Genome-Wide Identification and Classification of the AP2/EREBP Gene Family in the Cucurbitaceae Species

Sang-Choon Lee¹, Won-Kyung Lee¹, Asjad Ali², Manu Kumar³, Tae-Jin Yang¹*, Kihwan Song⁴*

¹Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea
²National Agrobiodiversity Center, National Institute of Agricultural Sciences, RDA, Jeonju 54875, Korea
³Department of Life Science, Sogang University, Seoul 04107, Korea
⁴Department of Bioresources Engineering, College of Life Sciences, Sejong University, Seoul 05006, Korea

ABSTRACT AP2/EREBP gene family consists of transcription factor genes with a conserved AP2 DNA-binding domain and is involved in various biological processes. AP2/EREBP gene families were identified through genome-wide searches in five Cucurbitaceae species including cucumber, wild cucumber, melon, watermelon, and bitter gourd, which consisted of more than 100 genes in each of the five species. The gene families were further divided into five groups including four subfamilies (ERF, DREB, AP2 and RAV) and a soloist group. Among the subfamilies, DREB subfamily which is known to be related to abiotic stress response was more analyzed and a total of 25 genes were identified as Cucurbitaceae homologues of Arabidopsis CBF/DREB1 genes which are important for abiotic stress-response and tolerance. In silico expression profiling using RNA-Seq data revealed diverse expression patterns of cucumber AP2/EREBP genes. AP2/EREBP gene families identified in this study will be valuable for understanding the stress response mechanism as well as facilitating molecular breeding in Cucurbitaceae crops.

Keywords AP2/EREBP gene family, CBF/DREB, Cucumber, Cucurbitaceae

INTRODUCTION APETALA 2/ethylene responsive element binding protein (AP2/EREBP) genes are members of a largest gene family in plants and encode transcription factors (TFs) with a conserved AP2 DNA-binding domain. These genes play important roles in various biological processes including flowering control, hormone response and environmental stress signaling. So far, AP2/EREBP gene families and their functions have been characterized in many plant species. For example, 147 AP2/EREBP genes were identified and their diverse biological roles were studied in Arabidopsis thaliana (Dietz et al. 2010). The family can be further divided into five groups including four subfamilies (ERF, DREB, AP2 and RAV) and a soloist group (Nakano et al. 2006; Dietz et al. 2010). The DREB subfamily itself is composed of six subgroups, of which the A1 subgroup includes C-repeat/dehydration-responsive element (CRT/DRE) binding factor 1 (CBF/DREB1) genes (Dietz et al. 2010; Lee et al. 2012). The CBF/DREB1 genes are known to play an important role in stress response and tolerance (Lata and Prasad 2011; Lee et al. 2012).

The Cucurbitaceae family consists of about 120 genera and more than 800 species distributed mostly in the tropical and subtropical regions. The family also includes many economically important vegetable and fruit crops such as cucumber, melon and watermelon (Lu and Jeffrey 2011). Therefore, the family is valued as the world’s second
largest vegetable producing family after Solanaceae family (Bhowmick and Jha 2015). So far, six species in the Cucurbitaceae family have been analyzed for their nuclear genome sequencing. Among those, *Cucumis sativus* var. *sativus* (cucumber) is an economically important crop and a model plant for sex determination and vascular biology study (Tanurdzic and Banks 2004; Lough and Lucas 2006; RDA 2013). The genomes of three cucumber cultivars were sequenced (Huang et al. 2009; Cavagnaro et al. 2010; Wóycicki et al. 2011). *Cucumis sativus* var. *hardwickii* is known as a wild cucumber and its genome was also sequenced (Qi et al. 2013). *Cucumis melo* (melon) and *Citrullus lanatus* (watermelon) are important fruit crops and genomes of both species were also sequenced (Garcia-Mas et al. 2012; Guo et al. 2013). *Momordica charantia* (bitter gourd) is an important vegetable and medicinal plant, whose genome was recently sequenced (Urasaki et al. 2016). Similarly, *Cucurbita maxima* (pumpkin) is an economically important crop and its genome sequence is available (Zhang et al. 2015; http://www.icugi.org/cgi-bin/ICuGI/index.cgi).

