Biochemical and Biophysical Research Communications 448 (2014) 365–371

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Genomic analysis and expression investigation of caleosin gene family in Arabidopsis

Yue Shen [a,b], Jun Xie [b], Rui-dan Liu [b], Xue-feng Ni [b], Xue-hao Wang [b], Zhi-xi Li [a,*], Meng Zhang [b,*]

[a] College of Food Science and Engineering, Northwest A&F University, Yangling 712100, Shaanxi, People's Republic of China
[b] College of Agronomy, Northwest A&F University, Yangling 712100, Shaanxi, People's Republic of China

ARTICLE INFO

Article history:
Received 16 April 2014
Available online 2 May 2014

Keywords:
Caleosin
Classification
Evolution
Expression
Function

ABSTRACT

Caleosin is a common lipid-droplet surface protein, which has the ability to bind calcium. Arabidopsis (*Arabidopsis thaliana*) is considered a model organism in plant researches. Although there are growing researches about caleosin in the past few years, a systemic analysis of caleosins in Arabidopsis is still scarce. In this study, a comprehensive investigation of caleosins in Arabidopsis was performed by bioinformatics methods. Firstly, eight caleosins in Arabidopsis are divided into two types, L-caleosin and H-caleosin, according to their molecular weights, and these two types of caleosin have many differences in characteristics. Secondly, phylogenetic tree result indicates that L-caleosin may evolve from H-caleosin. Thirdly, duplication pattern analysis shows that segmental and tandem duplication are main reasons for Arabidopsis caleosin expansion with the equal part. Fourthly, the expression profiles of caleosins are also investigated in silico in different organs and under various stresses and hormones. In addition, based on promoter analysis, caleosin may be involved in calcium signal transduction and lipid accumulation. Thus, the classification and expression analysis of caleosin genes in Arabidopsis provide facilities to the research of phylogeny and functions in this gene family.

1. Introduction

Plant oils are very important not only to supply essential nutrition for people, but also to provide energy and material for seed germination and seedling development. Triacylglycerols (TAGs) are the most abundant neutral lipids in the plant oils. Most of TAGs are partitioned in small spherical particles called lipid droplets (LDs, also known as oil bodies, lipid bodies, spherosomes, and oleosomes). LDs contain a hydrophobic core of TAGs, embraced by a monolayer of phospholipids, and unique proteins called oleosins, caleosins and steroleosins [1].

Caleosins have been discovered ubiquitously in plants and fungi [2]. As the structure of oleosin, caleosins consist of three domains: an N-terminal hydrophilic domain including EF-hand calcium-binding motif, a central hydrophobic domain containing proline knot for anchoring LDs, and a C-terminal hydrophilic domain with several phosphorylation sites [3].

Arabidopsis thaliana is considered a model organism in plant researches for its small genome (125 Mb) and short life cycle [4]. Eight caleosin genes have been identified in the Arabidopsis genome [5]. However, only limited analysis of their functions has been reported to date. At4g26740 has been found to be involved in post-germinative activity [6] and At5g55240 was believed to be associated with dormancy [7]. And more recently, At4g26740, At5g55240, At2g33380, At1g70670 are verified to be act as peroxygenases [5,8,9]. The term "caleosin" is given by their ability to bind calcium and similarity to oleosin in structure. Calcium signal is involved in various aspects of plant development. A study showed that caleosin expression was up-regulated in high yield and quality rapeseed [10]. Thus, whether caleosin has other functions, like regulation of oil content, still remains to be investigated. Potential functions of caleosin are proposed in this study.

The completion of *A. thaliana* genome sequence and development of various databases of gene expression provide a great opportunity to identify and analyze genes and predict gene function quickly by bioinformatics tools. *Cis*-acting elements are important switches taking part in the transcriptional regulation of activities of a gene in various biological processes. To the best of our knowledge, little attention has been devoted to the promoter elements analysis of caleosins.

In this work, a systemic analysis of eight caleosin genes in *A. thaliana* was performed. According to molecular weights (Mw), caleosin could be divided into two classes, H-caleosin and L-caleosin,

* Corresponding authors. Fax: +86 29 8708 2845 (M. Zhang).
E-mail addresses: shenyue1020@163.com (Y. Shen), xiejun1029@gmail.com (J. Xie), liurd1991@gmail.com (R.-d. Liu), nixuefeng123456@126.com (X.-f. Ni), xuehaowang1226@gmail.com (X.-h. Wang), lizhxi@nwsuaf.edu.cn (Z.-x. Li), zhangm@nwsuaf.edu.cn (M. Zhang).

http://dx.doi.org/10.1016/j.bbrc.2014.04.115
0006-291X/

for the first time. This classification was supported by the results of motif analysis, multiple alignment and phylogenetic tree. The properties, evolution and expression profiles of caleosins in Arabidopsis were also investigated. Moreover, potential functions of caleosin were speculated based on promoter analysis. Further studies on verifying the functions of caleosin in experiment will be reported in our next study.

