Available online at www.sciencedirect.com

ScienceDirect

GENOMICS

Genomics 90 (2007) 265–275

www.elsevier.com/locate/ygeno

Phylogenetic and expression analysis of ZnF-AN1 genes in plants☆

Ying Jin [a,1], Meng Wang [a,1], Junjie Fu [a], Ning Xuan [a], Yun Zhu [a], Yun Lian [a], Zhiwei Jia [a], Jun Zheng [b], Guoying Wang [a,b,*]

[a] *State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100094, China*
[b] *Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China*

Received 4 February 2007; accepted 27 March 2007
Available online 23 May 2007

Abstract

In plants, ZnF-AN1 genes are part of a multigene family with 13 members in *Arabidopsis thaliana*, 19 members in *Populus trichocarpa*, 17 members in *Oryza sativa*, at least 11 members in *Zea mays,* and 2 members in Chlamydomonas reinhardtii. All ZnF-AN1 genes contain the ZnF-AN1 domain. According to the phylogenetic analysis of the ZnF-AN1 domain, we divided plant ZnF-AN1 genes into two types. The coding sequences of most type I members do not possess any introns, while most type II members do possess intron(s). Through Northern blot analysis of maize members and digital Northern analysis of *Arabidopsis* members, we found that most ZnF-AN1 genes are involved in responses to abiotic stresses. The evolutionary analysis indicated that the expansion rate of type I was higher than that of type II. After expansion, some ZnF-AN1 genes may have gained new functions, some may have lost their functions, and some were specialized to perform their functions in stress-specific or tissue-specific modes. In addition, we propose an evolutionary model of type II ZnF-AN1 genes in plants.

Keywords: ZnF-AN1; Evolution; Expression; Abiotic stress; Gene expansion

The ZnF-AN1 family is composed of proteins containing the ZnF-AN1 domain. In the Conserved Domain Database (CDD), the traditional pattern of a ZnF-AN1 domain is CX(2)CX(9,12)CX(1,2)CX(4)CX(2)HX(5)HXC. An expanded pattern, CX(4)CX(9,12)CX(1,2)CX(4)CX(2)HX(5)HXC, was also named as a ZnF-AN1 domain in a later study [1]. In plants, we found that the structure of proteins with the traditional pattern is obviously different from that of proteins with the expanded pattern. The members with the traditional pattern have only one ZnF-AN1 domain, whereas the members with the expanded pattern have two ZnF-AN1 domains. These ZnF-AN1 genes are widely expressed in plants.

Recently, the ZnF-AN1 gene family was identified in rice with 18 members and in *Arabidopsis* with 14 members [2]. In rice, most genes are induced under abiotic stresses. This family seems to play an important role in rice tolerance to abiotic stresses [2]. In other plants, the function of this family is still unclear. Meanwhile, little is known about their evolution and their relationship to evolution and function in plants. In animals, two ZnF-AN1 proteins, ZNF216 [3,4] and AWP1 (associated with PRK1) [5], have been widely studied. ZNF216 plays a role in regulating NFκB activation and apoptosis [4]. AWP1 may perform functions in mammalian signal transduction pathways [5].

In our study, additional ZnF-AN1 genes were identified from *Populus trichocarpa*, *Zea mays,* and *Chlamydomonas reinhardtii*. In plants, the exon–intron structure differed between members with the traditional pattern and the expanded pattern. All ZnF-AN1 genes were divided into several subclasses using phylogenetic and computational methods. This classification indicated that some members expanded in a monocot- or eudicot-specific manner. To explore the functions of plant ZnF-AN1 genes, we have characterized the expression profiles of *Arabidopsis* and maize genes, respectively, using digital Northern and Northern blot. The expression analysis suggested that most plant ZnF-AN1 genes are involved in abiotic stress responses. In addition, the evolution of ZnF-AN1 genes in

☆ Sequence data from this article have been deposited with the GenBank Data Libraries under Accession Nos. EF396223, EF396224, EF396225, EF396226, EF396227, EF396228, EF396229, and EF396230.

* Corresponding author. Fax: +86 10 62732012.
E-mail address: gywang@cau.edu.cn (G. Wang).

[1] These authors contributed equally to this paper.

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.03.019

plants with the expanded pattern was proposed. Our analysis will be helpful in exploring the functions of each gene in plant ZnF-AN1 families.

Results

Plant ZnF-AN1 genes were divided into two types

To explore the ZnF-AN1 family in plants, we first collected ZnF-AN1 genes from two eudicots, *Arabidopsis thaliana* and *P. trichocarpa*, and two monocots, *Oryza sativa* and *Z. mays*; 14 *Arabidopsis* and 18 rice ZnF-AN1 genes were previously reported [2]. However, we found that AtSAP14 (At5g48205) and OsSAP18 (Os07g07370) have no typical ZnF-AN1 domain, so we excluded these two genes from the ZnF-AN1 gene family. In maize and poplar, multiple searches were used to identify 11 and 19 ZnF-AN1 genes, respectively (Table 1). Nine maize ZnF-AN1 genes were confirmed by RT-PCR. Based on the chromosomal locations and the discoverable order, we reset and renamed the ZnF-AN1 genes in *Arabidopsis*, poplar, rice, and maize (Table 1, Table S1).

Since all ZnF-AN1 family members possess the ZnF-AN1 domain, we performed multiple alignment analysis using the amino acid sequences of the ZnF-AN1 domain (Fig. S1) and constructed a phylogenetic tree based on this alignment. As shown in Fig. 1, all ZnF-AN1 genes were divided into two types: type I and type II. Type I genes contain the traditional pattern, CX(2)CX(9,12)CX(1,2)CX(4)CX(2)HX(5)HXC. Type II genes contain the expanded pattern of the ZnF-AN1 domain, CX(4)CX(9,12)CX(1,2)CX(4)CX(2)HX(5)HXC. In total, 10 members from *Arabidopsis*, 15 members from rice, 9 members from maize, and 17 members from poplar have the traditional pattern of the ZnF-AN1 domain (Table 1, Table S1). These 51 members belong to type I genes. Meanwhile, the type II ZnF-AN1 genes include 9 members, 3 from *Arabidopsis*, 2 from rice, 2 from maize, and 2 from poplar (Table 1, Table S1).

Between type I and type II, not only is the pattern of the ZnF-AN1 domain different but the gene and protein structures are