As other crops, the growth and yield of Cucurbitaceae crops are also affected by abiotic stresses, which are expected to become more frequent due to global climate changes. For example, cucumber growth is stopped at a temperature higher than 35°C and lower than 12°C and also very sensitive to low humidity in the air and high salt concentration in soil (RDA 2013; Ali et al. 2014a, 2014b). Although many studies have been performed to understand physiological and molecular mechanisms for stress response and tolerance in Cucurbitaceae crops, the progress is still in the early stage for molecular breeding in the crops.

Although more than 100 AP2/EREBP genes were identified in cucumber and watermelon genomes (Huang et al. 2009; Hu and Liu 2011; Guo et al. 2013), no comprehensive study on the AP2/EREBP gene family has been performed in Cucurbitaceae species. On this account, this study identified and compared AP2/EREBP gene families through genome-wide searches in five Cucurbitaceae species whose genomes are publically available. The gene information obtained in this study will be valuable for understanding stress response mechanism and facilitating molecular breeding in Cucurbitaceae crops.

**MATERIALS AND METHODS**

**Investigation of transcription factors and AP2/EREBP genes**

Seven gene sets of cucumber cultivars, wild cucumber, melon and watermelon were downloaded from genome database of respective plant. Gene set of bitter gourd was kindly provided by Prof. Hideo Matsumura in Shinshu University, Japan. Since gene set of pumpkin genome is not available, it was not used in this study. Deduced protein sequences of eight genomes including Arabidopsis genome (as a reference) were used to investigate TFs and AP2/EREBP genes (Table 1). iTAK standalone software (ver. 1.6b, http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi, Zheng et al. 2016) was utilized to identify and classify TFs, with default parameters. Among the identified TFs, AP2/EREBP domain-containing protein sequences were selected and analyzed further.

**Phylogenetic analysis of AP2/EREBP genes**

Multiple sequence alignment of the deduced protein sequences was performed using MUSCLE included in MEGA 7 software (http://www.megasoftware.net/, Kumar et al. 2016) and then the phylogenetic tree based on alignments was generated using the Neighbor-Joining (NJ) method by MEGA 7 software with the following parameters: Poisson model, pairwise gap deletion and 1,000 bootstraps. The 147 AP2/EREBP genes identified in Arabidopsis (Dietz et al. 2010) were also included in the phylogenetic tree for comparison. Four subfamilies and a soloist group were further classified based on classification of Arabidopsis AP2/EREBP genes (Nakano et al. 2006; Dietz et al. 2010).

**Identification of CBF/DREB1-homologues**

Cucurbitaceae genes closely grouped with six Arabidopsis *CBF/DREB1* genes (AT1G12610, AT1G63030, AT4G25470, AT4G25480, AT4G25490 and AT5G51990) in the phylogenetic tree were selected as *CBF/DREB1*-homologues. The selected genes were further analyzed by
| Scientific name (common name) | Chromosome number and genome size | Cultivar/accession/line | Total annotated genes | Reference and genome database |
|-------------------------------|---------------------------------|-------------------------|-----------------------|-------------------------------|
| Cucumis sativus var. sativus (Cucumber) | 2n = 2x = 14 367 Mb | Chinese cultivar Chinese long Inbred line 9930 | 24,522 (25,600) | Huang et al. (2009) http://www.icugi.org/cgi-bin/ICuGI/index.cgi |
| | | North American cultivar Gy14 gynoecious inbred line | 21,503 | Cavagnararo et al. (2010) https://phytozome.jgi.doe.gov/ |
| | | North-European cultivar Borszczagowski line B10 | 26,587 (29,789) | Wóycicki et al. (2011) http://csgenome.sggw.pl/en-us/ |
| Cucumis sativus var. hardwickii (Wild cucumber) | Not available | Accession PI183967 | 22,746 (26,548) | Qi et al. (2013) http://www.icugi.org/cgi-bin/ICuGI/index.cgi |
| Cucumis melo (Melon) | 2n = 2x = 24 450 Mb | Double-haploid line DHL92 | 27,432 (34,848) | Garcia-Mas et al. (2012) https://melonomics.net/ |
| Citrullus lanatus (Watermelon) | 2n = 2x = 22 425 Mb | East Asia watermelon cultivar 97103 | 23,440 | Guo et al. (2013) http://www.icugi.org/cgi-bin/ICuGI/index.cgi |
| Momordica charantia (Bitter gourd) | 2n = 2x = 22 339 Mb | A monoecious inbred line OHB3-1 | 45,873 | Urasaki et al. (2016) |
| Arabidopsis thaliana | 2n = 2x = 10 125 Mb | | 27,416 | Arabidopsis Genome Initiative (2000) https://www.arabidopsis.org/ |