2. Materials and methods

2.1. Identification and characteristics of caleosin genes in Arabidopsis

In order to survey whether there were other caleosin genes, exact name and HMM (Hidden Markov Model) searches were performed to search the putative caleosins in *A. thaliana* genome. Exact name searches were executed by using "caleosin" as a query in the Arabidopsis Information Resource (TAIR10.0, http://www.arabi-dopsis.org/) [11]. The raw HMM of caleosin domain (PF05042) was obtained from Pfam 26.0 (http://pfam.sanger.ac.uk/) [12] and as a query to retrieve the protein database in TAIR using "hmmsearch" in local HMMER 3.0 software package [13], with *E*-value $<10^{-10}$. After that, the gene sequences of predicted caleosin were collected and the redundant sequences were removed manually. All the deduced amino acid sequences of the putative caleosins were submitted to the InterProScan 4 (http://www.ebi.ac.uk/Tools/pfa/iprscan/) [14] to verify the presence of caleosin domain (IPR007736).

Physical and chemical properties of caleosins, Mw, isoelectric points (p*I*), and amino acid lengths, were performed on ExPASy (http://web.expasy.org/compute_pi/) [15]. The hydropathic plot was drawed in ProtScale (http://web.expasy.org/protscale/) [16] with Kyte and Doolittle method and the default setting. Ser, Thr and Tyr phosphorylation sites in proteins were retrieved by program NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) [17]. Protein stability changes for single-site mutations were predicted by MUpro (http://www.ics.uci.edu/~baldig/mutation.html) [18].

To investigate whether the caleosin protein owns a conserved domain other than "caleosin", the sequences were analyzed by MEME (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) [19] with 5 different motifs. The annotations of the identified motifs were carried out by InterProScan 4 (http://www.ebi.ac.uk/Tools/InterProScan/) [14].

2.2. Multiple alignment and phylogenetic tree construction

Multiple sequence alignment of the complete protein sequences of eight Arabidopsis caleosins and one cycad (*Cycas revoluta*) caleosin was done by software ClustalX 2.1 [20]. The cycad caleosin sequence was extract from GenBank with accession number FJ455154. Pretty output and shading of the alignment results were operated using online software Boxshade 3.21 in ExPASy (http://www.ch.embnet.org/software/BOX_form.html) [15].

Then, in order to build a neighbor-joining (NJ) tree, the alignment was submitted to MEGA 6.06 [21] using NJ method, Poisson model and pairwise deletion of gaps. The phylogeny was tested by bootstrap replication 1000 and the branch length count on phylogenetic distances.

To construct a maximum-likelihood (ML) tree, the alignment was employed to PhyML 3.0 on website (http://www.atgc-montpellier.fr/phyml/) [22] using LG model, 6 for substitution rate categories and SPR for tree improvement. The reliability of the tree was assessed with bootstrap replication 500.

The visualization of the two phylogenetic trees was both generated using MEGA 6.06.

2.3. Chromosome distribution and gene expansion analysis

Genes were mapped on five Arabidopsis chromosomes by Chromosome map tool on TAIR, using locus names.

The caleosin genes separated by a maximum of five genes were defined as tandem duplication [23]. Searching for segmental duplications were performed according to the study before [24], and also gained by searching locus name against the PGDD (plant genome duplication database, http://chibba.agtec.uga.edu/duplication/) [25].

Ka and Ks values were extracted from PGDD. The Ks values were used to estimate the date of the duplication occurrence following the formula $T = Ks/2\lambda$. For Arabidopsis, λ is 1.5×10^{-8} [26].

2.4. Upstream sequence element analysis

The 2000 bp of eight caleosins genomic sequences upstream of the start condon were obtained from Arabidopsis genome in TAIR. Putative transcription factor binding sites were predicted by using web-based program TFSEARCH 1.3 (http://www.cbrc.jp/research/db/TFSEARCH.html) in plant [27]. Potential *cis*-acting regulatory elements were searched by PLACE 30.0 (http://www.dna.affrc.go.jp/PLACE/) [28].

2.5. Expression patterns of caleosin genes

Expression patterns of caleosin genes were performed by analyzing ESTs (expressed sequence tags), full-length cDNAs, microarray data and MPSS (massively parallel signature sequencing) tags in the public databases.

ESTs and cDNAs were obtained from DFCI database (http://compbio.dfci.harvard.edu/tgi/plant.html) [29]. Caleosin gene sequences were used to search Arabidopsis DFCI database, using BLASTN, with *E*-value $<10^{-10}$, identity >95% and length >200 bp. The most matched one was selected when a gene corresponds to multiple TCs. To analysis gene expression in different tissues, all the libraries were classified into eight synthetic libraries (Table S2). Then, for each gene, the number of corresponding ESTs in each synthetic library was counted and was normalized to TPM (Transcripts per million).

MPSS tags of caleosin genes were acquired from Arabidopsis MPSS Database (http://mpss.udel.edu/at/mpss_index.php) with normalized signature MPSS (TPM) of 17-base signatures.

Microarrays analysis and visualization were done through AtGenExpression Visualization Tool (AVT, http://jsp.weigelworld.org/expviz/expviz.jsp) [30,31], using locus search with mean-normalized value.

The heatmap was generated for displaying expression by R script [32] on website http://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap.html.