Table 1
ZnF-AN1 genes in poplar and maize

Species	Gene name	Type	Accession No.[a]	Locus[b]	ZnF-AN1 domain[e]	ZnF-A20 domain[e]	Intron in ORF[c]	EST[d]
Poplar	*PtAN11*	I	grail3.0161001401	LG_I_2698226_2700798	1	1	NO	YES
	PtAN12	I	eugene3.00011004	LG_I_8651082_8653636	1	1	NO	YES
	PtAN13	I	gw1.I.1868.1	LG_I_19555679_19558284	1[f]	1	–	NO
	PtAN14	I	eugene3.00011778	LG_I_19562545_19565180	1	1	NO	YES
	PtAN15	I	estExt_Genewise1_v1.C_LG_III10	LG_III_11275753_11278328	1	1	NO	YES
	PtAN16	I	estExt_fgenesh4_kg.C_LG_III005	LG_III_16853931_16856479	1	1	NO	YES
	PtAN17	I	grail3.0045025401	LG_IV_16550767_16553443	1	1	NO	YES
	PtAN18	II	estExt_fgenesh4_pg.C_LG_VI0432	LG_VI_3478454_3482637	2	0	YES	YES
	PtAN19	I	gw1.VII.2246.1	LG_VII_5356920_5359518	1	1	NO	YES
	PtAN110	I	grail3.0001021601	LG_IX_1504850_1507493	1	1	NO	YES
	PtAN111	I	estExt_Genewise1_v1.C_LG_IX323	LG_IX_6439509_6442224	1	1	NO	YES
	PtAN112	II	gw1.XI.527.1	LG_XI_13387806_13390523	2	0	YES	NO
	PtAN113	I	eugene3.00111200	LG_XI_13720435_13724557	1	0	–	NO
	PtAN114	I	eugene3.00121217	LG_XII_13623964_13626446	1	1	NO	YES
	PtAN115	I	fgenesh4_pg.C_LG_XII001244	LG_XII_13626388_13628927	1	1	NO	NO
	PtAN116	I	gw1.XV.2999.1	LG_XV_9533052_9535651	1	1	NO	YES
	PtAN117	I	gw1.XV.3016.1	LG_XV_9559692_9562213	1	1	NO	NO
	PtAN118	I	estExt_Genewise1_v1.C_LG_XVI16	LG_XVI_3257671_3260368	1	1	NO	YES
	PtAN119	I	estExt_Genewise1_v1.C_1520011	scaffold_152_30634_33334	1	1	NO	YES
Maize	*ZmAN11*	I	DQ244747	–	1	1	NO	NO
	ZmAN12	I	EF396224	ZmGSStuc11-12-04.59743.1	1	1	NO	YES
	ZmAN13	I	EF396225	ZmGSStuc11-12-04.10004.1	1	1	NO	YES
	ZmAN14	I	EF396226	ZmGSStuc11-12-04.57324.2	1	1	NO	YES
	ZmAN15	I	EF396227	ZmGSStuc11-12-04.90226.1	1	1	NO	YES
	ZmAN16	I	–	ZmGSStuc11-12-04.7428.1	1	0	NO	NO
	ZmAN17	II	EF396228	ZmGSStuc11-12-04.35068.1	2	0	NO	YES
	ZmAN18	I	EF396229	ZmGSStuc11-12-04.19079.1	1	1	NO	YES
	ZmAN19	I	EF396230	ZmGSStuc11-12-04.41832.1	1	1	NO	YES
	ZmAN110	II	EF396223	ZmGSStuc11-12-04.15687.1	2	0	YES	YES
	ZmAN111	I	–	ZmGSStuc11-12-04.16050.1	1	1	NO	YES

The information that we can not confirm is marked by – symbol.

[a] Accession No. means the protein accession number in poplar, while it means GenBank accession number for nucleotide sequences in maize.

[b] In poplar, locus shows the genomic regions containing the ZnF-AN1 gene. In maize, locus represents the Genome Survey Sequences (GSSs) contigs from PlantGDB.

[c] Shows whether the genomic region of the coding sequences has introns or not.

[d] Corresponding EST is available or not.

[e] Represents the number of the domain in the ZnF-AN1 protein.

[f] Represents the defective ZnF-AN1 domain.

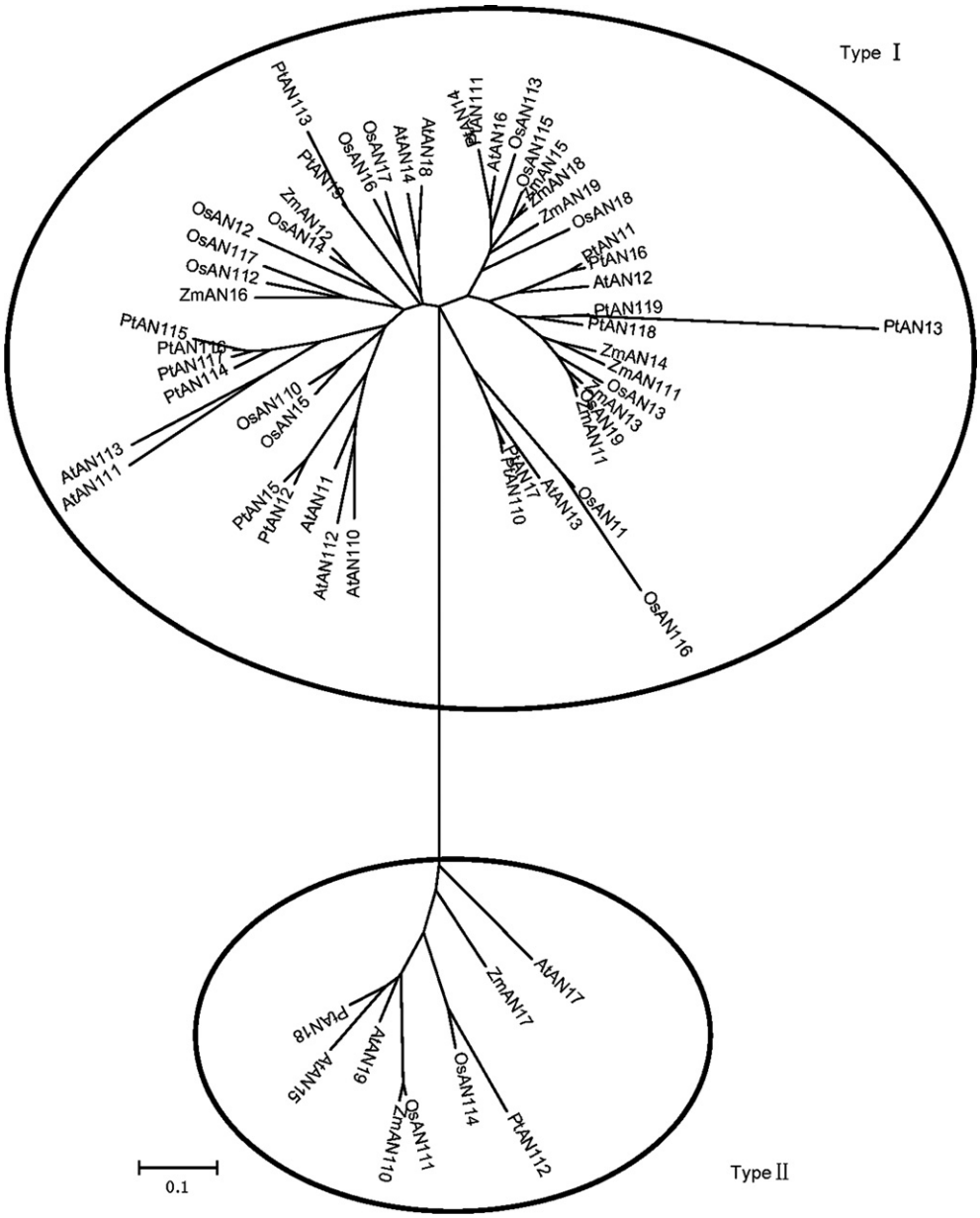


Fig. 1. An unrooted phylogenetic tree of 60 ZnF-AN1 proteins. The amino acid sequences of the ZnF-AN1 domain were aligned by ClustalX1.83 (Fig. S2), and the phylogenetic tree was constructed using the NJ method in MEGA 3.1.

also not alike. The data from the *Arabidopsis* genome annotation in TAIR, the poplar genome annotation in JGI, the rice genome annotation in TIGR, and the results of PCR in maize revealed that most type I members lack introns, while most type II members have introns (Table 1, Table S1, Fig. S2).

