, variations dispersed at 5' end and occurred more frequently at 3' end of the coding region. However, the segment defining the DNA-binding domain in the TF was conserved. Among such variations, one deletion of TCG codon at position 159 and two insertion mutations including GGCCCG at position 574 and TCG at position 627 increased the nucleotide number of *ZmCBF3tv* gene by six. Noticeably, one substitution mutation happened at position 67 which formally was the 5'UTR of *ZmCBF3* from B73 maize created a new ATG codon. In fact, the B73-origin *ZmCBF3* consists of an open reading frame sequence expanding 804 bp segment. The appearance of an upstream ATG codon probably led to the addition of 24 nucleotides to the beginning of *ZmCBF3tv* coding region. Nevertheless, these variations did not affect the reading frame of ZmCBF3tv protein.

Sequence analysis of *ZmCBFtv* protein in silico

The deduced ZmCBF3tv protein contained 274 amino acids and was rich in Ala (17.2%), Ser (12.0%), and Pro (9.5%). Besides, the protein contained a low number of Asn and Ile, with the figure of one and two residues, respectively. The molecular weight of ZmCBF3tv was predicted at 28.6 kDa with the theoretical isoelectric point of 5.58. The instability index and GRAVY were calculated of 68.20 and -0.314, respectively, suggesting that the ZmCBF3tv protein had an unstable and non-polar nature.

By comparing the ZmCBF3tv protein to the homolog one originated from B73 maize, we found many changes in the composition of the polypeptide (Fig. 2). In term of protein

length, due to the appearance of an ATG codon at position 67 and three InDel mutations, the ZmCBF3tv protein was 10 amino acids longer than the ZmCBF3 TF of B73 maize. By ways of illustration, an additional MVTVTAAA segment was detected at the most N-terminus while additional Gly-Pro and Ser residues were observed at position 170 and 188, respectively. On the other hand, the 7xSer cluster was replaced by a 6xSer sequence as a TCG codon located at position 159 was removed from ZmCBF3tv coding region. Furthermore, 14 nonsynonymous substitutions including eight transversion mutations and six transition mutations also impact the amino acid content of ZmCBF3tv. For instances, 102G>C, 189C>T, and 733A>G caused the transition of G13R, A42V, and E223G, respectively (Fig. 2).

The alignment of ZmCBF3tv with other CBF/DREB1 proteins from different species revealed functional regions and "signature" motifs of CBF3 proteins (Fig. 3). Because of being a TF, the putative ZmCBF3tv protein had a basic-amino-acid-rich sequence of PKKRPAGRTKFRETRHP as a nuclear localization signal (NLS) that help the protein entered the nucleus. After the NLS segment, the AP2 DNA-binding domains of AP2/ERF superfamily expand to a sequence of about 60 amino acids (Fig. 3). Additionally, several conserved regions rich in D and E and LWSY motif were identified near the C-terminus of ZmCBF3tv protein. Nevertheless, besides the high similarity between functional regions, the composition and structure of CBF/DREB1 proteins are quite diverse among species (Fig. 3).

All TFs belonging to AP2/ERF superfamily contain the AP2 motif that highly conserved within the family. The AP2 domains were analyzed further and used for phylogenetic analysis (Fig. 4, 5). The AP2 sequence consisted of two local regions: YRG and RAYD. While the 20-amino-acid YRG region that was predicted to be important to the DNA-binding activity rich in basic and hydrophobic residues, the RAYD region