*Including alternative spliced forms.
*Including genes encoding less than 50 amino acid sequence.
*Including allelic forms and transposable elements.
*Only used to identify and compare TF gene families (Supplementary Table S1).

In silico expression profiling of cucumber AP2/EREBP genes

The expression of cucumber AP2/EREBP genes was investigated using RNA-Seq data (Bioproject accession no. PRJNA80169 (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA80169, Li et al. 2011) deposited in GenBank SRA database. First, the RNA-Seq data were retrieved and trimmed using NGS OC Tool kit (v2.3.1). Trimmed RNA-Seq reads of high quality (Phred score > 20) were mapped on protein-coding sequences of *C. sativus* cv. Chinese long gene set (25,600 genes) and then Fragment Per Kilobase of transcript per Million mapped reads (FPKM) values were calculated using RSEM program with default parameters. Next, FPKM values for the cucumber AP2/EREBP genes were selected and analyzed further. Hierarchical clustering based on expression pattern and heatmap drawing were performed using MeV s/w (Saeed et al. 2003) with default parameters.

RESULTS

Transcription factor and AP2/EREBP gene families in five Cucurbitaceae species

Eight gene sets from genome sequence data, three cucumber cultivars, wild cucumber, melon, watermelon, bitter gourd, and *A. thaliana*, were retrieved from public database and analyzed in this study (Table 1). First, all TF gene families were searched through genome-wide identification based on conserved TF domains. As a result, more than 1,200 TF genes were identified in each of the five Cucurbitaceae species (Supplementary Table S1). Gene numbers of most TF families were similar among five Cucurbitaceae species and *A. thaliana*, with some exceptions. Among the TF families, AP2/EREBP gene family was selected and analyzed further.

In cucumber, 144, 132, and 142 AP2/EREBP genes were identified from cucumber cv. Chinese long, cv. Gy14, and cv. Borszczagowski, respectively (Table 2). Among those, 144 genes in cucumber cv. Chinese long were reduced to 136 by removing alternative spliced (AS) forms. In wild
Fig. 1. Phylogenetic tree of AP2/EREBP genes identified in five Cucurbitaceae species and A. thaliana. Tree was based on similarity among deduced protein sequences of AP2/EREBP genes identified in four Cucurbitaceae species and A. thaliana. Four subfamilies, DREB, ERF, RAV, and AP2, and a soloist group were divided based on the classification of Arabidopsis genes (Nakano et al. 2006; Dietz et al. 2010). Cucumber AP2/EREBP genes identified in cucumber cv. Chinese long were marked with red circles, while ones in cucumber cv. Gy14, cv. Borszczagowski, and wild cucumber were not marked in the tree. The A1 subgroup including CBF/DREB1 transcription factors of the DREB subfamily is marked with a yellow triangle. Multiple sequence alignment of the protein sequences were performed using MUSCLE included in MEGA 7 software and then the phylogenetic tree was generated using the Neighbor-Joining (NJ) method by MEGA 7 software. Scale bar represents the number of amino acid substitution per site. The bootstrap support values are omitted for a legible illustration.