3. Result

3.1. Identification, classification and characteristics of caleosins in Arabidopsis

Eight caleosins in Arabidopsis were identified through exact name search in TAIR and HMM search in local, which is consistent with the reports before [5]. Since only names AtCLO1 to AtCLO5 were given in the literatures, additional caleosin genes were named from AtCLO6 to AtCLO8 (Table 1). The size of ORF and deduced amino acid, calculated Mw and p*I* of eight caleosin genes varied in a wide range (Table 1). According to Mw, eight caleosins were classified into two groups, high-Mw and low-Mw caleosin, named as H-caleosin and L-caleosin.

Cycad is a relative lower plant, multiple proteins alignment of eight caleosins in Arabidopsis and one caleosin in cycad were done to investigate structural features. The results showed that the insertion of 29 residues in the N-terminal of H-caleosins was the main difference between L- and H-isoform caleosin. This insertion made N-terminal larger, hydrophobic region in the medium in the H-isoform, which were the classification methods used before [5,33]. It is noteworthy that the cycad caleosin belongs to H-caleosin (Fig. 1).

Alignment revealed that EF-hand domain was quite conserved except At1g23250 (Fig. 1), which might suggest the At1g23250 lost the ability to bind calcium.

Central hydrophobic domain, PX3PSX3P, was essential for anchoring LDs [9]. The second proline (P) of proline knot in L-caleosins was found to be substituted by phenylalanine (F). And in addition, the serine (S) in At1g23250 was found to be replaced by glycine (G) (Fig. 1). To figure out whether these changes affect the structure of caleosins, hydropathic plot was generated and the change of stability for amino acid mutation was predicted. The results indicated that the hydrophobic region of L-caleosins were smaller (Fig. S1) and the stability were decreased.

Possible Ser, Thr and Tyr phosphorylation sites were predicted by NetPhos 2.0 and marked in Fig. 1. S65, S106, T119, S187, S188 (according to At1g70670) were conserved in L-caleosin, while T187, S140, S214, S225 (according to At4g26740) were conserved in H-caleosin. Y145 (according to At4g26740) were conserved in the all examined caleosins (Fig. 1). Except that Y145, S225, T119 were reported before [3], other seven phosphorylation sites were newly identified.

Additional 24 residues were invariable in the tested sequences, suggesting these sites might play roles in keeping functions of caleosin (Fig. 1). Moreover, previous study has point out that C221 (according to At4g26740) was highly conserved [3]. However, our results displayed that C221 was only conserved in H-caleosin without cycad, indicating C221 might be conserved in H-caleosin of angiosperm.

Table 1
Classification and characteristics of caleosin genes in Arabidopsis.

Gene name	TAIR locus ID	Size (ORF, bp)	Size (AA)	pI	Mw
H-caleosin					
AtCLO1	At4g26740	738	245	5.81	28038.04
AtCLO2	At5g55240	732	243	5.62	27875.58
AtCLO3	AT2G33380	711	236	5.17	26600.04
AtCLO8	At5g29560	663	220	8.78	24855.3
L-caleosin					
AtCLO4	AT1G70670	588	195	9.39	22092.12
AtCLO5	AT1G23240	633	210	9.62	23851.25
AtCLO6	AT1G70680	579	192	9.04	21518.56
AtCLO7	AT1G23250	618	205	9.77	23802.67

3.2. Motif analysis of caleosins in Arabidopsis

MEME was used to survey whether there were other motifs. The result revealed that eight caleosin genes all contained motif 1 and motif 2, while L-caleosins missed motif 3, and At5g29560 missed motif 4 (Fig. S2). The annotation of the motifs (Table S1) showed that these motifs were a part of "caleosin".

3.3. Phylogenetic analysis, chromosomal distribution and duplication analysis of caleosin genes

To gain insight into the evolutionary relationships of caleosin genes in Arabidopsis, both NJ method by MEGA 6.06 and ML method by PhyML were used to construct the phylogenetic tree. These two methods led to the similar tree topologies and which were supported by their high bootstrap values (>80%), suggesting that a reliable tree was generated. Following the tree, eight caleosins were divided into two groups and the cycad caleosin was closer to H-caleosin group in Arabidopsis (Fig. 3). It was worth noting that these two groups completely matched the two groups divided by Mw.

To survey the expansion pattern of caleosin gene family, chromosomal distribution of eight caleosins was investigated. As shown in Fig. 2, caleosins were distributed over all other chromosomes expect chromosome 3. Four genes (50%), were present on chromosome 1, whereas two genes was on chromosome 5 and only one gene was on chromosome 2 and 4 respectively.