contained about 40 amino acids contributing to amphipathic property was proposed to mediate protein-protein interaction (Okamuro et al., 1997). Therefore, the AP2 DNA-binding domain was composed of three anti-parallel β-sheets and one α-helix. The alignment of AP2 domain in CBF/DREB1 TFs from *Z. mays* var. *Tevang-1* and other species led to a consistent result with previous studies. For instance, a number of highly conserved amino acids, in which, some are critical to DNA specificity-binding and activation activity, such as Arg (6,8,27), Trp (14,29), Val (15), and Glu (17) (Allen et al., 1998) were noticed. These residues were remain unchanged in ZmCBF3tv, suggesting that the C-repeat binding factor from *Tevang-1* might have the same target gene and bind to the *cis*-element on such gene promoter with the same affinity in comparison to other CBF/DREB1 proteins.
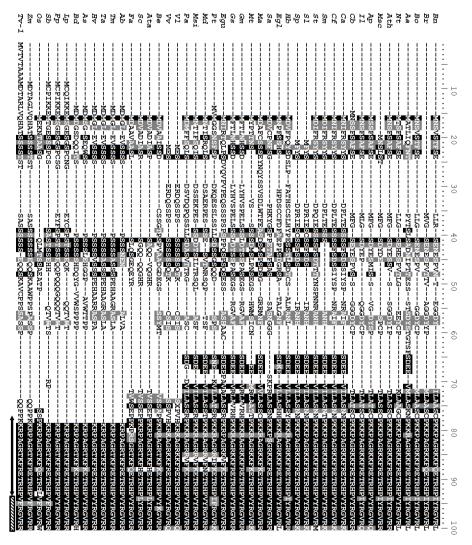

Figure 3. Amino acid sequence alignment of the putative CBF/DREB1 TFs of *Tevang-1* maize (Tv-1) and those from various species. The nuclear localization signal (NLS) was demonstrated by a double-headed arrow. The DNA-binding domain was presented by rectangle box filled with upward diagonal pattern. Highly and moderately similar residues across species were shade in black and gray, respectively

Figure 3. Amino acid sequence alignment of the putative CBF/DREB1 TFs of *Tevang-1* maize (Tv-1) and those from various species. The nuclear localization signal (NLS) was demonstrated by a double-headed arrow. The DNA-binding domain was presented by rectangle box filled with upward diagonal pattern. Highly and moderately similar residues across species were shade in black and gray, respectively (next)
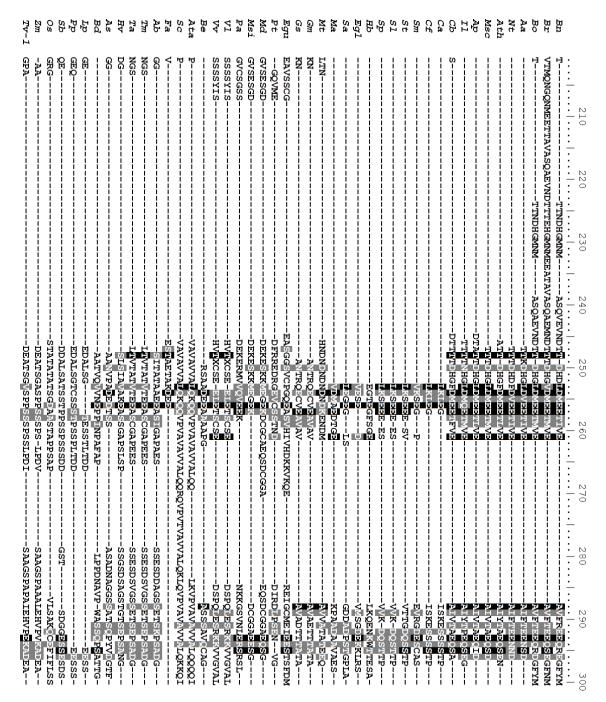
Figure 3. Amino acid sequence alignment of the putative CBF/DREB1 TFs of *Tevang-1* maize (Tv-1) and those from various species. The nuclear localization signal (NLS) was demonstrated by a double-headed arrow. The DNA-binding domain was presented by rectangle box filled with upward diagonal pattern. Highly and moderately similar residues across species were shade in black and gray, respectively (next)

