Molecular cloning, sequence and expression analysis of *ZmArf2*, a maize ADP-ribosylation factor

Yanyang Liu · Junzhou Li · Yuling Li · Mengguan Wei · Qingxin Cui · Qilei Wang

Received: 19 January 2009/Accepted: 17 June 2009/Published online: 27 June 2009 © Springer Science+Business Media B.V. 2009

Abstract A full-length cDNA encoding a maize GTPbinding protein of the ADP-ribosylation factor family was cloned by suppression subtractive hybridization and an in silico cloning approach. The cDNA was 938 bp in length and contained a complete ORF of 612 bp, which encodes a protein of 203 amino acid residues. Its deduced amino acids sequence had an 83% identity with that of a GTPbinding protein in rice. The gene was designated ZmArf2. The ZmArf2 gene consists of G1, G2, G3, G4 and G5 boxes, and Switch I and Switch II regions. Eight nucleotides differed and five amino acids changed between the popcorn inbred N04 and the dent corn inbred Dan232. One changed amino acid was in the G1 box. RT-PCR analysis showed that ZmArf2 expression increased in the early stages of endosperm development and was not tissuespecific.

Keywords Maize · Endosperm · GTP-binding protein · ADP-ribosylation factor · In silico cloning

Abbreviations

ARF	ADP-ribosylation factor
EST	Expressed sequence tag
GTP	Guanosine triphosphate
ORF	Open reading frame
RT-PCR	Reverse transcription-polymerase
	chain reaction
SSH	Suppression subtractive hybridization

Introduction

The candidate gene approach, which consists of looking for genes segregating around a locus putatively responsible for the variation of a trait, has been proposed as a means of initiating QTL characterization [1, 2]. Over 9 million expressed sequence tags (ESTs) from plant tissues are currently lodged in GenBank (http://www.ncbi.nlm.nih. gov/dbEST/dbEST_summary.html; December 5, 2008). These represent a large set of candidate sequences involved in the physiological processes involved in plant development. The isolation of gene sequences specifically or differentially expressed during one developmental stage provides another source of candidate genes [3, 4]. Expressed sequence tag (EST) databases are valuable resources for discovering novel genes through in silico cloning [5–7].

Guanosine triphosphate (GTP) binding proteins (G proteins) participate in a wide range of biological processes, including signal transduction, protein synthesis and secretion, and cellular proliferation. Ras-homologous GTPases constitute a large family of signal transducers that alternate between an activated, GTP-binding state and an inactivated, GDP-binding state [8–11]. To date, five subfamilies of the Ras superfamily are known: Ras, Rho, Rab, Ran, and ARF1 proteins.

The ADP-ribosylation factor (ARF) belongs to the Ras superfamily of low molecular weight GTP-binding proteins (21–24 kDa) and regulates a diverse range of cellular processes. Like all small GTP-binding proteins in the Ras superfamily, ARF proteins cycle between inactive GDPbound and active GTP-bound forms that bind selectively to effectors [12]. The ARF family plays a major role in membrane trafficking in eukaryotic cells [13]. ARF activation is facilitated by specific guanine nucleotide exchange factors (ARF-GEFs). Several ARF-GEFs have

Y. Liu · J. Li · Y. Li (⊠) · M. Wei · Q. Cui · Q. Wang College of Agriculture, Henan Agricultural University, 95 Wenhua Rd, Zhengzhou, China e-mail: yuling_li@126.com

been identified, varying in size, structure and subcellular distribution [14–18].

In maize, an ARF-like protein has only been identified by Verwoert et al. [19]. Another new candidate EST (PE215C3) putatively encoding GTP-binding proteins has been isolated from the suppression subtractive hybridization (SSH) library and mapped on the location of a QTL for grain weight by in silico mapping in our laboratory [20]. In this study, repeated EST searching, multiple sequence comparisons, and other data-mining techniques were employed to identify the Arf gene from maize endosperm. Certain bioinformatics tools were also used to analyze the genomic organization and the encoded protein of the isolated Arf gene. The ARF cDNA sequence of two maize inbreds with large- and small-sized grains was compared. Its expression characteristics were also analyzed both at different stages of endosperm development and in other tissues. The results of this study provide useful information for future studies on the maize Arf gene.