Conserved motifs outside of the ZnF-AN1 domain

To examine the protein structures, we used MEME [6] to search for conserved motifs. A total of 11 conserved motifs was identified in all 60 ZnF-AN1 proteins (Table S2). Through the Conserved Domain Search service (CD-Search) [7], we further characterized these motifs. Motif 1 was identified as the ZnF-AN1 domain, motif 2 was identified as the ZnF-A20 domain, and motif 11 was identified as a C_2H_2-type zinc finger. Although many motifs are shared by monocots and eudicots, specific motifs at the N terminus are also found (Table S3). Motifs 8 and 9 are eudicot specific, and motif 10 is monocot specific. In addition, some members (AtAN111, PtAN13, OsAN116, and OsAN117) have a defective ZnF-AN1 domain (Fig S3), which lost one or more zinc-coordinating Cys and/or His residues.

As shown in Fig. 2, the motif organization of type I is different from that of type II. In type I, most members have one intact ZnF-AN1 domain (motif 1) and one intact ZnF-A20 domain (motif 2). There are also several members (PtAN113, ZmAN16, OsAN17, OsAN18, and OsAN116) without any ZnF-A20 domains and one member (OsAN112) with two ZnF-A20 domains. In type II, all members have two intact ZnF-AN1

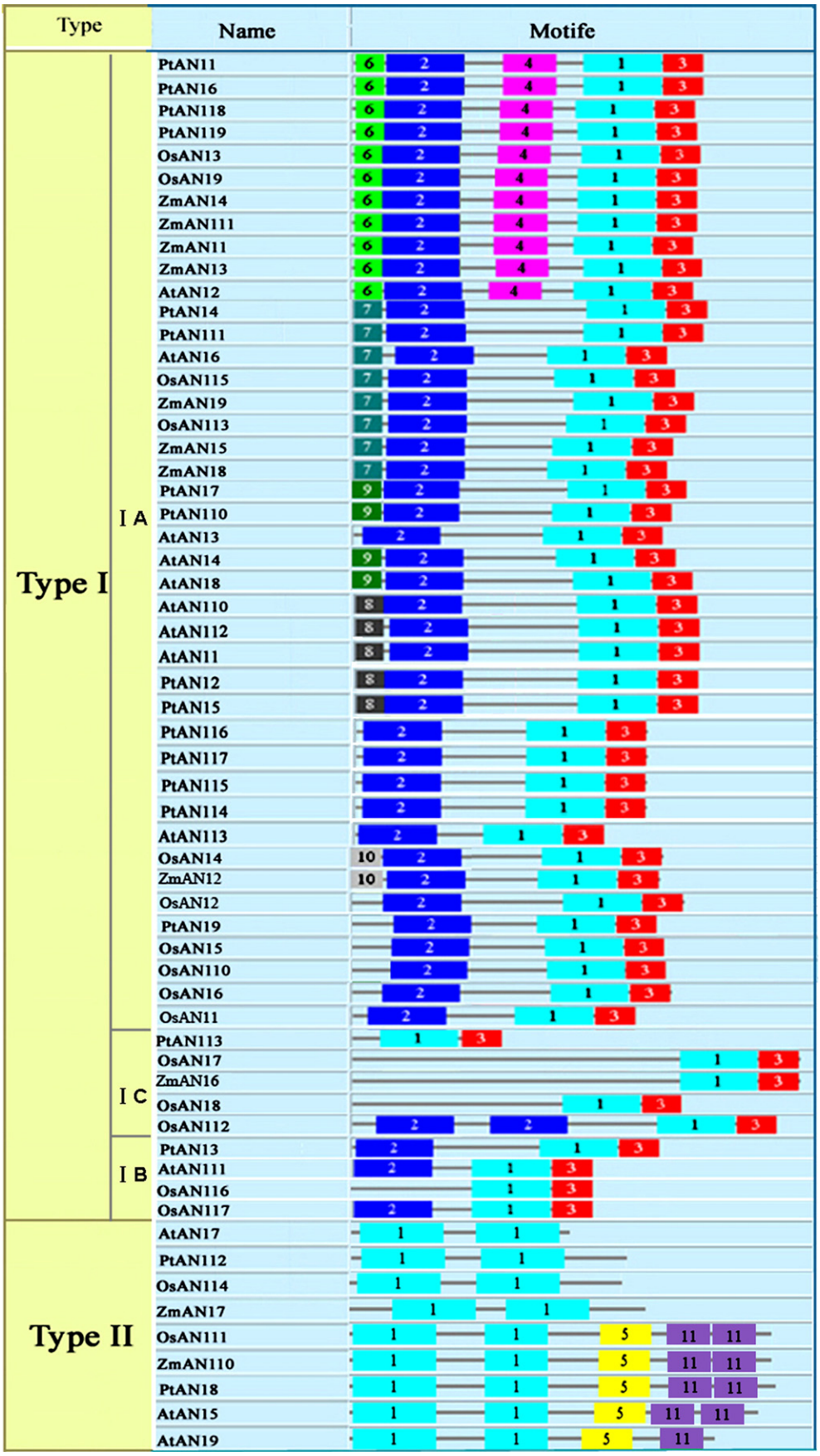


Fig. 2. Motif organization of two types of ZnF-AN1 proteins in plants.

domains (motif 1) but no ZnF-A20 domains. ZnF-AN1 (motif 1) is the only common domain for both type I and type II.

Characteristics of type I ZnF-AN1 proteins

According to the motif organization in Fig. 2, all type I ZnF-AN1s were classified into three subclasses, class IA, class IB, and class IC. Class IA includes 42 members, which contain one intact ZnF-A20 domain at the N terminus and one intact ZnF-AN1 domain at the C terminus. Class IB includes 4 members with a defective ZnF-AN1 domain and a ZnF-A20 domain except for OsAN116. Class IC includes 5 members containing zero or two ZnF-A20 domains.

To further evaluate the characteristics of class IA ZnF-AN1s, we constructed an unrooted tree by multiple alignment using amino acid sequences from the ZnF-A20 domain and the whole C terminus (the ZnF-AN1 domain and motif 3) (Fig. S4). According to the phylogenetic tree, we classified class IA into several subclasses (Fig. 3A). Each subclass has well-supported bootstrap values (>50%) except class IA4, all members of which have a conserved motif (motif 8) at the N terminus (Fig. 2). Classes IA1 and IA2 include members from both monocots and eudicots, classes IA3, IA4, and IA5 are eudicot specific, and class IA6 is monocot specific. This finding suggested that most ZnF-AN1 genes expanded in a monocot- or eudicot-specific manner. In addition, many duplication pairs, which have very strong bootstrap (>70%), were identified in class IA, especially in poplar (Fig. 3A).

Characteristics of type II ZnF-AN1 proteins

To further examine the characteristics of type II, we constructed a phylogenetic tree using a multiple alignment of protein sequences from all nine members (Fig. 3B, Fig. S5). According to the phylogenetic tree, we divided type II into two subclasses: class IIA and class IIB (Fig. 3B). Each subclass includes members from monocots and eudicots, which suggested that the origination of type II genes in plants occurred before the divergence of monocots and eudicots.