Figure 3. Amino acid sequence alignment of the putative CBF/DREB1 TFs of *Tevang-1* maize (Tv-1) and those from various species. The nuclear localization signal (NLS) was demonstrated by a double-headed arrow. The DNA-binding domain was presented by rectangle box filled with upward diagonal pattern. Highly and moderately similar residues across species were shade in black and gray, respectively (next)
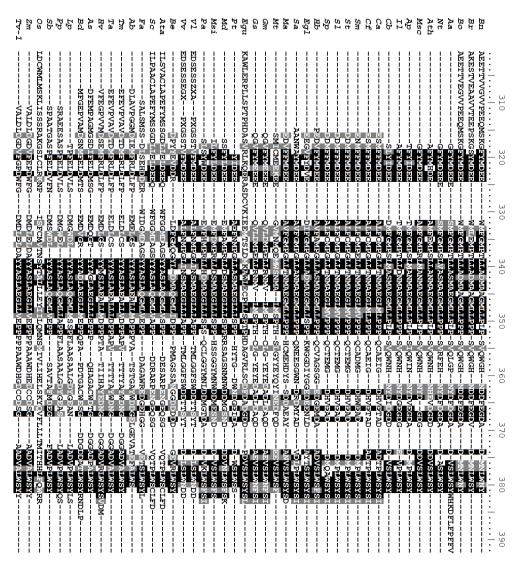
Whereas, there were several differences in amino acid content of the DNA-binding domain between the monocots and dicots CBF/ DREB1 proteins. For instance, Glu (20) was observed in almost dicot plants except for *Vitis spp.* contain Asp (20) while in monocots, the position was replaced by Val (20) in spite of *B. emeiensis* had Glu (20) (Fig. 4). Other moderately conserved positions in dicots were Lys (13), Asn (22), and Ile (28), but those in monocot were Arg (13), Gly (22), and Leu (28), respectively. Because of originating from maize, the ZmCBF3tv contained such monocots-typical residues. Noticeably, in the position 35, the CBF/DREB1 transcription factor from *Z. mays* including B73 and *Tevang-1* maize had Ala residue instead of Thr as those from other plants (Fig. 4). It was also identified that the DVAA (44) and DSAWR (66) motif were highly conserved

among dicots. Meanwhile, most of the monocots including *Tevang-1* contained DAA(M/V) motif and DSAWL, respectively (Fig. 4).

```
                10        20        30        40        50        60        70
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Bn    IYRGVRLR-KSGKWVCEVREPNKK-SRIWLGTFKTAEIAARAHDVAALALRGRG-----ACLNFADSAWR
Br    .........-................-............M.........------....Y......
Bo    .........-................-............-..........------...........
Aa    .........-N...............-......L...........I...KS----............
Nt    V........-N...............-......L...........I...KS----............
Ath   ......R.-N..............-T......Q...M.........S----...........
Msc   ......R.-N..............-T......P...M.........S----...........
Ap    V.....Q.-N.....S.L......-T......Q..M.......I...S-----.......V..
Il    ......Q.-N.....S.........-T......Q..M.......I...S-----............
Cb    V.....R.-N.............-......P..DM.......I...S-----............
Ca    V.....K.-N.............-......P..M........I...S-----............
Cf    V.....K.-N.............-......P..M........I...S-----............
Sm    V.....K.-N.............-......P..M........S-----.........G.
St    V.....M.-N..............-T......P..M........I...S-----............
Sl    V.....K.-N..............-T......P..M........I...S-----............
Sp    V.....K.-N..............-T......P..M........I...S-----............
Hb    ......R.-N.......I.......-......P.E.M.........S-----.........S.
Egl   V........-D..........R..-......P.V.M.........S-----............
Sa    V.....R.-D.........S.Q-A....P..DM....Y.....S...S-----..H..E.....
Ma    V.....R.-N..............-T......P..M..........I...S-----T.........
Mt    V.....K.-NLD....M......-TK....P...M.......M...Y-----............
Gm    V.....R.-N.D.............-T....P.P.M.......M...Y-----............
Gs    V.....R.-N.D.............-T....P.P.M.......M...Y-----............
Egu   V.....R.-NR......L..S..-T.V....YP...M..........HG----V..........
Pt    .F....K.-NGD.....L.......-......YP.P.M.........F..KS-----........
Md    V.....R.-NND.....M......K........YP...M.........F..L-----........
Msi   V.....R.-NNN.....M......K........YQ...M.........F..KL-----.......V..
Pa    V.....R.-NND.....M......K........YP...M.........F..KL-----..I........
Vl    V.....Q.-NGNR......D.KX.-........X.P.M.........F..DF----A.-...P..TS.
Vv    V.....Q.-NGNR......D.KX.-........X.P.M....X......F..BF----A.-..P..XS.
Be    V.K...R.-NP.R........HG.-Q......E...M...........-----A.....PR.
Ata   L.....R.GRV.Q......V.GI.G..L....N...M....A.V...S..-----A.........
Sc    L.....R.GRV.Q......V.GI.G..L....N...M....A.V...S..A-----.........
Fa    V.....R.GRA.Q...M.VHGT.G..L....D...M......S..D-----.........
Ab    V.....R.GNAER.....V.G.RGA.L....YA........N.A.M...G..S----A.R..P....L
Tm    V.....R.GNAER.....V.G.RGA.L....YA........N.A.M...G..S----A...........L
Ta    V.....R.GNAER.....V.G.RGA.L....YA........N.A.M...G..S----A...........L
Hv    V.....R.GNTER.....V.G.RGA.L....YA...V....N.A.M...G..S----A...........L
As    V.....R.GNA.R.....V.GQRGE.L....YL...S.....A.M.G.L..SA--SS...........
Bd    V.....H.GNA.R.....V.GTSGK.L....HL...S.....A.M...H.PSAAAAARL...P....L
Lp    V.....R.GNA.R.....V.GRRG..L.V..D..........A.M...A-A.----DV.......EL
Pp    V.....R.GNA.R.....V.GRRG..L.V..D..........A.M...A.A.S----V...V..EL
Sb    VF....R.GNA.R.....V.GRRGC.L....D..DA....A.M..IA.A.-----........L
Os    VF....W.GCA.R...K..V.GSRGD.F.I..SD...ET..T..A.M...C.AS-----.S...........L
Zm    VF....R.GAA.R.....V.GRRGA.L....YLA..A.....A.M...Q..----GAGR...P...RL
Tv-1  VF....R.GAA.R.....V.GRRGA.L....YLA..A.....A.M...Q..----GAGR...P...RL
Cons     RGVR R    WVCE R P    R W GT  T   AARA D A  A   G        LNF DS
```
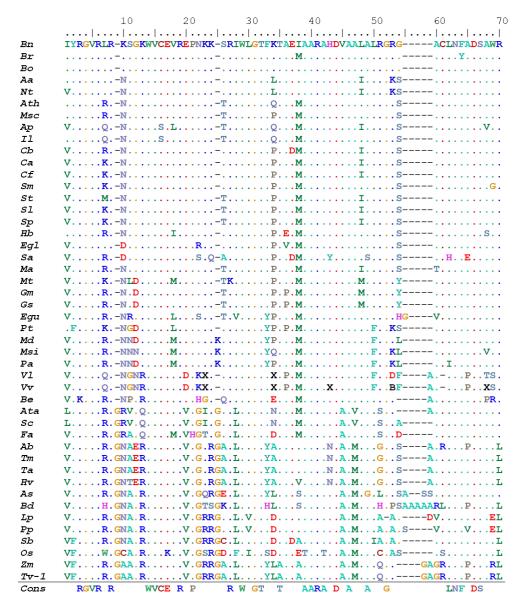
Figure 4. The comparison of the AP2 domain sequence between CBF/DREB1 transcription factors from various species. The *Cons*, or consensus sequence shown underneath was created using BioEdit software