Materials and methods

Plant materials

A dent maize inbred Dan232 with large-sized grains and a popcorn inbred N04 with small-sized grains were planted at the Scientific Research and Education Center of Henan Agricultural University near Zhengzhou, Henan, China in 2006. Dan232 was derived from Lu 9 kuan × Dan340. N04 was derived from a Chinese popcorn variety BL03. Each plant was self-pollinated by hand. Ears were harvested at 3, 5, 7, 10, 15, 20, 25, 30 and 35 days after pollination (DAP), respectively. To increase the uniformity of the isolated kernels, the upper half and the lower approximately one-sixth of the ears were discarded. Grains were collected from at least six ears and pooled at each time point. Some of the collected samples were immediately frozen in liquid nitrogen and stored at -70° C.

Isolation of total RNA and mRNA

Total RNA was isolated using a hot phenol extraction method [21]. For PCR-select DNA subtraction, mRNA was purified from total RNA using an OligotexTM mRNA Purification Kit (QIAGEN).

In silico cloning of complete open reading frames

Based on the candidate EST putatively encoding GTPbinding proteins from the SSH library established in our laboratory [20], the putative full-length TaCRT cDNA was obtained by in silico cloning. The differentially expressed EST obtained by SSH was selected for BLAST search in the National Center for Biotechnology Information (NCBI) EST database. The overlapping ESTs were assembled into contigs to obtain the open reading frames (ORFs). The specific primers P1 (5'-CATCGAGTCAACCGAACCCAA GC-3'; sense) and P2 (5'- GATAATCCCGGAATGCAG CAAAT-3'; antisense) were designed for amplification of the complete ORF. PCR was carried out using 1 µl of the obtained cDNA, 2.5 μ l 10 × PCR buffer, 2.5 μ l dNTPs mixture (2.5 mM each), 0.1 µl of each primer (10 µM), 0.125 µl Takara La Taq, and distilled H₂O was added to make up the final volume of 25 µl. The PCR conditions were 1 min at 94°C, then 30 cycles of 40 s at 94°C, 40 s at 58°C and 1.5 min at 72°C, and a final extension of 10 min at 72°C. PCR products were separated on 1% agarose gels and the single specific PCR product band was cloned into the pGEM-T easy vector (Promega) for sequencing.

Bioinformatics analysis

Nucleotide sequence and protein similarity analyses were carried out using Danman version 5.2.2 and BLAST programs (http://www.ncbi.nlm.nih.gov/BLAST/), respectively. The ORF Finder (Open Reading Frame Finder) was used to identify the ORF in the nucleotide sequence (http:// www.ncbi.nlm.nih.gov/projects/gorf/) [22]. To establish the genomic organization, the cDNA sequence was blasted to the contigs of the maize genome in GenBank. SIM4 (http://pbil.univ-lyon1.fr/sim4.php) was used to align the cDNA sequence with the genomic sequences to search for potential introns.

Reverse transcription-polymerase chain reaction (RT-PCR) analysis

RNA from maize endosperm and several other organs/tissues (embryo, pericarp, root, stem and leaf) was used for the RT-PCR, which was performed using the P1 and P2 primers. The β -actin gene was amplified as an internal control with the following primers: 5'-CGATTGAGCATGG-CATTGTCA-3' and 5'-CCCACTAGCGTACAACGAA-3'. The samples were resolved on a 1% agarose gel with 1 µg/ml of ethidium bromide and were analyzed with the software Quantity one 4.30 (BioRad, Hercules, CA).

Results

Cloning of ZmArf2 cDNA

An EST highly homologous to GTP-binding protein of rice was obtained from the SSH libraries during identification for differentially expressed genes between 10 DAP and 20 DAP in popcorn inbred N04 endosperm. This 275 bp EST was chosen as a query probe for in silico cloning. BLASTN searches against NCBI maize EST databases revealed that more than 50 EST hits were returned for this EST. The overlapping ESTs were assembled into a 1,259 bp extended sequence. To verify the result of in silico cloning, specific primers were designed for RT-PCR amplification and a 938 bp cDNA fragment was obtained from maize endosperm 10 DAP (Fig. 1). This fragment was fully sequenced and identified as a new maize GTP (ZmArf2) cDNA clone (GenBank accession no. EU816421).