Table 2. AP2/EREBP gene families identified in this study.

| Plant name | Cultivar/accession/line | Total AP2/EREBP (ERF x) (DREB (CBF/DREB1) (AP2 RAV Soloist) |
|------------|-------------------------|---------------------------------------------------------------|
| Cucumber   | Chinese long            | 136 (144)                                                     |
|            | Gy14                    | 142 (145)                                                     |
|            | Borszczagowski          | 132 (135)                                                     |
|            | PII183967               | 146 (153)                                                     |
| Wild cucumber | DHL92                  | 140 (148)                                                     |
| Melon      | 97103                   | 145 (148)                                                     |
| Watermelon | OHB3-1                  | 109 (109)                                                     |
| Bitter gourd | OHB3-1                 | 147 (147)                                                     |
| Arabidopsis|                         |                                                               |

x) Including alternative spliced forms.

y) Based on classification by phylogenetic analysis without alternative spliced forms (Fig. 1).

z) CBF/DREB1-homologues identified by phylogenetic analysis (Fig. 2).

w) Reported in Dietz et al. (2010).

v) Only single gene (AT4G13040) was reported in A. thaliana (Nakano et al. 2006; Dietz et al. 2010).
cucumber and melon, 153 and 148 AP2/EREBP genes were identified respectively and 146 and 140 genes were utilized by removing AS forms. In addition, 145 and 109 AP2/EREBP genes were identified in watermelon and bitter gourd, respectively. These AP2/EREBP gene numbers were similar to those identified in this study and reported previously in *A. thaliana* [138 in TAIR database (https://www.arabidopsis.org/browse/genefamily/AP2EREBP.jsp); 141 in this study; 147 in Dietz et al (2010)]. Among the identified genes, only ones except for AS forms were used for further analysis.

Through phylogenetic analysis of the AP2/EREBP genes in five Cucurbitaceae species with 147 AP2/EREBP genes in *A. thaliana*, four subfamilies and a soloist group could be divided based on classification of Arabidopsis genes. Based on this classification, the gene members belonging to each of subfamilies/group could be identified (Table 2, Fig. 1, Supplementary Tables S2a-S2h). As reported in other plant AP2/EREBP gene family studies, most of the genes were included in ERF subfamily, followed by DREB and AP2 subfamilies. Only two or four genes were included in RAV subfamily and soloist group. Gene numbers in each of the subfamilies were not largely different among cucumber cultivars as well as among five Cucurbitaceae species and *A. thaliana*.

**CBF/DREB1**-homologous genes in five Cucurbitaceae species

After selection of genes closely grouped with Arabidopsis **CBF/DREB1** genes in the A1 subgroup of DREB subfamily in phylogenetic tree (Fig. 1), 25 Cucurbitaceae genes were identified as **CBF/DREB1**-homologues (Fig. 2). Among those, 14 genes were derived from four cucumber genomes, four, four, two and four genes from cucumber cv. Chinese long, cv. Gy14, cv. Borszczagowski, and wild cucumber, respectively. In addition, four, four, and three genes were identified in melon, watermelon, and bitter gourd, respectively. These genes showed 29% to 100% similarity to each other and 30% to 55% similarity to Arabidopsis **CBF/DREB1** genes at amino acid level.

The Cucurbitaceae **CBF/DREB1**-homologous genes could be divided into four clades (Fig. 2). Clade I and IV consisted of seven genes from each of used Cucurbitaceae species, while Clade II and III consisted of five and six genes, respectively, because the absence of gene from some Cucurbitaceae species. Among the four clades, Clade II was closely located with Arabidopsis CBF4/DREB1D protein sequence and Clade III was with Arabidopsis CBF1/DREB1B and CBF3/DREB1A protein sequences. **CBF/DREB1**-homologues derived from four cucumber genomes were almost identical to each other at amino acid level, with an exception of Cucsa.302750.1 in cucumber cv. Gy14 that had no translational start codon (Supplementary Table S3). Therefore, these **CBF/DREB1**-homologues were grouped together within each of four clades (Fig. 2). Furthermore, the cucumber genes were more closely located with those from congeneric species such as melon, than those from watermelon and bitter gourd.