In addition to the ZnF-AN1 domain, other conserved motifs, motif 5 and motif 11 (C_2H_2-type zinc finger), were identified in class IIA proteins but not in class IIB proteins (Fig. 2). Except for AtAN19, which contains only one C_2H_2-type zinc finger [2], all members of class IIA have two C_2H_2-type zinc fingers at their C terminus. However, we could not confirm the structure of PtAN112 because we did not obtain the full-length amino acid sequence of PtAN112.

Expression analysis of type I genes

Most rice type I genes have been demonstrated to be induced by one or more abiotic stresses [2]. Publicly available data from rice massively parallel signature sequencing (MPSS) (20-bp signatures) (http://mpss.udel.edu/rice/) further supported their abiotic-stress-inducing expression profiles (Table S4). This result indicates that most rice type I genes may perform functions to defend against abiotic stresses. To explore the functions of type I ZnF-AN1 genes in other plants, we examined their expression profiles under abiotic stresses in maize and *Arabidopsis* using Northern blot and digital Northern analysis, respectively.

Under cold stress, most type I genes of maize are induced with different expression patterns (Fig. 4A). For example, at 0.5 h after cold treatment, *ZmAN15* was induced, while *ZmAN18* was induced at after 8 h of cold treatment. Most maize type I genes have not only different temporal expression patterns but also different tissue expression patterns. EST profile viewer in NCBI was used to discover tissue-specific expression profiles. The majority of maize genes are expressed at varying levels in various tissues (Table S5). As shown in the table, *ZmAN15*, *ZmAN18*, and *ZmAN19* are expressed in the glume, flower, and ear, respectively.

In *Arabidopsis*, the expression database on the AtGenExpress Visualization Tool (http://jsp.weigelworld.org/expviz/) was used. We surveyed their expression profiles under cold, drought, salt, and heat stresses (Fig. 5). Three main patterns were observed: (1) no response to any stresses, (2) response to several stresses, and (3) response to specific stress. For example, *AtAN12* and *AtAN16* did not show responses to the stresses analyzed, *AtAN18* could be induced by cold, drought, and salt, and *AtAN13* was induced only by heat stress. In addition, *AtAN113* was elicited only in roots under all abiotic stresses.

Overall, in plants, most type I genes are involved in abiotic stress responses. They perform functions in a stress-specific or tissue-specific manner.

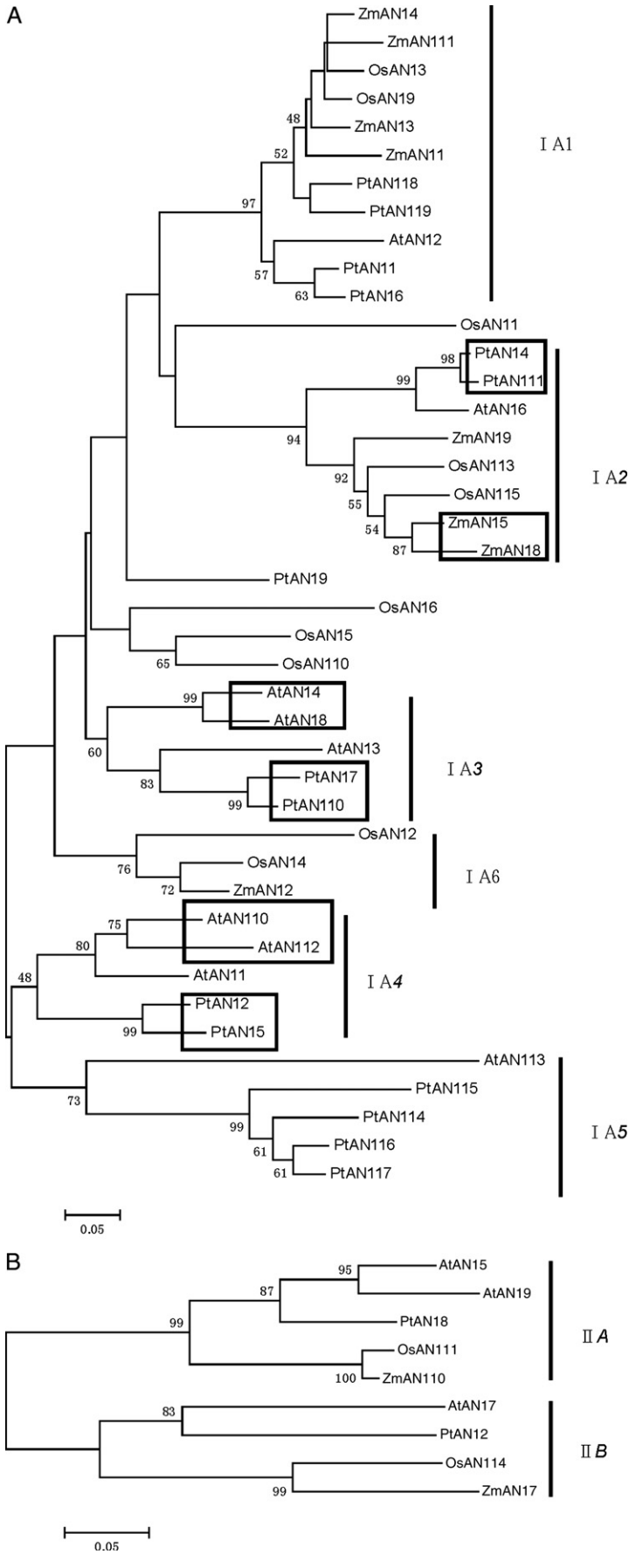
Expression analysis of type II genes

Two rice type II genes, *OsAN111* and *OsAN114*, have been demonstrated to be induced by salt and drought stresses [2]. In *Arabidopsis*, according to the expression database on the AtGenExpress Visualization Tool, the expression of *AtAN19* is stable under various abiotic stresses, whereas *AtAN17* can be induced by many stresses (Fig. 5). In maize, Northern blot analysis showed that *ZmAN17* was elicited by cold stress and that the expression of *ZmAN110* was stable (Fig. 4B). Combining the phylogenetic analysis of type II genes, it is suggested that all class IIB members possibly play roles in response to abiotic stresses.

Discussion

History of two type ZnF-AN1 genes

It is believed that land plants originated from the green alga. Chlamydomonas reinhardtii is a kind of single-celled green algae. Therefore, it is important to investigate the expression of ZnF-AN1 genes in *C. reinhardtii*. The C. reinhardtii genome version 3.0 (http://genome.jgi-psf.org/Chlre3/) was searched using the 13 *Arabidopsis* ZnF-AN1 amino acid sequences as direct queries. We found only two nonredundant gene models, e_gwH.25.97.1 and e_gwW.22.154.1. They were named *CrAN11* and *CrAN12*, respectively. *CrAN11* contains the traditional pattern of a ZnF-AN1 domain and has no introns in the


A
ZmAN14
ZmAN111
OsAN13
OsAN19
ZmAN13
ZmAN11
PtAN118
PtAN119
AtAN12
PtAN11
PtAN16
I A1
OsAN11
PtAN14
PtAN111
AtAN16
ZmAN19
OsAN113
OsAN115
ZmAN15
ZmAN18
I A2
PtAN19
OsAN16
OsAN15
OsAN110
AtAN14
AtAN18
AtAN13
PtAN17
PtAN110
I A3
OsAN12
OsAN14
ZmAN12
I A6
AtAN110
AtAN112
AtAN11
PtAN12
PtAN15
I A4
AtAN113
PtAN115
PtAN114
PtAN116
PtAN117
I A5
0.05
B
AtAN15
AtAN19
PtAN18
OsAN111
ZmAN110
II A
AtAN17
PtAN12
OsAN114
ZmAN17
II B
0.05