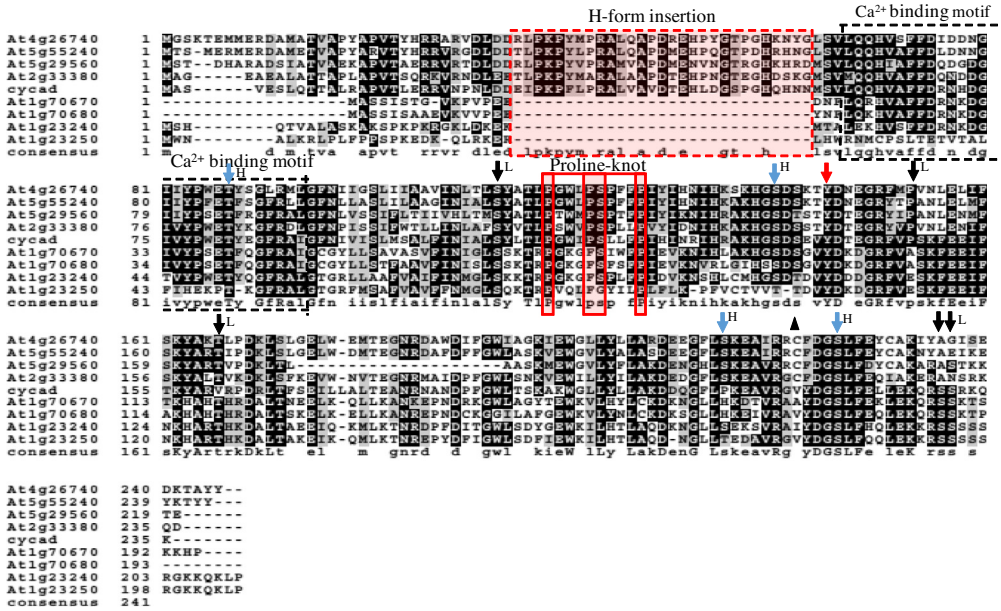

Fig. 1. Multiple alignment of the eight caleosins in Arabidopsis and one caleosin in cycad. H-form insertion is boxed with dotted red line. Calcium binding motif is boxed with dotted black line. Proline-knot domain is boxed with red line. Putative phosphorylation sites of L-caleosin are marked by upper black arrow with "L" and H-caleosin by blue arrow with "H". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

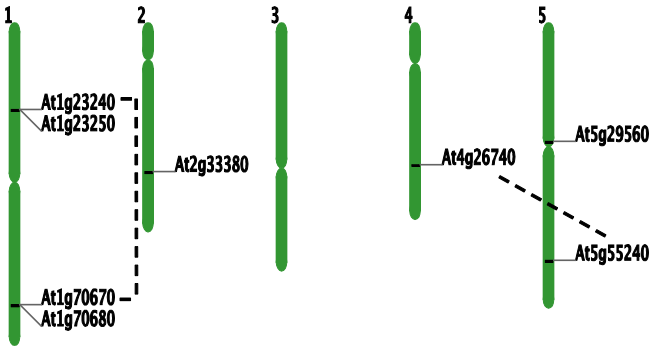

Fig. 2. Chromosomal localization of caleosin genes in Arabidopsis. The number indicated at the top represents the chromosome number. The segmentally duplicated genes are indicated by dotted lines.

Two pairs including four genes were found to be located on chromosome 1 in tandem with high sequence identity at the protein level (Fig. 2, Table 2). Two pairs containing four genes were found to be segmental duplication (Fig. 2, Table 3).

Furthermore, the approximate date of the segmental duplicated caleosins in Arabidopsis was assessed. Based on the Ks value, gene pair, At4g26740 and At5g55240, might undergo a duplication ~40.33 MYA, while another gene pair, At1g70670 and At1g23240, might evolve ~26.33 MYA (Table 3).

3.4. Promoter sequences analysis of caleosin genes

Transcription factor binding sites were retrieved for 2 kb of the caleosin gene 5′-upstream region, using TFSEARCH. As show in Text S1, four binding sites for the transcription factors MYB.Ph3 (Myb-like protein of Petunia hybrida), SBF-1 (silencer-binding factor 1), P (maize activator P of flavonoid biosynthetic genes) and Athb-1 (*A. thaliana* homeo box protein 1) were identified. Among these four binding sites, the former three were common in Arabidopsis caleosin genes, the last one, Athb-1, was only present in At5g55240 (Table 4).

To survey whether other motifs present in the promoter, the potential *cis*-elements were further investigated by PLACE. The motifs GATA, GT-1, phytochrome, RY and RAV1 were found out (Table S3).

3.5. Expression profiles of Arabidopsis caleosin genes in different organs

The expression profiles of caleosin genes in Arabidopsis were surveyed by searching public available resources: ESTs, full-length cDNA, microarrays and MPSS tags. In the three resources, the data from MPSS was limited, only four organs (Table S4). The expression data showed that each caleosin gene had at least one corresponding evidence, indicating that eight caleosin gene in Arabidopsis were all expressed (Table S4). The expression patterns from three