Phylogenetic analysis of *ZmCBFtv* proteins

A phylogenetic analysis based on the AP2 domain sequence was carried out to clarify the phylogenetic relationship between ZmCBF3tv and other CBF/DREB1 TFs from diverse species. We compared the amino acid similarities in 48 proteins from 44 species, in which three DREB2 proteins served as outgroup samples. As shown in Fig. 5, the CBF/DREB1 transcription factors were well separated into two groups from monocots and dicots with the exception of *B. emeiensis*. The

ZmCBF3tv shared the highest homology with one from *O. sativa* and had close relationships with those from monocots including barley, sorghum, wheat, etc (Fig. 5). The result was in agreement with the alignment of the AP2 domain (Fig. 4).
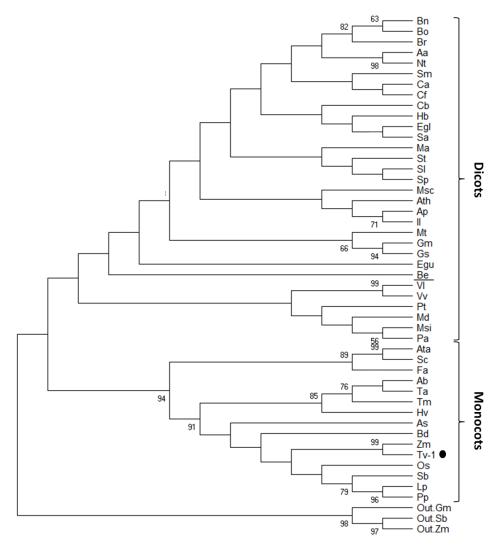