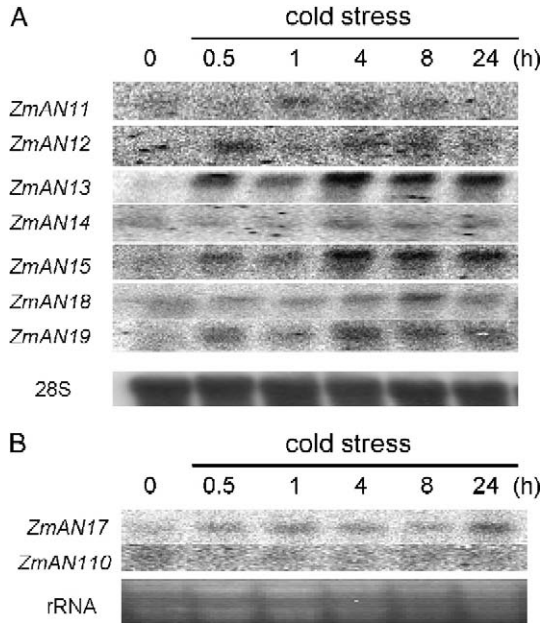


Fig. 4. Northern blot analysis of maize ZnF-AN1 genes under cold stress. (A) Expression pattern of the maize type I genes. 28S rRNA was used as an internal standard. (B) Expression pattern of the maize typeII genes. Ethidium-bromide-stained rRNA was used for equivalent loading.

coding sequences. *CrAN12* contains the expanded pattern of a ZnF-AN1 domain and has introns. Because we could not obtain the full-length cDNA of *CrAN12*, the sequence of CrAN12 was deduced from partial CDS. This fragment of CrAN12 has two intact ZnF-AN1 domains and one intact and one partial C_2H_2-type zinc fingers. Perhaps the entire CrAN12 protein contains two intact C_2H_2-type zinc fingers. Protein structures of CrAN11 and CrAN12 are similar to higher plant type I and type II genes, respectively. For exploring the expression of ZnF-AN1 genes in other green algae, ESTs of green algae excluding *Chlamydomonas*, were searched. We obtained three nonredundant type I ESTs and two nonredundant type II ESTs (Table S7). These results indicated that the two types of ZnF-AN1 genes came into being before the generation of higher plants and they expanded in different ways during the long history of evolution.

In yeast and humans, the two types of ZnF-AN1 genes also exist. Yeast contains one type I gene and one type II gene. Humans have six nonredundant ZnF-AN1 genes in which there are three type I genes and three type II genes. However, in yeast and humans, only two human type I genes have protein structures similar to those of plants, and all type II genes have protein structures different from those of plants. Thus, we considered that the protein structures of plant type II genes are specific to green plants.

It is well known that most genes are duplicated along with species evolution. In plants, partly due to the frequent occurrence of genomic tandem duplication [8], segmental duplications [9], and polyploidization events [10], a large fraction of their genomes consist of duplicate loci [11]. Under evolutionary selection, some duplicated genes were retained and were inclined to functional diversity by mutation. These gene duplications enable plants to survive in the face of environmental changes. Our study revealed that ZnF-AN1 genes in plants are part of a multigene family with 13 members in *Arabidopsis*, 19 members in poplar, 17 members in rice and at least 11 members in maize. Among the 60 ZnF-AN1 genes from the four plant species, 51 members are type I genes and only 9 members are type II genes. This indicated that the expansion rate of type I genes is much higher than that of type II genes.

After gene expansion, most members of type I genes (42/51) retained their ancestral protein structure with one intact ZnF-A20 domain at the N terminus and one intact ZnF-AN1 domain at the C terminus. Five members lost the ZnF-A20 domain and four members lost one or more zinc-coordinating Cys and/or His residues of the ZnF-AN1 domain.

During gene evolution, some members of type II genes lost their C_2H_2-type zinc fingers, which were divided into Class IIB. To further investigate the evolutionary process of type II genes, we searched the PUT database of moss *Physcomitrella patens* in plantGDB and obtained a member (PUT-155a-Physcomitrella_patens-23098) that contains the full-length CDS. The protein predicted by this PUT has two ZnF-AN1 domains, while it has no C_2H_2-type zinc fingers. This result indicated that divergence of type II ZnF-AN1 genes may have occurred before the generation of land plants.

Combining the motif organization, we predicted an evolutionary model of type II genes. As shown in Fig. 6, the ancestor of type II genes contains two ZnF-AN1 domains and two C_2H_2-type zinc fingers. After duplication, the different evolutionary forces produced the diversity of type II genes. High selection pressure made Class IIA members retain the protein structure of their ancestor, while low pressure made Class IIB members lose the C_2H_2-type zinc fingers.

Gain or lose function by mutation

According to the gene duplication theory of Ohno [32], one of the duplicates either loses (pseudogenazation) or gains (neofunctionalization) a new function by mutation [12]. In class IB, the ZnF-AN1 domain was mutated. In class IC, the ZnF-A20 domain was mutated. BLAST searches in the NCBI EST database showed that all of them have no EST perfectly matched, except two class IC members, *OsAN17* and *OsAN18* (Table 1, Table S1). *OsAN17* can be highly induced by salt and dehydration stresses, and *OsAN18* can be induced by drought stress [2]. However, the expression of all rice class IB members cannot be detected by rice MPSS (Table S4). In addition, an intact CDS-containing EST (Accession No. CN007875), which

Fig. 3. Phylogenetic trees of class I A ZnF-AN1 proteins and type II ZnF-AN1 proteins. (A) Unrooted phylogenetic tree of 42 class IA ZnF-AN1 proteins. Bootstrap values (>40%) from 1000 replicates are shown at the nodes. The classification of some proteins is indicated beside the tree. Some duplication pairs are showed in rectangles. (B) Unrooted phylogenetic tree of 9 type II ZnF-AN1 proteins. Bootstrap values (>50%) from 1000 replicates are shown at the nodes. The classification is indicated beside the tree.

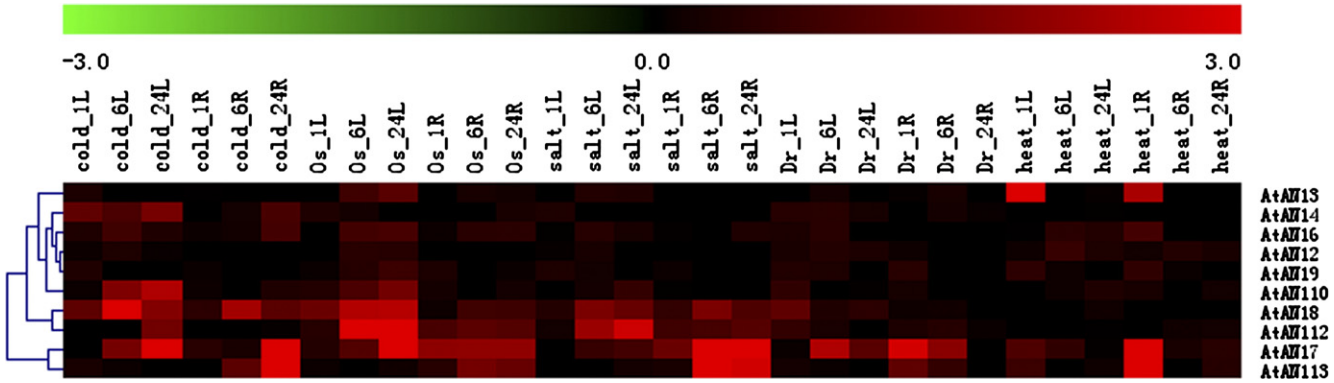


Fig. 5. Many AtAN1 genes respond to abiotic stresses. Microarray data from the AtGenExpress abiotic stress dataset for AtAN1 genes was extracted using AtGenExpress Visualization Tool and clustered using hierarchical clustering with average linkage. Expression is indicated as log2 of the fold changes.

was from abiotic stress cDNA libraries, was identified with BLAST searches in the NCBI *Triticum aestivum* EST database and the predicted protein from it was highly similar to OsAN18. According to these findings, we could draw two hypotheses: (1) proteins losing ZnF-A20 domains may gain new functions and (2) proteins containing the defective ZnF-AN1 domains may lose their functions.