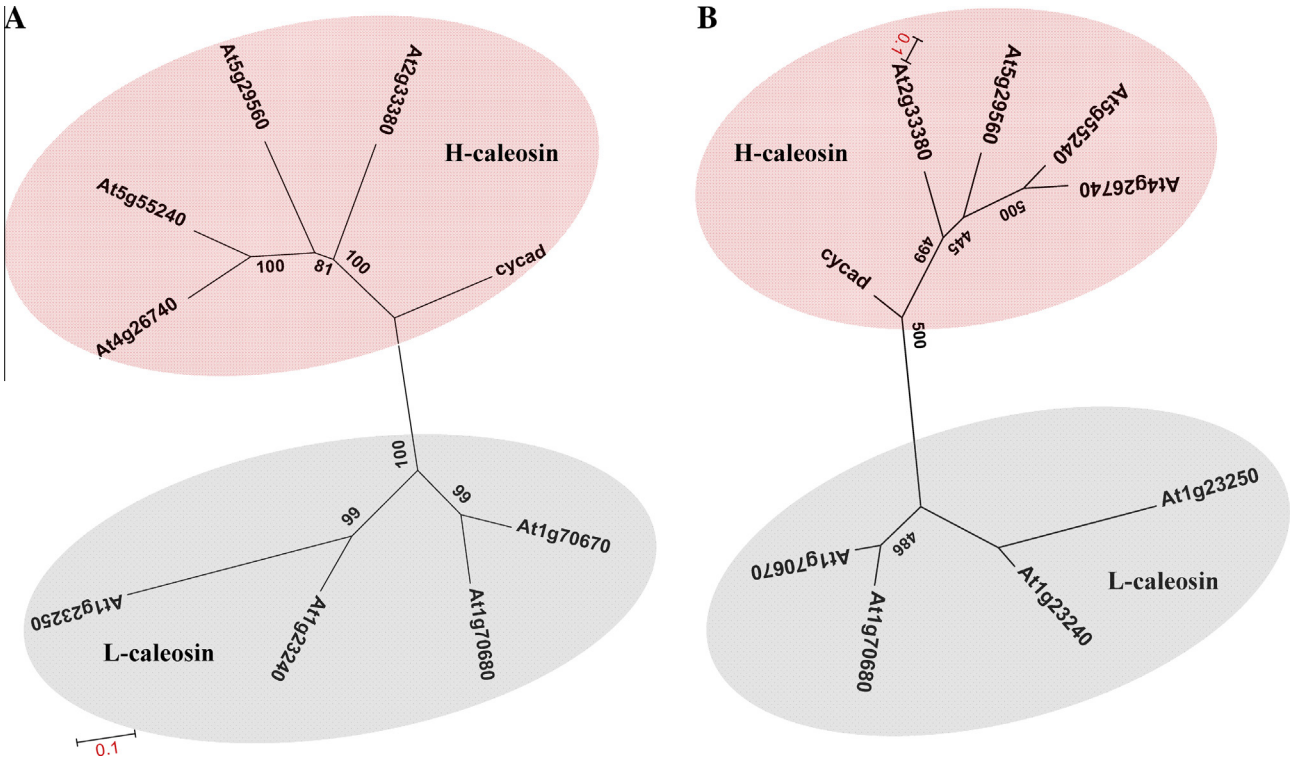

Fig. 3. Phylogenetic trees of eight caleosins in Arabidopsis and one caleosin in cycad by neighbor-joining mehod in MEGA 6.06 (A) and maximum likelihood method in PhyML 3.0 (B).

Table 2
Tandemly duplicated caleosin genes in *A. thaliana*.

Duplicated caleosin 1	Location	Duplicated caleosin 2	Location	Identity (%)
AT1G70670	Chr1: 26644582–26646173	AT1G70680	Chr1: 26647220–26648734	81.15
AT1G23240	Chr1: 8252715–8254625	AT1G23250	Chr1: 8255233–8256770	72.96

Table 3
Segmentally duplicated caleosin genes in *A. thaliana*.

Duplicated caleosin 1	Duplicated caleosin 2	Flanking genes	Ks	Ka/Ks	Age (MYA)	Identity (%)
AT4G26740	AT5G55240	12	0.79	0.15	26.33	79.92
AT1G23240	AT1G70670	10	1.21	0.25	40.33	64.94

Table 4
Transcription factor binding sites predicted by TFSEARCH.

TF-binding site	MYB.Ph3	SBF-1	P	Athb-1
At1g23240	2	2	1	0
At1g23250	0	1	1	0
At1g70670	0	1	2	0
At1g70680	0	4	1	0
At2g33380	1	1	1	0
At5g29560	0	1	1	0
At5g55240	0	1	3	1
At4g26740	0	0	4	0

resources were similar. As showed in Fig. 4A, the expression of At4g26740, At1g70670 and At2g33380 were relatively high in the examined tissues. At4g26740 and At5g55240 display preferential expression in seed, while At1g23240 in bud and At1g70670 in leaf. At1g70670 and At2g33380 were found to be expressed in various organs.

3.6. Expression analysis of Arabidopsis caleosin genes in response to stress conditions

The ESTs analysis indicated that At1g70670 and At2g33380 were expressed in senescence leaf and under stress conditions. And the MPSS tags showed that only At2g33380 expressed subjected to the SA after 4 h and 52 h with moderate abundance and low abundance, respectively (Table S4).

The expression patterns of caleosin genes under different hormones and stresses were also analyzed using microarray data. All Arabidopsis caleosin genes have microarray data except At5g29560 and At1g70680. However, the microarray data of other tissue, like seed, under various stresses and hormones were still scarce. The results indicated that At2g33380 was super-sensitive to ABA (abscisic acid), and MJ (methyl jamonate), sensitive to GA (gibberellin) in seedling, whereas At5g55240 and At4g26740 were sensitive to ABA in seed (Fig. 4B, Table S4). And At2g33380 was induced in aerial part with cold, osmotic, salt, drought, UV-B, wounding, and in root with osmotic, salt, while At1g70670 was induced in root with heat and sensitive to heat in cell culture (Fig. 4C, Table S4).