Figure 5. The phylogenetic tree base on AP2 sequence of CBF/DREB1 transcription factors from different species

CBF/DREB1 TF genes were induced by various abiotic stresses, such as low-temperature condition. However, the expression pattern of CBF/DREB1 genes to stress were varied upon species. Wang et al. (2008) reported that the *ZmCBF3* of *Z. mays* var. *Qi319* was strongly induced after 48 hours of cold stress. Nonetheless, *BjDREB1B* gene expressed immediately when *B. juncea* plants confronting to stresses - in other words after 20-minute treatment, and reached the highest transcription level between 2 and 12 hours (Cong et al., 2008). In this study, *ZmCBF3tv* gene was expressed after one and two days under 4°C, in which the expression level at time-point of 24-hours was higher. However, there were no transcripts of *ZmCBF3tv* observed when using a longer cold-exposing time (72 hours) (Fig. 1).

The alignment of *ZmCBF3* locus between *Tevang-1* and B73 maize revealed 25 genetic variations (3.51%) including 20 substitutions and five InDels upon a segment of 908 bp long (Fig. 2). The percentage of variation occurred in *CBF/DREB1* loci were analyzed in many plants and different according to genes and species. In wheat, the ratio of polymorphic sites to the coding sequence length of *TaCBFs* ranged from 0.00% to 3.54% (Babben et al., 2018). Meanwhile, a research examining 87 different sequences of *VvCBF3* showed 8.9% variation in the coding region (Londo & Garris, 2014). Although 3.51% of *ZmCBF3* sequence contained genetic differences between two varieties (B73 and *Tevang-1*), this figure was not representative for maize due to the limitation in sample collection and the data size as only two accessions were included in the analysis. Nevertheless, the changes in the 5'UTR and the coding sequence of *ZmCBF3tv* gene did alter the amino acid composition of C-repeat binding factor. The changes in ZmCBF3tv polypeptide sequence might play a role in enhancing the cold-tolerance phenotype observed in *Z. mays* var. *Tevang-1*.

The study of putative CBF/DREB1 proteins from *Tevang-1* maize and other species using bioinformatic tools suggested that ZmCBF3tv was a typical CBF/DREB1 TF that could induce the expression of various stress-responsive genes due to the presence of many functional and "signature" regions (Fig. 3, 4). Additionally, the putative ZmCBF3tv protein had a close phylogenetic relationship to other monocots and had the highest similarity to one from rice (Fig. 5). To clarify the function of the ZmCBF3tv transcription factor, it is necessary to conduct further experimental studies. Still, the results from this *in sillico* analysis of the putative protein gave clues on the ZmCBF3tv activity.

CONCLUSIONS

In this research, we isolated from *Tevang-1* maize variety a 908 bp cDNA called *ZmCBF3tv* that was 96.49% homology to the reference genome B73. The isolated fragment consisted of many variations in the 5'UTR

and coding sequences. Furthermore, the putative ZmCBF3tv protein had the conserved DNA-binding domain - AP2 and some signature regions such as NLS and C-terminus acidic domains. In addition, the ZmCBF3tv had the close phylogenetic relationship with other CBF/DREB1 transcription factors from monocots. Therefore, it was implemented that the *ZmCBF3tv* originated from *Z. mays* var. *Tevang-1* carried the genetic information coding for a cold-response C-repeat binding factor.

REFERENCES

Akhtar M., Jaiswal A., Taj G., Jaiswal J. P., Qureshi M. I., Singh N. K., 2012. DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. *J. Genet.*, 91(3): 385–395.

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

Babben S., Schliephake E., Janitza P., Berner T., Keilwagen J., Koch M., Arana-Ceballos F. A., Templer S. E., Chesnokov Y., Pshenichnikova T., Schondelmaier J., Borner A., Pillen K., Ordon F., Perovic D., 2018. Association genetics studies on frost tolerance in wheat (*Triticum aestivum* L.) reveal new highly conserved amino acid substitutions in *CBF-A3, CBF-A15, VRN3* and *PPD1* genes. *BMC Genomics*, 19(1): 409.