Characterization of ZmArf2 cDNA sequence

The determined nucleotide sequence was 938 bp, which consists of a 237 bp 5'-untranslated region (UTR), a 92 bp 3' UTR, and a 612 bp ORF. The ORF of the ZmArf2 gene encodes 203 deduced amino acid residues with a calculated molecular mass of 22.77 kDa and a predicted pI of 6.13. The BLASTN and BLASTX results showed that it was highly homologous to the GTP-binding protein-like gene in Oryza sativa L. with 80 and 83% identities, respectively.

By using the ZmArf2 cDNA sequence to BLAST search the NCBI High-Throughput Genomic Sequences (HTGS) database of maize (http://www.ncbi.nlm.nih.gov/HTGS/),

a partial maize genomic sequence was identified as being present on BAC clone AC206679.3. Intron-exon boundaries were determined by aligning the cDNA sequence with the partial genomic sequence. Using SIM4, five exons were found in the relevant DNA sequence. The length of each exon was 93, 141, 109, 62, 101 and 106 bp, respectively (Fig. 2a, b).

For further study the ARF cDNA sequence was similarly cloned from the dent corn inbred Dan232. Comparison of the two ARF cDNA sequences from N04 and Dan232 revealed that eight nucleotides differed, which resulted in five amino acid changes (Table 1).

Protein prediction and phylogenetic analysis

Using the online software CDD v2.16 of NCBI for prediction of the gene functional domain (http://www.ncbi. nlm.nih.gov/Structure/cdd/cdd.shtml/), sequence analysis showed that the protein had a high similarity to that of other plant ARFs. Several conserved sequences were revealed. The G1 box, VVDRAGKT, constitutes the phosphatebinding loop [consensus sequence GXXXXGK(S/T)]. The G2 box, T only, is conserved throughout the superfamily, but neighboring residues are conserved within families. The G3 box, DLGG (consensus DXXG), interacts with the 7-phosphate of GTP. The G4 box, NKQD [consensus (N/T)KXD],

Fig. 1 Sequence of the <i>ZmArf2</i>		P1 primer
gene and its deduced amino acid	1	CATCGAGTCAACCGAACCCAAGCCGTGTGGCGACTGGCGAGGCGAGTGCCCCAGTTCCTA
sequence	61	ACTCCGGTGGGCGCGCTCCCACCGCCGCGCGCGCTCAAAACCCGCCCTCAGCCTCCCGCGC
	121	TCCAGTCCACACGGGAGCGGGTGGTGTCGTCTGAAGCGGCGCGATCAAGGAGTCTTCGGG
	181	CGCTCCGGACGTAACGTGCTGGCGGTTGCTTATCTCCGGATGTATATAAAGCGGAAATG
		М
	241	TTCTCCTTGTTCTATGGCCTGTGGAAGTATGTGTTCGCCAAGGACGAGTTCCGTGTTCTG
		F S L F Y G L W K Y V F A K D E F R V L
	301	ATTCTTGTTGTTGACAGAGCTGGCAAGACGACTTTGCTGGAGAAGTTGAAATCGATATAT
		I L V V D R A G K T T L L E K L K S I Y
	361	CTCAAGGGGGAAGGACTTCCGCCTGACCGTGTCGTTCCAACAGTTGGGCTCAACATTGGC
		L K G E G L P P D R V V P T V G L N I G
	421	CGCATCGAAGACGCAAAGGCAAAACTTGTTTTCTGGGATCTAGGTGGTCAGGTTAGCCTA
		R I E D A K A K L V F W D L G G O V S L
	481	CGAACAATCTGGGAGAAATACTATGAAGAGGCCTATGCCATAATGTACGTTATTGACGCT
	541	GCCACAGCATCGTCATTTGAAGATTCCAAAATCTGCTCTGGAGAAAGGTTATTCGCCATGAA
		A T A S S F F D S K S A I. F K V T R H F
	601	
	001	H L P G & P L L T Y & N K O D L P G Y T
	661	
	001	
	721	
	101	O & W S & V D G B G T K S G T D W L W F
	781	
	101	OMEKSKA DTETIONDOGOODOCOODOCOODOCO
	841	
	UTI	
	001	+ "F2 primer
	AOT	ATCGCGTAAGCAGCCATTIGCTGCATTCCGGGATTATC