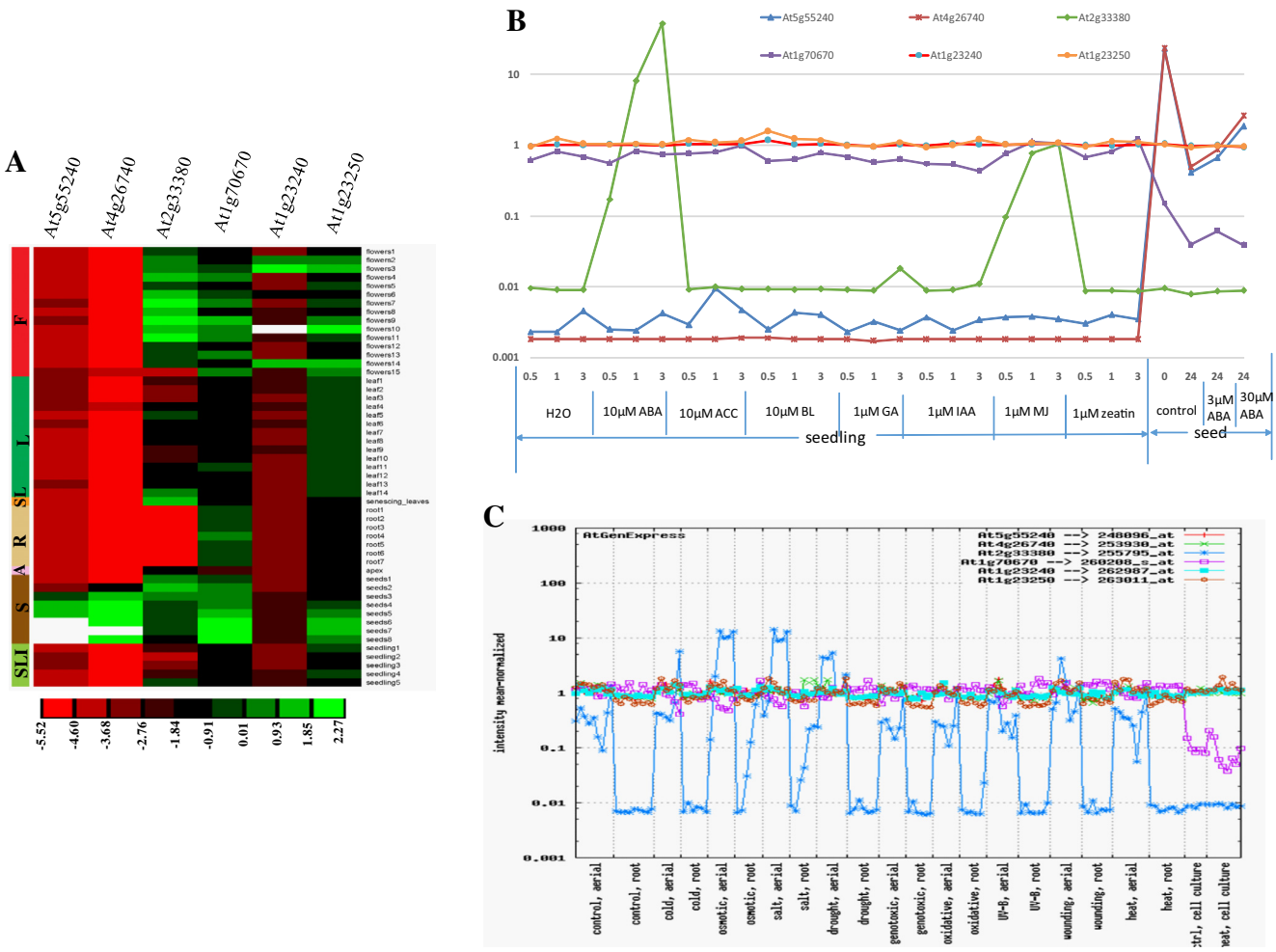

Fig. 4. Expression profiles of Arabidopsis caleosin genes using microarray data. (A) Expression patterns in different development tissues. The heat map were generated by average log signal values of caleosin genes (F, flower; L, leaf; SL, senescing leaf; R, root; A, apex; S, seed; SLI, seedling). The color scale representing average log signal values is shown at the bottom. (B) Expression patterns under various hormones (ABA, abscisic acid; ACC, 1-aminocyclopropane-1-carboxylic acid; BL, brassinolide; GA, gibberellin; IAA, indol-yl-3-acetic acid; MJ, methyl jamonate). (C) Expression profiles under various stresses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Taken together, At2g33380 was sensitive to various stresses and hormones, At1g70670 was sensitive to some stresses in root and cell culture, At5g55240 and At4g26740 were sensitive to some hormones in seed.

4. Discussion

4.1. Two types of caleosins in Arabidopsis

In this study, eight caleosins are sorted into two classes, H-caleosin and L-caleosin, according to the Mw, as the oleosins are classified [34]. Motif analysis, protein alignment and phylogenetic analysis all support this classification. Protein alignment shows that the insertion of 29 residues in the N-terminal domain of H-caleosins is the main difference between the two types of caleosins.