The deduced protein sequences of all **CBF/DREB1**-homologues had a conserved nuclear localization signal (NLS), an AP2/EREBP domain, and CBF/DREB1 signatures, which were similar to those in Arabidopsis **CBF/DREB1** protein sequences (Supplementary Fig. S1). In particular, six genes in Clade III had almost same amino acid residues in the conserved regions as Arabidopsis CBF1/DREB1B and CBF3/DREB1A (Fig. 3). Furthermore, they showed high similarity of 55% to 60% to the two Arabidopsis **CBF/DREB1** genes at whole amino acid level.

**Expression profiles of cucumber AP2/EREBP genes**

To investigate expression patterns of cucumber AP2/EREBP genes, published cucumber cv. Chinese long RNA-Seq data including transcriptomes for ten cucumber tissues (leaf, stem, root, tendril, tendril base, female flower, male flower, ovary, unfertilized expanded ovary, and fertilized expanded ovary) were used. As a result of *in silico* expression profiling of 136 AP2/EREBP genes identified in cucumber cv. Chinese long, 125 genes were identified to be expressed (FPKM value > 0) in at least one of the ten examined tissues, whereas the remaining 11 genes showed no expression (FPKM value = 0) in all tissues (Supplementary Table S4). Therefore, the 125 genes were used for further analysis. The expression patterns of the 125 genes varied among ten tissues, which were roughly divided into 12 groups by hierarchical clustering (Fig. 4).
Fig. 2. Phylogenetic tree of CBF/DREB1-homologous genes identified in five Cucurbitaceae species. Phylogenetic tree of the DREB A1 subgroup in Fig. 1 was enlarged to show Cucurbitaceae genes closely grouped with Arabidopsis CBF/DREB1 genes at amino acid level. The four clades are marked with yellow boxes. The bootstrap support values more than 50 are shown in the branches. Scale bar represents the number of amino acid substitution per site. Clade I consists of Csa5M155570.1 (cv. Chinese long), Cucsa.302750.1 (cv. Gy14), gene_1#CSB10A_v1_contig_4381 (cv. Borszczagowski), evm.model.Chr5.580 (wild cucumber), MELO3C005629T1 (melon), Cla011488 (watermelon), and MOMC2_838 (bitter gourd). Clade II consists of Csa5M174570.1 (cv. Chinese long), Cucsa.166340.1 (cv. Gy14), evm.model.Chr5.834 (wild cucumber), MELO3C005367T1 (melon), and Cla002330 (watermelon). Clade III consists of Csa3M180260.1 (cv. Chinese long), Cucsa.251030.1 (cv. Gy14), evm.model.Chr3.1638 (wild cucumber), MELO3C006869T1 (melon), and MOMC16_201 (bitter gourd). Clade IV consists of Csa3M751440.1 (cv. Chinese long), Cucsa.378450.1 (cv. Gy14), gene_2#CSB10A_v1_contig_4841 (cv. Borszczagowski), evm.model.Chr3.3373 (wild cucumber), MELO3C009442T1 (melon), Cla017719 (watermelon), and MOMC32_103 (bitter gourd). Gene IDs of Arabidopsis CBF/DREB1 protein sequences in this tree are AT1G12610 (DREB1F), AT1G63030 (DREB1E), AT4G25470 (CBF2/DREB1C), AT4G25480 (CBF3/DREB1A), AT4G25490 (CBF1/DREB1B), and AT5G51990 (CBF4/DREB1D).
Fig. 3. Multiple alignment of the deduced protein sequences of CBF/DREB1-homologous genes in Clade III with Arabidopsis CBF1/DREB1B and CBF3/DREB1A protein sequences. The AP2/EREBP domain is marked by an over-line and the CBF/DREB1-conserved amino acid residues, PKRPAGRTKFRETRHP (in NLS), DSAWR (after AP2/EREBP domain), and LWSY (in C-terminal) are marked by black dots. Shaded boxes indicate conserved residues among compared protein sequences. The alignment was made using MUSCLE in MEGA 7 program and visualized using GeneDoc software. Gene IDs of protein sequences used for multiple alignment are Csa3M180260.1 (cv. Chinese long), Cucsa.251030.1 (cv. Gy14), evm.model.Chr3.1638 (wild cucumber), MELO3C006869T1 (melon), Cla006212 (watermelon), MOMC16_201 (bitter gourd), AT4G25480 (CBF3/DREB1A), and AT4G25490 (CBF1/DREB1B).