Fig. 2 a Schematic representation of the structure of the $ZmArf^2$ gene. The numbered exons (*E*) are indicated as *boxes* with the length in nucleotides above the *box*. The introns (*I*) are *thin lines* with length in nucleotides listed below the *line*. **b** Representation of the structure

Table 1 Changes in the Arf cDNA sequence and the correspondingamino acid changes between maize inbred N04 and Dan232endosperm

Exon	Base number	Nucleotide (N04→Dan232)	AA change (N04→Dan232)
1	71	T→G	Val→Gly
2	222	T→C	N/A
3	277	T→C	Tyr→His
3	340	$C \rightarrow T$	N/A
5	419	T→C	Val→Ala
6	507	C→G	N/A
6	565	A→G	Thr→Ala
6	592	C→A	Val→Ile

The base number refers to the cDNA sequences. AA change refers to amino acid changes

A adenosine, C cytidine, T thymidine, G guanosine, N/A unchanged amino acid

was shown to be specific for guaninyl binding. The G5 box, SAY, has the consensus sequence [C/S]A[K/L/T]. The Switch I region, KLKS and PPDRVVPTVGL, and the Switch II region, DLGGQVSLRTIWEKYYEEA, undergo conformational changes upon GTP binding (Fig. 3).

Phylogenetic analysis of a multiple alignment showed that six *Arf* sequences from different plant species were classified into two groups. The *ZmArf2* sequence belonged to the same group as predicted ARFs from rice and *Arabidopsis*, with similarities of 84 and 71%, respectively (Fig. 4a, b). However, the similarity between this *Arf*

for the corresponding cDNA product. The 5'-untranslated region (*UTR*), open reading frame (*ORF*), and 3'-UTR are indicated and the length in nucleotides is indicated below each *box*

sequence and another maize *Arf* sequence was only 35%, which were classified into different groups.

Expression analysis of the ZmArf2 gene

The expression of ZmArf2 at different developmental stages in the endosperm and other tissues for inbred N04 was measured. As shown in Fig. 5a, ZmArf2 was expressed more strongly during the early development of maize endosperm and subsequently expression was reduced slightly. Moreover, expression of the ZmArf2 gene was not tissue-specific but was also expressed in the embryo, pericarp, root, stem and leaf (Fig. 5b). Expression was higher in the stem than in the embryo, pericarp, root and leaf.

Discussion

Using an in silico cloning technique an ADP-ribosylation factor (ARF) gene from the popcorn inbred N04 was cloned in this study. The full length of this cDNA was 938 bp, which included an ORF of 612 nucleotides encoding a polypeptide of 203 amino acids with an estimated molecular mass of about 22.77 kDa. RT-PCR expression showed that the mRNA of ZmArf2 was detectable at all stages during endosperm development and in all other tissues. Therefore, the expression of ZmArf2 was ubiquitous. Previous studies also showed that ARF proteins were ubiquitous and have been found in all eukaryotic



Fig. 3 The protein sequence for the ZmArf2 gene as deduced by CDD in NCBI



Fig. 5 RT-PCR analysis of the ZmArf2 gene expressed at different stages of endosperm development (a) and in other tissues (b) for popcorn inbred N04. The *numbers* 1–7 refer to 3 days after

pollination (DAP), 5, 7, 10, 15, 20 and 25 DAP, respectively. The *numbers* 8–12 refer to the embryo, pericarp, root, stem and leaf, respectively

cells, including humans, bovines, rat, mouse, chicken, plants, yeast and slime mold, and *Drosophila melanogaster* [23–27].

It is known that ARF belongs to the Ras superfamily, which regulate a diverse range of cellular processes. In the present study, expression of ZmArf2 was higher in the earliest four stages than in later stages of endosperm development, and were highest in the first two stages. Since cellularization, cell division and cell proliferation are characteristically involved in the early developmental stages for

maize endosperm [28-31], such results suggested that the cloned *ZmArf2* gene may play a critical role in these phases. Although the function of ARF proteins has not yet been fully elucidated, similar roles have been implicated in previous studies. Ras proteins, related to ARFs, are thought to be involved in basic processes such as cell growth and cell proliferation [8]. ARF proteins are implicated as regulators of vesicle-mediated protein trafficking [32]. A relationship between ARFs and membrane lipids has been demonstrated [33]. ARFs can influence the phospholipid content of

membranes by activating phospholipase D (PLD), yielding phosphatidic acid (PA) and choline. PA has been found to induce DNA synthesis and cell proliferation [34].