A comparison between H- and L-caleosin shows that the p*Is* of H-caleosins are below 6 except At5g29560, and L-caleosins are above 9 (Table 1), suggesting that two types of caleosins might operate in different microenvironments. Meanwhile, L-caleosin has smaller hydrophobic region (Fig. S1) and decreased stability since proline knot changed. Further, two types of caleosins have different conserved Ser, Thr phosphorylation sites (Fig. 1).

Two oleosin isoforms have different function [34]. Whether two types of caleosin have diverse functions remains to be determined.

4.2. Evolution of the caleosin genes in Arabidopsis

The duplicated pair, At1g23240 and At1g70670, is speculated to be occurred 40.33 MYA, which is consistent with the emergence of crucifers (24–40 MYA) [35]. The other pair, At4g26740 and At5g55240, may evolve 26.33 MYA, just before the split between Arabidopsis and genus Brassica (12–20 MYA) [36].

Cycad is a relatively primitive gymnosperms species. Classification result shows that cycad caleosin belonged to H-caleosin (Fig. 1) and phylogenetic analysis indicates that cycad has closer relationship to the H-caleosin (Fig. 3). Taken together, H-caleosin is presumably more primitive than L-caleosin.

4.3. Expansion pattern of caleosins in Arabidopsis

Segmental duplication, tandem duplication, and transposition play important roles in the gene family expansion. In Arabidopsis, At1g23240 and At1g70670 (25%) undergo both segmental duplication and tandem duplication. At4g26740 and At5g55240 (25%) undergo segmental duplication, and At1g23250 and At1g70680 (25%) undergo tandem duplication. Thus, segmental and tandem duplication (75%) are main reasons for Arabidopsis caleosin expansion with the equal part.

At1g23250, tandem gene of At1g23240, has mutations in the calcium-binding domain of N-terminal, indicating the function lose for calcium-binding. So, it is likely that At1g23250 undergo non-functionalization. At4g26740 and At5g55240 are segmental duplication pair, they both specially express in seeds, however, the expression of At4g26740 was much higher than At5g55240, suggesting sub-functionalization has happened.

4.4. The potential roles of caleosins in signal transduction and lipid accumulation

Three common putative transcription factor-binding sites were identified in the promoter of caleosin genes (Table 4). Two of them, MYB.Ph3 and P, participate in flavonoid biosynthesis [37,38], which are available to respond to various abiotic stresses in plant [39].

The third one, SBF-1, which is nearly identical to GT-1 [40], is discovered in all Arabidopsis caleosins except At4g26740. GT-1 and as-1 or GATA motif work together to respond to the calcium or calmodulin signal [41].

PLACE analysis shows GATA and GT-1 are existed in all caleosins in Arabidopsis. Interestingly, phytochrome regulation element, the target of Calcium [42], are also found out. Moreover, caleosins have conserved calcium-binding domain and phosphorylation sites. Therefore, it is speculated that caleosin may response to calcium signal, or, may transduce calcium signal to the extracellular space.

PLACE find motif RY and RAV1, too. RY can bind B3 domain which is in the transcript factors of seed development: ABI3, FUS3 and LEC2 [43]. RAV1 transcription factor contains AP2 and B3-like domain, which is related with storage oil synthesis in castor bean [44]. In addition, a study shows that caleosin is up-regulated in high yield and quality rapeseed [10]. Thus, caleosin may be involved in the process of oil-accumulation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31270295) and the starting funding of Northwest A&F University, China.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.04.115.

References

[1] K.D. Chapman, J.M. Dyer, R.T. Mullen, Biogenesis and functions of lipid droplets in plants thematic review series: lipid droplet synthesis and metabolism: from yeast to man, J. Lipid Res. 53 (2012) 215–226.
[2] I. Hernandez-Pinzon, K. Patel, D.J. Murphy, The *Brassica napus* calcium-binding protein, caleosin, has distinct endoplasmic reticulum- and lipid body-associated isoforms, Plant Physiol. Biochem. 39 (2001) 615–622.
[3] J.C.F. Chen, C.C.Y. Tsai, J.T. Tzen, Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds, Plant Cell Physiol. 40 (1999) 1079–1086.
[4] D.W. Meinke, J.M. Cherry, C. Dean, et al., *Arabidopsis thaliana*: a model plant for genome analysis, Science 282 (1998) 662–682.
[5] A. Hanano, M. Burcklen, M. Flenet, et al., Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif, J. Biol. Chem. 281 (2006) 33140–33151.
[6] M. Poxleitner, S.W. Rogers, A. Lacey Samuels, et al., A role for caleosin in degradation of oil-body storage lipid during seed germination, Plant J. 47 (2006) 917–933.
[7] P.E. Toorop, R.M. Barroco, G. Engler, et al., Differentially expressed genes associated with dormancy or germination of *Arabidopsis thaliana* seeds, Planta 221 (2005) 637–647.
[8] M. Partridge, D.J. Murphy, Roles of a membrane-bound caleosin and putative peroxygenase in biotic and abiotic stress responses in Arabidopsis, Plant Physiol. Biochem. 47 (2009) 796–806.
[9] Y.Y. Kim, K.W. Jung, K.S. Yoo, et al., A stress-responsive caleosin-like protein, AtCLO4, acts as a negative regulator of ABA responses in Arabidopsis, Plant Cell Physiol. 52 (2011) 874–884.
[10] Y. Hu, G. Wu, Y. Cao, et al., Breeding response of transcript profiling in developing seeds of *Brassica napus*, BMC Mol. Biol. 10 (2009) 49.
[11] E. Huala, A.W. Dickerman, M. Garcia-Hernandez, et al., The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant, Nucleic Acids Res. 29 (2001) 102–105.
[12] M. Punta, P.C. Coggill, R.Y. Eberhardt, et al., The Pfam protein families database, Nucleic Acids Res. 32 (2012) D290–D301.
[13] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, Nucleic Acids Res. 39 (2011) W29–W37.
[14] S. Hunter, P. Jones, A. Mitchell, et al., InterPro in 2011: new developments in the family and domain prediction database, Nucleic Acids Res. 40 (2011) D306–D312.
[15] E. Gasteiger, A. Gattiker, C. Hoogland, et al., ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31 (2003) 3784–3788.
[16] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.