In addition, ten tissues could be divided into three groups based on similarity of gene expression patterns among tissues. Group I included only leaf tissue. Group II included six tissues such as root, ovary, expanded ovary (fertilized and unfertilized) and flower (female and male). Group III included three tissues such as stem, tendril and tendril base. Among the 125 genes, 8 [Csa2M279250.1 (AP2), Csa3M073900.1 (ERF), Csa3M646560.1 (ERF), Csa5M167110.1 (ERF), Csa6M124180.1 (DREB), Csa6-M167230.1 (ERF), Csa6M518040.1 (ERF), and Csa7-M448110.1 (ERF)] showed FPKM values more than 10 in all ten tissues, indicating they were constitutively expressed. Among four cucumber CBF/DREB1-homologous genes, three (Csa3M180260.1, Csa5M155570.1, and Csa5M174570.1) showed high expression mainly in tendril and tendril base, while the remaining one (Csa3M751440.1) was not highly expressed in any tissues examined.

DISCUSSION

Cucurbitaceae family includes many important crops such as cucumber and melon. The growth and yield of the crops are adversely affected by environmental stresses. However, gene family, in particular, TF important for stress-response, has not been actively studied in Cucurbitaceae species. On this account, the genome-wide study identified Cucurbitaceae AP2/EREBP gene family which is known to play various roles including stress response. To the best of our knowledge, this study is the first to present the results of comprehensive analysis of AP2/EREBP gene
Fig. 4. Expression profiles of 125 cucumber AP2/EREBP genes. Cucumber RNA-Seq data were retrieved from GenBank SRA database (Bioproject acc. PRJNA80169, Li et al. 2011) and employed to calculate FPKM values using RSEM program with default parameters (Supplementary Table S4). Hierarchical clustering based on expression pattern and heatmap drawing were performed using MeV s/w with default parameters. Color scale for expression level is shown at the bottom of heatmap. Eleven cucumber genes with FPKM values of zero were not included in this analysis. L, leaf; R, root; O1, ovary; O2, fertilized expanded ovary (7 days after flowering); O3, unfertilized expanded ovary (7 days after flowering); F1, female flower; F2, male flower; S, stem; T1, tendril; T2, tendril base. Three groups for tissues and 12 groups for genes are indicated on top and the right of heatmap, respectively.
families in Cucurbitaceae species.

In this study, more than 100 AP2/EREBP genes were identified from each of the five Cucurbitaceae species (Table 2). The 136 AP2/EREBP genes were identified in genome of cucumber cv. Chinese long, which was larger in gene number than those (110 and 103 genes) reported in previous cucumber genome-wide identification studies (Huang et al. 2009; Hu and Liu, 2011). This different gene number was probably due to difference in chosen tool for identification and/or improvement of cucumber genome annotation. In fact, gene annotation of cucumber cv. Chinese long genome has been improved using RNA-Seq data (Li et al. 2011) after the initial genome sequence was reported (Huang et al. 2009). This study used the improved version of annotated gene set of cucumber cv. Chinese long genome and thus presented more precise AP2/EREBP gene sequence in cucumber. Two ERF and a DREB gene were characterized in melon (Mizuno et al. 2006), all of which were included in 140 AP2/EREBP genes of melon identified in this study (Supplementary Table S2e). In addition, this study identified 145 watermelon AP2/EREBP genes, which were in accordance with those simply reported in watermelon genome paper (Guo et al. 2013). Overall, these results indicated AP2/EREBP genes identified in this study could cover almost all corresponding genes present in genomes of five Cucurbitaceae species.