From the comparison of the nucleotide and amino acid sequences between the popcorn inbred N04 and the dent corn inbred Dan232, eight nucleotides differed and five amino acids changed between the two inbreds. One amino acid change was from Val to Gly in the G1 box. The Tyr to His and Val to Ala changes were close to the Switch II and G4 box, respectively. The study by Barbacid [35] showed that mutation of Gly to Val or other amino acids resulted in deficient GTPase activity and increased transforming activity. Verwoert et al. [19] speculated that the ARF protein is indirectly involved in transformation of the lipid composition by base exchange. In the present study, the two inbreds are greatly different in endosperm weight and 100-grain weight, with Dan232 31.89-165.94% and 26.46-168.20% higher, respectively, than N04 during grain development [20]. Whether this sequence change could lead to deficient GTPase activity and increased transforming activity, or transformation of the lipid composition, and ultimately lead to the difference in grain weight between the two inbreds should be revealed in a future study. Furthermore, the roles of the other four changes in amino acids should be studied simultaneously.

Acknowledgments This work was funded by the Henan Innovation Project for University Prominent Research Talents (2005HANCET-12), the Henan Natural Science Foundation (0511032900) and Henan Development and Reform Commission.

References

- Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291. doi: 10.1023/A:1011605013259
- Etienne C, Rothan C, Moing A et al (2002) Candidate genes and QTLs for sugar and organic acid content in peach (*Prunus persica* L. Batsch). Theor Appl Genet 105:145–159. doi:10.1007/s00122-001-0841-9
- Wang K, Gan L, Jeffery E et al (1999) Monitoringgeneexpression profile changes in ovarian carcinomas using cDNA microarray. Gene 229:101–108. doi:10.1016/S0378-1119(99)00035-9
- Aharoni A, Keizer LCP, Bouwmeester HJ et al (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12:647–661
- Prigent C, Gill R, Trower M et al (1999) In silico cloning of a new protein kinase, Aik2, related to *Drosophila aurora* using the new tool: EST Blast. In Silico Biol 1:123–128
- Schultz J, Doerks T, Ponting CP et al (2000) More than 1, 000 putative new human signalling proteins revealed by EST data mining. Nat Genet 25:201–204. doi:10.1038/76069
- 7. Rinner O, Morgenstern B (2002) AGenDA: gene prediction by comparative sequence analysis. In Silico Biol 2:195–205
- Hall A (1990) The cellular functions of small GTP-binding proteins. Science 249:635–640. doi:10.1126/science.2116664
- Bourne HR, Sanders DA, McCormick F (1991) The GTPase superfamily: conserved structure and molecular mechanism. Nature 349:117–127. doi:10.1038/349117a0