[17] N. Blom, S. Gammeltoft, S. Brunak, Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362.
[18] J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins 62 (2006) 1125–1132.
[19] T.L. Bailey, N. Williams, C. Misleh, et al., MEME: discovering and analyzing DNA and protein sequence motifs, Nucleic Acids Res. 34 (2006) W369–W373.
[20] M.A. Larkin, G. Blackshields, N.P. Brown, et al., Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.
[21] K. Tamura, G. Stecher, D. Peterson, et al., MEGA6: molecular evolutionary genetics analysis version 6, Mol. Biol. Evol. 30 (2013) 2725–2729.
[22] S. Guindon, J.F. Dufayard, V. Lefort, et al., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321.
[23] H. Ma, J. Zhao, Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.), J. Exp. Bot. 61 (2010) 2647–2668.
[24] C. Maher, L. Stein, D. Ware, Evolution of Arabidopsis microRNA families through duplication events, Genome Res. 16 (2006) 510–519.
[25] T.H. Lee, H. Tang, X. Wang, et al., PGDD: a database of gene and genome duplication in plants, Nucleic Acids Res. 41 (2013) D1152–D1158.
[26] G. Blanc, K.H. Wolfe, Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes, Plant Cell 16 (2004) 1667–1678.
[27] Y. Akiyama, TFSEARCH: Searching Transcription Factor Binding Sites, Real World Computing Partnership, Japan, 1995.
[28] K. Higo, Y. Ugawa, M. Iwamoto, et al., Plant *cis*-acting regulatory DNA elements (PLACE) database, Nucleic Acids Res. 27 (1999) 297–300.
[29] J. Quackenbush, J. Cho, D. Lee, et al., The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species, Nucleic Acids Res. 29 (2001) 159–164.
[30] H. Goda, E. Sasaki, K. Akiyama, et al., The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access, Plant J. 55 (2008) 526–542.
[31] J. Kilian, D. Whitehead, J. Horak, et al., The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses, Plant J. 50 (2007) 347–363.
[32] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2005. www.R-project.org.
[33] H. Næsted, G.I. Frandsen, G.Y. Jauh, et al., Caleosins: Ca2+-binding proteins associated with lipid bodies, Plant Mol. Biol. 44 (2000) 463–476.
[34] Y.Y. Wu, Y.R. Chou, C.S. Wang, et al., Different effects on triacylglycerol packaging to oil bodies in transgenic rice seeds by specifically eliminating one of their two oleosin isoforms, Plant Physiol. Biochem. 48 (2010) 81–89.
[35] G. Blanc, K. Hokamp, K.H. Wolfe, A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome, Genome Res. 13 (2003) 137–144.
[36] M.A. Koch, B. Haubold, T. Mitchell-Olds, Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae), Mol. Biol. Evol. 17 (2000) 1483–1498.
[37] R. Solano, C. Nieto, J. Paz-Ares, MYB.Ph3 transcription factor from Petunia hybrida induces similar DNA-bending/distortions on its two types of binding site, Plant J. 8 (1995) 673–682.
[38] E. Grotewold, B.J. Drummond, B. Bowen, et al., The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset, Cell 76 (1994) 543–553.
[39] T. Iwashina, Flavonoid function and activity to plants and other organisms, Uchu Seibutsu Kagaku 17 (2003) 24–44.
[40] M.A. Lawton, S.M. Dean, M. Dron, et al., Silencer region of a chalcone synthase promoter contains multiple binding sites for a factor, SBF-1, closely related to GT-1, Plant Mol. Biol. 16 (1991) 235–249.
[41] D.X. Zhou, Regulatory mechanism of plant gene transcription by GT-elements and GT-factors, Trends Plant Sci. 4 (1999) 210–214.
[42] Y. Wu, K. Hiratsuka, G. Neuhaus, et al., Calcium and cGMP target distinct phytochrome-responsive elements, Plant J. 10 (1996) 1149–1154.
[43] G. Monke, M. Seifert, J. Keilwagen, et al., Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon, Nucleic Acids Res. 40 (2012) 8240–8254.
[44] D. Tajima, A. Kaneko, M. Sakamoto, et al., Wrinkled 1 (WRI1) homologs, AP2-type transcription factors involving master regulation of seed storage oil synthesis in castor bean (*Ricinus communis* L.), Am. J. Plant Sci. 4 (2013) 333–339.