Meanwhile, 109 AP2/EREBP genes were identified in bitter gourd, which were less in number than those (132-147) in other four Cucurbitaceae species and A. thaliana. This might be due to incomplete annotation of bitter gourd genome, considering that the bitter gourd genes were annotated only using ab initio prediction and included a large number of transposable elements (Urasaki et al. 2016).

AP2/EREBP genes identified in each Cucurbitaceae species were divided into five groups (Table 2, Fig. 1), as reported in other plant studies (Dietz et al. 2010; Lee et al. 2012). Among examined species, total AP2/EREBP gene number was not much different and also gene member number in each group was similar. This indicated the AP2/EREBP gene family was well conserved in genomes of Cucurbitaceae species and A. thaliana. The conservation of TF genes including AP2/EREBP genes was already well demonstrated in many plant genomes, such as Chinese cabbage (Lee et al. 2012) and carrot (Iorizzo et al. 2016).

In DREB subfamily, CBF/DREB1-homologous genes were identified in each of the five Cucurbitaceae species (Fig. 2). A CBF/DREB gene encodes TF which plays an important role in signaling and tolerance under abiotic stresses (Lata and Prasad 2011). Among the four clades of CBF/DREB1-homologous genes, Clade II and III were closely located with Arabidopsis CBF4/DREB1D and CBF1/DREB1B & CBF3/DREB1A, respectively. In addition, the protein sequences of the homologous genes had also well conserved amino acid residues as found in Arabidopsis CBF/DREB1 proteins. In A. thaliana, CBF4/DREB1D was induced by osmotic stresses such as drought and salt while CBF1/DREB1B and CBF3/DREB1A were induced by cold stress but not by drought and salt stress (Liu et al. 1998; Shinwari et al. 1998; Haake et al. 2002; Nakashima et al. 2009). Therefore, it was suggested that the CBF/DREB1-homologous genes might play similar role in Cucurbitaceae species as observed in case of Arabidopsis genes. It was also evidenced by Talanova et al. (2008) that CBF/DREB genes were highly expressed in cucumber plants treated with cold stress. Considering that over-expression of CBF/DREB genes enhanced tolerance against abiotic stresses to plants (Lata and Prasad 2011), CBF/DREB1-homologous genes identified in this study can be good molecular targets for developing a stress-tolerant Cucurbitaceae crops.

Although transcriptome data of Cucurbitaceae species are limited, in silico expression profiling of cucumber AP2/EREBP genes revealed various expression patterns in normal cucumber tissues (Fig. 4). The expression patterns reflect various biological roles of the AP2/EREBP genes. Further expression study and finding molecular evidence will be necessary to know stress-responsiveness of AP2/EREBP genes in Cucurbitaceae plants.

In conclusion, this study identified AP2/EREBP gene families and CBF/DREB1-homologous genes in five Cucurbitaceae species through genome-wide searches and phylogenetic analysis. In addition, various expression patterns of cucumber AP2/EREBP genes were revealed by in silico expression profiling. The AP2/EREBP gene families identified in this study will be valuable for
understanding stress response mechanism and facilitating molecular breeding in Cucurbitaceae crops.

ACKNOWLEDGEMENTS

We thank Prof. Hideo Matsumura in Shinshu University, Japan for kindly providing gene set of bitter gourd. This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Agri-Bio industry Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant number 116076-03-2-HD0b0).

REFERENCES

Ali A, Bang SW, Yang EM, Chung SM, Staub JE. 2014a. Putative paternal factors controlling chilling tolerance in Korean market-type cucumber (Cucumis sativus L.). Scientia Horticulturae 167: 145-148.

Ali A, Yang EM, Bang SW, Chung SM, Staub JE. 2014b. Assessment of chilling injury and molecular marker analysis in cucumber cultivars (Cucumis sativus L.). Kor. J. Hort. Sci. Technol. 32: 227-234.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815.