According to the Northern blot, digital Northern, and EST searches, all members of type II genes are expressed. Moreover, class IIB members without C_2H_2-type zinc fingers possibly obtain new biochemical functions different from those of class IIA.

Divergence of expression patterns in class IA

In type I, class IA genes exhibited high expansion rates. After gene expansion, duplication pairs must perform at least partial nonoverlapping functions [12]. A large number of ZnF-AN1 class IA genes in plants suggested that these duplication products possibly play different roles.

Diverse expression patterns of duplication pairs

Recently, Force et al. [13] proposed a duplication-degeneration-complementation (DCC) evolutionary model. Under this model, the majority of duplicated genes undergo functional specialization through complementary partition of ancestral functions. We selected several duplication pairs (paralogs) to explore their expression patterns (Fig. 7). For example, in *AtAN14* and *AtAN18* pairs, the expression of *AtAN18* is induced under several stresses, while *AtAN14* is induced only by cold stress, in *AtAN110* and *AtAN112* pairs, salt stress elicits only the expression of *AtAN112*, and, although the expression patterns of *OsAN113* and *OsAN115* pairs are similar under abiotic stresses, their tissue expression patterns are diverse (Table S4) [30]. These results indicated that gene duplications expanded their expression models. The expansion of expression models makes

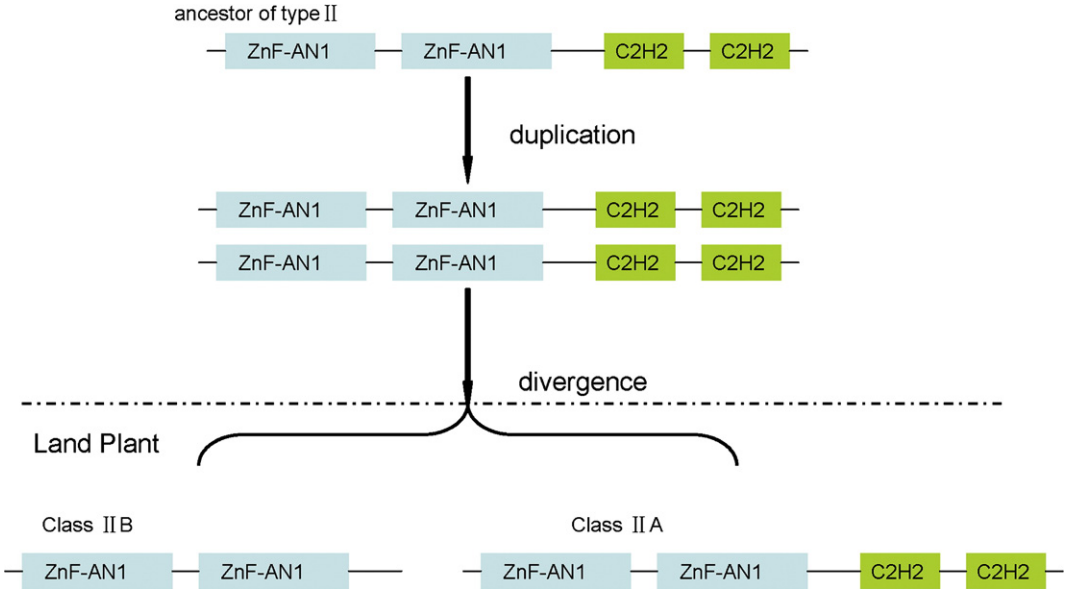


Fig. 6. Possible evolution model of type II ZnF-AN1 proteins.

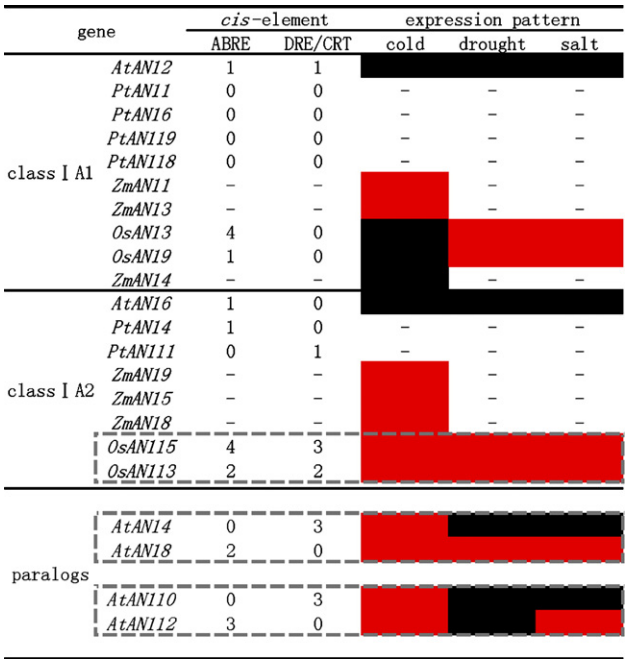


Fig. 7. Diverse expression patterns in class IA. In *Arabidopsis*, the expression patterns were obtained from AtGenExpress Visualization Tool. In rice, the expression patterns were obtained from previous reports. In maize, the expression patterns were obtained from our Northern blot results. The increased level more than twice is indicated by red. The stable expression level is indicated by black. It was also considered to be stable that the changed level is more than 0.5-fold and lower than 2-fold. The results that we did not get or could not confirm are indicated by – symbol. Three duplication pairs are indicated in rectangles. The number of *cis*-elements, ABRE and DRE/CRT, were obtained from PLACE.

the function of each sister gene nonoverlapping by distinct stress-specific or tissue-specific expression.

Diverse abiotic stress expression patterns in monocots and eudicots

In class IA, class IA1 and class IA2 are nonspecific in monocots and eudicots. Although monocots and eudicots have similar protein structures in each subclass, the expression patterns of genes from monocots and eudicots are different (Fig. 7). Under many abiotic stresses, the expression of two eudicot members, *AtAN12* and *AtAN16*, are unchangeable; nevertheless most monocot genes can be induced by one or more abiotic stresses.

To further recognize the expression pattern under abiotic stresses, 1000-bp genomic sequences, upstream of the translation start site, were used to search for ABA-responsive elements (ABRE) and C-repeat (CRT)/dehydration responsive elements (DRE) [14,15] by PLACE [16]. The results indicated that all eudicot members of classes IA1 and IA2 did not have any ABRE or DRE in their upstream genomic sequences, but most monocot members had them.