- 10. Bourne HR, Sanders DA, McCormick F (1990) The GTPase superfamily: a conserved switch for diverse cell functions. Nature
- 348:125–132. doi:10.1038/348125a0
 11. Boguski MS, McCormick F (1993) Proteins regulating Ras and its relatives. Nature 366:643–654. doi:10.1038/366643a0
- Vetter IR, Wittinghofer A (2001) The guanine nucleotide-binding switch in three dimensions. Science 294:1299–1304. doi:10.1126/ science.1062023
- Moss J, Vaughan M (1998) Molecules in the ARF orbit. J Biol Chem 273:21431–21434. doi:10.1074/jbc.273.34.21431
- Chardin P, Paris S, Antonny B et al (1996) A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. Nature 384:481–484. doi:10.1038/384481a0
- Klarlund JK, Guilherme A, Holik JJ et al (1997) Signaling by phosphoinositide -3,4,5-trisphosphate through proteins containing pleckstrin and Sec7 homology domains. Science 275:1927–1930. doi:10.1126/science.275.5308.1927
- Meacci E, Tsai SC, Adamik R et al (1997) Cytohesin-1, a cytosolic guanine nucleotide-exchange protein for ADP-ribosylation factor. Proc Natl Acad Sci USA 94:1745–1748. doi:10.1073/pnas. 94.5.1745
- Morinaga N, Moss J, Vaughan M (1997) Cloning and expression of a cDNA encoding a bovine brain brefeldin A-sensitive guanine nucleotide-exchange protein for ADP-ribosylation factor. Proc Natl Acad Sci USA 94:12926–12931. doi:10.1073/pnas.94.24. 12926
- Peyroche A, Paris S, Jackson CL (1996) Nucleotide exchange on ARF mediated by yeast Gea1 protein. Nature 384:479–481. doi: 10.1038/384479a0
- Verwoert Ira IGS, Brown A, Slabas AR et al (1995) A Zea mays GTP-binding protein of the ARF family complements an Escherichia coli mutant with a temperature-sensitive malonyl-coenzyme A: acyl carrier protein transacylase. Plant Mol Biol 27:629– 633. doi:10.1007/BF00019329
- 20. Liu YY (2008) Identification and cloning of differentially expressed genes in endosperm at two key development stages of different maize inbreds. Doctoral dissertation, Henan agricultural university
- Kay R, Chan A, Daly M et al (1987) Duplication of CaMV 35S promoter sequences creats a strong enhancer for plant genes. Science 236:1299–1302. doi:10.1126/science.236.4806.1299
- Wheeler DL, Church DM, Federhen S et al (2003) Database resources of the National Center for Biotechnology. Nucleic Acids Res 31:28–33. doi:10.1093/nar/gkg033
- Erickson FL, Hannig EM, Krasinskas A et al (1993) Cloning and sequence of ADP-ribosylation Factor 1 (ARF1) from *Schizosaccharomyces pombe*. Yeast 9:923–927. doi:10.1002/yea.320090812
- Memon AR, Clark GB, Thompson GA (1993) Identification of an ARF type low molecular mass GTP-bmding piotein in Pea. Biochem Biophys Res Commun 193:809–813. doi:10.1006/bbrc. 1993.1697
- Murtaugh JJ, Lee FS, Deak P et al (1993) Molecular characterization of a conserved, guanine nucleotide dependent ADPribosylation factor in *Drosophila melanognster*. Biochemistry 32:6011–6018. doi:10.1021/bi00074a012
- Regad F, Bardet C, Tremousaygue D et al (1993) cDNA cloning and expression of an *Arabidopsis* GTP-binding protein of the ARF family. FEBS Lett 316:133–136. doi:10.1016/0014-5793(93) 81201-A
- Tsuchiya M, Price SR, Tsai SC et al (1991) Molecular identification of ADP-ribosylation factor mRNAs and their expression in mammalian cells. J Biol Chem 266:2772–2777
- Lopes MA, Larkins BA (1993) Endosperm origin, development and function. Plant Cell 5:1383–1399
- Berger F (1999) Endosperm development. Curr Opin Plant Biol 2:28–32. doi:10.1016/S1369-5266(99)80006-5

- Olsen OA (2004) Nuclear endosperm development in cereals and Arabidopsis thaliana. Plant Cell 16:S214–S227. doi:10.1105/tpc. 017111
- Randolph LF (1936) Developmental morphology of the caryopsis in maize. J Agric Res 53:881–961
- 32. Serafini T, Orci L, Amherdt M et al (1991) ADP-ribosylation factor is a subunit of the coat of Golgi-derived COP-coated vesicles: a novel role for a GTP-binding protein. Cell 67:239–253. doi:10.1016/0092-8674(91)90176-Y
- Kahn RA, Yucel JK, Malhotra V (1993) ARF signaling: a potential role for phospholipase D in membrane traffic. Cell 75:1045–1048. doi:10.1016/0092-8674(93)90314-G
- 34. Yu CL, Tsai MH, Stacey DW (1988) Cellular ras activity and phospholipid metabolism. Cell 52:63–71. doi:10.1016/0092-8674 (88)90531-4
- Barbacid M (1987) Ras genes. Annu Rev Biochem 56:779–827. doi:10.1146/annurev.bi.56.070187.004023