Bhowmick BK, Jha S. 2015. Dynamics of sex expression and chromosome diversity in Cucurbitaceae: a story in the making. J. Genet. 94: 793-808.

Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, et al. 2010. Genome-wide characterization of simple sequence repeats in cucumber (Cucumis sativus L.). BMC Genomics. 11: 569.

Dietz KJ, Vogel MO, Viehhauser A. 2010. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. Protoplasma. 245: 3-14.

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, et al. 2012. The genome of melon (Cucumis melo L.). Proc. Natl. Acad. Sci. USA. 109: 11872-11877.

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, et al. 2013. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat. Genet. 45: 51-58.

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ. 2002. Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant Physiol. 130: 639-648.

Hu L, Liu S. 2011. Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers. Genet. Mol. Biol. 34: 624-633.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. 2009. The genome of the cucumber, Cucumis sativus L. Nat. Genet. 41: 1275-1281.

Iorzio M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, et al. 2016. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat. Genet. 48: 657-666.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870-1874.

Lata C, Prasad M. 2011. Role of DREBs in regulation of abiotic stress responses in plants. J. Exp. Bot. 62: 4731-4748.

Lee SC, Lim MH, Yu JG, Park BS, Yang TJ. 2012. Genome-wide characterization of the CBF/DREB1 gene family in Brassica rapa. Plant Physiol. Biochem. 61: 142-152.

Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K. 2011. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. BMC Genomics 12: 540.

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, et al. 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.

Lough TJ, Lucas WJ. 2006. Integrative plant biology: role of phloem long-distance macromolecular trafficking. Annu. Rev. Plant Biol. 57: 203-232.

Lu A, Jeffrey C. 2011. Cucurbitaceae. In: ZY. Wu, PH. Raven, DY. Hong (eds.). Flora of China. Vol. 19. Science Press,
Mizuno S, Hirasaw Y, Sonoda M, Nakagawa H, Sato T. 2006. Isolation and characterization of three DREB/ERF-type transcription factors from melon (Cucumis melo). Plant Sci. 170: 1156-1163.

Nakano T, Suzuki K, Fujimura T, Shinshi H. 2006. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol. 140: 411-432.

Nakashima K, Ito Y, Yamaguchi-Shinozaki K. 2009. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant Physiol. 149: 88-95.

Qi J, Liu X, Shen D, Miao H, Xie B, Li X, et al. 2013. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat. Genet. 45: 1510-1515.

Rural Development administration (RDA). 2013. Cucumber cultivation-Agricultural technique guide 107. RDA. Korea

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374-378.

Shinwari ZK, Nakashima K, Miura S, Kasuga M, Seki M, Yamaguchi Shinozaki K, Shinozaki K. 1998. An Arabidopsis gene family encoding DRE/CRT binding s involved in low-temperature responsive gene expression. Biochem. Biophys. Res. Commun. 250: 161-170.

Talanova VV, Titov AF, Topchieva LV, Malysheva IE. 2008. Effect of stress factors on expression of the gene encoding a CBF transcription factor in cucumber plants. Doklady Biological Sciences, 423: 419-421.

Tanurdzic M, Banks JA. 2004. Sex-determining mechanisms in land plants. Plant Cell 16: S61-S71.

Urasaki N, Takagi H, Natsume S, Uemura A, Taniai N, Miyagi N, et al. 2016. Draft genome sequence of bitter gourd (Momordica charantia), a vegetable and medicinal plant in tropical and subtropical regions. DNA Res. 24: 51-58.

Wóycicki R, Włoszko J, Gawroński P, Dąbrowska J, Lomsadze A, Pawelkowicz M, et al. 2011. The genome sequence of the North-European cucumber (Cucumis sativus L.) unravels evolutionary adaptation mechanisms in plants. PLoS One 6: e22728.

Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, et al. 2016. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol. Plant. 9: 1667-1670.