In addition, overexpression of *OsAN115* (*OSISAP1*), one monocot member of class IA2, could facilitate the tolerance to many abiotic stresses in the transgenic tobacco plants [1]. This implied that the proteins with similar structures in monocots and eudicots may have similar biochemical functions despite the diverse expression patterns.

Rapid evolutionary rate in the internal regions of class IA genes

According to the protein motif organization, the internal regions of all class IA members are very diverse (Fig. 2). Even in recent duplication pairs, such as AtAN14/AtAN18 and PtAN12/PtAN15 pairs, the internal regions are also highly variable (Fig. S6). Variability of the internal regions is possibly due to low alternative measurements. It is well known that the nonsynonymous/synonymous substitution ratio (Ka/Ks) indicates an alternative measure of evolutionary pressure. It has been observed that the elevated Ka/Ks ratio can result in rapid molecular evolution [17]. So we calculated the Ka/Ks ratio in some recent duplication paralogs (Fig. 3A) using SWAKK (http://oxytricha.princeton.edu/SWAKK/) [18]. The Ka/Ks ratio of the internal regions is much higher than that of other regions (data not shown). This result implied that low alternative pressure induced the rapid divergence of internal regions. Such rapid divergence was also found in C-terminal regions of many R2R3 Myb proteins, and the high degree of sequence divergence was coupled with the absence of functional means [19]. Perhaps, the diverse regions in class IA are not important to their functions.

There were Ser or other amino-acid-rich sequences in the internal regions of some ZnF-AN1 proteins, such as AtAN18, PtAN19, OsAN113, ZmAN19, etc. Previous studies have found that many domain linkers contain specific amino-acid-rich sequences, for example, polyserine linkers (PSLs) contain Ser-rich sequences [20]. Domain linkers were reported to allow proteins to optimize the spatial structures of active domains with a flexible region [21,22]. Moreover, Ser-rich linkers of SMAD proteins are diverse and can be phosphorylated [23]. The meanings of internal diverse regions with specific amino-acid-rich sequences in ZnF-AN1 proteins are still unclear, which need further in-depth study.

Putative biochemical functions of ZnF-AN1 proteins in plants?

Most class IA ZnF-AN1 proteins are similar in animals and plants [1]. There are many homologous proteins in animals, such as the human ZNF216 [4] and AWP1 (associated with PRK1) [5]. Domain mapping experiments have indicated that the ZnF-AN1 domain of ZNF216 interacts with TRAF6 and the ZnF-A20 domain interacts with Ikkγ and RIP. Thereby ZNF216 can inhibit the activity of transcription factor NFκB and is involved in immune regulation and inflammation. Moreover, the N-terminal ZnF-A20 and C-terminal ZnF-AN1 domains of ZNF216 can interact with each other [4]. In muscle cells, ZNF216 directly binds to polyubiquitin chains through its N-terminal ZnF-A20 domain and associates with the 26S proteasome to degrade specific cellular proteins [3]. Little research has been done in plant ZnF-AN1 genes at present. Mukhopadhyay and colleagues [1] reported that overexpression of a ZnF-AN1 gene (*OsSAP1*) could increase the tolerance of

transgenic tobacco plants to abiotic stresses. Because ZnF-AN1 proteins lack any typical nuclear localization signal, it is suggested that they may use their ZnF-AN1 or ZnF-A20 domains for protein-protein interactions to perform their functions [1]. Whether some factors can interact with ZnF-AN1 or ZnF-A20 domains in plants, whether plant class IA proteins play the same biochemical function as ZNF216, and how these processes occur still require further investigation.

Due to the similar protein structures of class IA members, they may perform the same biochemical functions, despite the diversity of expression patterns. When one member is knocked out, another may substitute it to play the same roles. We have ordered some *AtAN1* loss-of-function mutants from ABRC. Unfortunately, all homozygous mutants failed to show obviously visible phenotypic differences from wild type under drought, salt, osmotic, and ABA treatments (data not shown). Biochemical functional redundancy may lessen the effects of a single mutant. Analysis of gain-of-function plants may be a useful method to study this gene family.

In Class IIA, all members have a C_2H_2-type zinc finger. Many studies reported that most C_2H_2-type zinc finger genes are involved in transcriptional regulation [24]. So, we presumed that members of class IIA possibly act as transcription factors. Due to the absence of any typical nuclear localization signal, class IIA members may use the ZnF-AN1 domain to bind other proteins to perform transcriptional functions.

Overall, some type I members may play a role in the ubiquitin pathway, and some type II members may play a role in the process of transcriptional regulation.

Conclusions

We described the phylogenetic analysis and expression profiles of ZnF-AN1 genes in plants. Most ZnF-AN1 genes are involved in responses to abiotic stresses. All genes are obviously divided into two types. Both types of genes expanded in different manners. The expansion rate of type I is higher than that of type II. After expansion, some genes of type I may gain new functions by the absence of ZnF-A20 domains, some may lose function by the mutation of ZnF-AN1 domains, and some may be specialized to perform ancestral functions in stress-specific or tissue-specific modes. Class IIB genes of type II lost two C_2H_2-type zinc fingers after expansion and probably developed novel functions.

Materials and methods

Plant materials and growth conditions

Seeds of maize inbred line Han 21 were sterilized in 5% sodium hypochloride for 20 min, washed six times (2 min each) in distilled water, and then germinated on Whatman paper saturated with water for 6 days with a 16-h photoperiod at 26 °C. The seedlings were transferred to hydroponic growth conditions. Plants were grown in troughs filled with aerated nutrient solution under controlled conditions (28 °C day/26 °C night, 16-h photoperiod, 500 mmol m^{-2} s^{-1} photons, and 80% relative humidity). The solution, which was described in a previous report [25], was changed every 2 days.

Three-leaf-stage seedlings were treated with low temperature. Seedlings were transferred to an incubator at 4 °C with the same light and humidity conditions. Cold-stressed leaves were harvested at 0, 0.5, 1, 4, 8, and 24 h of treatment and prepared for the Northern blot.

ZnF-AN1 gene discovery

Arabidopsis and rice

A recent study reported 14 SAP genes in *Arabidopsis* and 18 SAP genes in rice [2]. However, we found that AtSAP14 (At5g48205) and OsSAP18 (Os07g07370) have no typical ZnF-AN1 domain. Thus we excluded these 2 genes from the ZnF-AN1 gene family.

Poplar

To obtain poplar ZnF-AN1 genes, all AtAN1 protein sequences were used to search the *Populus* genome assembly version1.1 (http://genome.jgi-psf.org/) using blastp and tblastn. Six gene models did not have full-length amino acid sequences. We reannotated genomic sequences of two members, *PtAN116 and PtAN117*, according to the protein similarity and obtained their full-length protein sequences. After searching the EST database in NCBI, the full-length cDNA of *PtAN111* was obtained. We did not obtain the full-length amino acid sequences of three others, *PtAN12, PtAN112,* and *PtAN113*. Another 13 gene model sequences were directly used. The information on PtAN1 EST was obtained from GenBank. To determine whether *PtAN113* contained the ZnF-A20 domain, we tested 3-kb genomic sequences that were located upstream of the *PtAN113* coding sequences. We did not find any pattern of ZnF-A20.