Century K., Reuber T. L., Ratcliffe O. J., 2008. Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. *Plant Physiol.*, 147: 20–29.

Choi D. W., Rodriguez E. M., Close T. J., 2002. Barley *Cbf3* gene identification,

expression pattern, and map location. *Plant Physiol.*, 129(4): 1781–1787.

Cong L., Chai T. Y., Zhang Y. X., 2008. Characterization of the novel gene *BjDREB1B* encoding a DRE-binding transcription factor from *Brassica juncea* L. *Biochem. Biophys. Res. Commun.*, 371(4): 702–706.

Gilmour S. J., Fowler S. G., Thomashow M. F., 2004. *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. *Plant Mol. Biol.*, 54: 767–781.

Jaglo K. R., Kleff S., Amundsen K. L., Zhang X., Haake V., Zhang J. Z., Deits T., Thomashow M. F., 2001. Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol.*, 127: 910–917.

Kasuga M., Liu Q., Miura S., Yamaguchi-Shinozaki K., Shinozaki K., 1999. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress inducible transcriptionfactor. *Nat. Biotechnol.*, 17: 287–291.

Kitashiba H., Ishizaka T., Isuzugawa K., Nishimura K., Suzuki T., 2004. Expression of a sweet cherry DREB1/CBF ortholog in *Arabidopsis* confers salt and freezing tolerance. *J. Plant Physiol.*, 161(10): 1171–1176.

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: 391–406.

Londo J. P., Garris A. J., 2015. Genetic and protein sequence variation of CBF1-4 in cold hardy wild grapevine germplasm. *VITIS,* 53(4): 201–206.

Miller A. K., Galiba G., Dubcovsky J., 2006. A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-A m 2 in *Triticum monococcum. Mol. Genet. Genomics*, 275(2): 193–203.

Navarrete-Campos D., Le-Feuvre R., Balocchi C., Valenzuela S., 2017. Overexpression of three novel CBF transcription factors from *Eucalyptus globulus* improves cold tolerance on transgenic *Arabidopsis thaliana. Trees*, 31(3): 1041–1055.

Nguyen H. C., Cao P. B., San-Clemente H., Ployet R., Mounet F., Ladouce N., Harvengt L., Marque C., Teulieres C., 2017. Special trends in CBF and DREB2 groups in *Eucalyptus gunnii* vs *Eucalyptus grandis* suggest that CBF are master players in the trade-off between growth and stress resistance. *Physiol. Plant.*, 159(4): 445–467.

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. USA, 94: 7076–7081.

Umezawa T., Fujita M., Fujita Y., Yamaguchi-Shinozaki K., Shinozaki K., 2006. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Curr. Opin. Biotechnol.*, 17: 113–122.

Wang L., Luo Y., Zhang L., Zhao J., Hu Z., Fan Y., Zhang C., 2008. Isolation and characterization of a C-repeat binding transcription factor from Maize. *J. Integr. Plant Biol.*, 50(8): 965–974.

Wang X., Liu S., Liu X., Chen Z., Liu X., Pang Y., Sun X., Tang K., 2004. Molecular cloning and characterization of a *CBF* gene from *Capsella bursa-pastoris. DNA Seq.*, 15(3): 180–187.

Xiong Y., Fei S. Z., 2006. Functional and phylogenetic analysis of a DREB/CBF-like gene in perennial ryegrass (*Lolium perenne* L.). *Planta*, 224(4): 878–888.

Yang S., Tang X. F., Ma N. N., Wang L. Y., Meng Q. W., 2011. Heterology expression of the sweet pepper *CBF3* gene confers elevated tolerance to chilling stress in transgenic tobacco. *J. Plant Physiol.*, 168(15): 1804–1812.

Zhang X., Fowler S. G., Cheng H., Lou Y., Rhee S. Y., Stockinger E. J., Thomashow M. F., 2004. Freezing-sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing-tolerant *Arabidopsis*. *Plant J.*, 39(6): 905–919.

Zhuang L., Yuan X., Chen Y., Xu B., Yang Z., Huang B., 2015. *PpCBF3* from cold-tolerant kentucky bluegrass involved in freezing tolerance associated with up-regulation of cold-related genes in transgenic *Arabidopsis thaliana*. *PloS One*, 10(7): e0132928.