Maize

To identify ZnF-AN1 genes in maize, tblastn in PlantGDB (http://www.plantgdb.org) was performed using the ZnF-AN1 domains of all OsAN1 protein sequences as queries. According to the information on maize sequences (EST, PUT, and GSS), we designed specific primers to amplify maize cDNA. *ZmAN11* was obtained from our previous maize full-length cDNA library [26]. Through RT-PCR, we obtained eight *ZmAN1* cDNA sequences, which had been deposited at GenBank. We obtained these nine ZmAN1 protein sequences from their full-length cDNA. The full-length amino acid sequences of two putative ZmAN1s, ZmAN16 and ZmAN111, were deduced from their genomic sequences. Primers of eight *ZmAN1* genes are shown in Table S6.

Phylogenetic analysis

Multiple sequence alignment was performed using ClustalX1.83 [27] with default parameters and was manually adjusted with Jalview (2.07) [28]. A phylogenetic tree was constructed using the neighbor joining method in MEGA (3.1) [29]. The reliability of different phylogenetic groupings was evaluated by using the bootstrap test (1000 bootstrap replications) available in MEGA (3.1).

Motif identification

We identified the protein motifs of two types of ZnF-AN1 genes using MEME (http://meme.sdsc.edu/meme/meme.html) with the motif length set at 6–80, motif sites 2–200, and e-value < 1e-6. The results were then manually adjusted. Furthermore, we characterized the motifs using the CD-Search [7].

Promoter analysis

For all *Arabidopsis*, poplar, and rice ZnF-AN1 genes, 1000 bp genomic upstream of the translation start site was considered to be the promoter. We used PLACE (http://www.dna.affrc.go.jp/PLACE/) [16], a database of plant *cis*-elements, to search ABRE and DRE elements.

Gene expression

The abiotic response expression analysis of *Arabidopsis* genes was done with the AtGenExpress Visualization Tool (http://jsp.weigelworld.org/expviz/expviz.jsp). The expression is indicated as log2 of fold-change relative to a control treatment. Data were clustered using hierarchical clustering, Euclidean distance, and average linkage with the TIGR MeV software package [30]. For the cold, salt, drought, and heat treatments, experiments at 1, 6, and 24 h of treatment were

combined relative to controls. The expression analysis of rice genes was performed using MPSS (20-bp signatures) (http://mpss.udel.edu/rice).

Abiotic response expression analysis of maize genes was done with Northern blot. Total RNA for cold treatment was extracted from the leaves of the samples using a hot phenol method [31]. Analysis by Northern blot was followed as described in the previous report [25]. All nine cDNAs that we obtained through PCR were used as probes. Primers of eight ZmAN1 genes are the same as the primers in Table S6. Primers of ZmAN11 gene are also shown in Table S6. Each lane contained 20 μg total RNA.

Acknowledgments

This work was supported by the National High-tech Program (2006AA10A106) and the Cultivation Fund of the Key Scientific and Technical Innovation Project, China Ministry of Education (No: 705009).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.03.019.

References

[1] A. Mukhopadhyay, S. Vij, A.K. Tyagi, Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco, Proc. Natl. Acad. Sci. USA 101 (2004) 6309–6314.

[2] S. Vij, A.K. Tyagi, Genome-wide analysis of the stress associated protein (SAP) gene family containing A20/AN1 zinc-finger(s) in rice and their phylogenetic relationship with Arabidopsis, Mol. Genet. Genomics 276 (2006) 565–575.

[3] A. Hishiya, et al., A novel ubiquitin-binding protein ZNF216 functioning in muscle atrophy, EMBO J. 25 (2006) 554–564.

[4] J. Huang, et al., ZNF216 Is an A20-like and $I_{\kappa}B$ Kinase γ-interacting inhibitor of $NF_{\kappa}B$ activation, J. Biol. Chem. 279 (2004) 16847–16853.

[5] W. Duan, et al., Cloning and characterization of AWP1, a novel protein that associates with serine/threonine kinase PRK1 in vivo, Gene 256 (2000) 113–121.

[6] T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proc. Int. Conf. Intell. Syst. Mol. Biol. 2 (1994) 28–36.

[7] A. Marchler-Bauer, S.H. Bryant, CD-Search: protein domain annotations on the fly, Nucleic. Acids Res. 32 (2004) W327–W331.

[8] M. Jain, A.K. Tyagi, J.P. Khurana, Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*), Genomics 88 (2006) 360–371.

[9] J. Kim, et al., Patterns of expansion and expression divergence in the plant polygalacturonase gene family, Genome Biol. 7 (2006) R87.

[10] K.L. Adams, J.F. Wendel, Polyploidy and genome evolution in plants, Curr. Opin. Plant Biol. 8 (2005) 135–141.

[11] R.C. Moore, M.D. Purugganan, The evolutionary dynamics of plant duplicate genes, Curr. Opin. Plant Biol. 8 (2005) 122–128.

[12] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes, Science 290 (2000) 1151–1155.

[13] A. Force, et al., Preservation of duplicate genes by complementary, degenerative mutations, Genetics 151 (1999) 1531–1545.

[14] H. Knight, et al., Abscisic acid induces CBF gene transcription and subsequent induction of cold-regulated genes via the CRT promoter element, Plant. Physiol. 135 (2004) 1710–1717.

[15] Y. Narusaka, et al., Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses, Plant J. 34 (2003) 137–148.

[16] K. Higo, Y. Ugawa, M. Iwamoto, T. Korenaga, Plant cis-acting regulatory DNA elements (PLACE) database, Nucleic. Acids Res. 27 (1999) 297–300.

[17] C.D. Bingle, et al., Phylogenetic and evolutionary analysis of the PLUNC gene family, Protein Sci. 13 (2004) 422–430.

[18] H. Liang, W. Zhou, L.F. Landweber, SWAKK: a web server for detecting positive selection in proteins using a sliding window substitution rate analysis, Nucleic. Acids Res. 34 (2006) W382–W384.

[19] A.P. Dias, E.L. Braun, M.D. McMullen, E. Grotewold, Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication, Plant Physiol. 131 (2003) 610–620.

[20] M.B. Howard, et al., Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer degradans*, Protein Sci. 13 (2004) 1422–1425.

[21] L.M. Ferreira, et al., Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase, Biochem. J. 269 (1990) 261–264.

[22] H. Shen, et al., Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity, J. Biol. Chem. 266 (1991) 11335–11340.

[23] M. Kretzschmar, J. Doody, J. Massague, Opposing BMP and EGF signalling pathways converge on the TGF-beta family mediator Smad1, Nature 389 (1997) 618–622.

[24] C.C. Englbrecht, H. Schoof, S. Bohm, Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome, BMC.Genomics 5 (2004) 39.

[25] J. Zheng, et al., Isolation and analysis of water stress induced genes in maize seedlings by subtractive PCR and cDNA macroarray, Plant Mol. Biol. 55 (2004) 807–823.

[26] J. Jia, et al., Annotation and expression profile analysis of 2073 full-length cDNAs from stress-induced maize (*Zea mays L.*) seedlings, Plant J. 48 (2006) 710–727.

[27] J.D. Thompson, et al., The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic. Acids Res. 25 (1997) 4876–4882.

[28] M. Clamp, et al., The Jalview Java alignment editor, Bioinformatics 20 (2004) 426–427.

[29] S. Kumar, K. Tamura, M. Nei, MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment, Brief. Bioinform. 5 (2004) 150–163.

[30] A.I. Saeed, et al., TM4: a free, open-source system for microarray data management and analysis, Biotechniques 34 (2003) 374–378.

[31] R. KaY, et al., Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes, Science 236 (1987) 1299–1302.