Identification, duplication, evolution and expression analyses of caleosins in Brassica plants and Arabidopsis subspecies

Shen, Yue; Liu, Mingzhe; Wang, Lili; Li, Zhuowei; Taylor, David C.; Li, Zhixi; Zhang, Meng

Publisher's version / Version de l'éditeur:
https://doi.org/10.1007/s00438-015-1156-x
Molecular Genetics and Genomics, pp. 1-18, 2016-01-19

Mol Genet Genomics
DOI 10.1007/s00438-015-1156-x

ORIGINAL ARTICLE

Identification, duplication, evolution and expression analyses of caleosins in Brassica plants and Arabidopsis subspecies

Yue Shen[1,2] · Mingzhe Liu[2] · Lili Wang[2] · Zhuowei Li[2] · David C. Taylor[3] · Zhixi Li[1] · Meng Zhang[2]

Received: 16 June 2015 / Accepted: 8 December 2015

Abstract Caleosins are a class of Ca^{2+} binding proteins that appear to be ubiquitous in plants. Some of the main proteins embedded in the lipid monolayer of lipid droplets, caleosins, play critical roles in the degradation of storage lipids during germination and in lipid trafficking. Some of them have been shown to have histidine-dependent peroxygenase activity, which is believed to participate in stress responses in Arabidopsis. In the model plant *Arabidopsis thaliana*, caleosins have been examined extensively. However, little is known on a genome-wide scale about these proteins in other members of the Brassicaceae. In this study, 51 caleosins in Brassica plants and *Arabidopsis lyrata* were investigated and analyzed in silico. Among them, 31 caleosins, including 7 in *A. lyrata*, 11 in *Brassica oleracea* and 13 in *Brassica napus*, are herein identified for the first time. Segmental duplication was the main form of gene expansion. Alignment, motif and phylogenetic analyses showed that Brassica caleosins belong to either the H-family or the L-family with different motif structures and physicochemical properties. Our findings strongly suggest that L-caleosins are evolved from H-caleosins. Predicted phosphorylation sites were differentially conserved in H-caleosin and L-caleosins, respectively. 'RY-repeat' elements and phytohormone-related *cis*-elements were identified in different caleosins, which suggest diverse physiological functions. Gene structure analysis indicated that most caleosins (38 out of 44) contained six exons and five introns and their intron phases were highly conserved. Structurally integrated caleosins, such as BrCLO3-3 and BrCLO4-2, showed high expression levels and may have important roles. Some caleosins, such as BrCLO2 and BoCLO8-2, lost motifs of the calcium binding domain, proline knot, potential phosphorylation sites and haem-binding sites. Combined with their low expression, it is suggested that these caleosins may have lost function.

Communicated by S. Hohmann.

Y. Shen and M. Liu have contributed equally to this publication.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-1156-x) contains supplementary material, which is available to authorized users.

✉ Zhixi Li
lizhxi@nwsuaf.edu.cn

✉ Meng Zhang
zhangm@nwsuaf.edu.cn

Yue Shen
shenyue1020@163.com

Mingzhe Liu
lmz920621@163.com

Lili Wang
wanglily1005@gmail.com

Zhuowei Li
lizhuowei1014@126.com

David C. Taylor
David.Taylor@nrc-cnrc.gc.ca

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

[2] College of Agronomy, Northwest A&F University, Yangling 712100, Shaanxi, People's Republic of China

[3] National Research Council of Canada, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada

Published online: 19 January 2016

Keywords Caleosin · Brassica · Characteristics · Evolution · Expression · Function

Introduction

In plants, neutral lipids provide energy and substrates/skeletons for seedling establishment and development. Triacylglycerols (TAGs), the main component of neutral storage lipids, are packaged in organelles called lipid droplets. The lipid droplets are covered by a monolayer of phospholipid molecules in which proteins are embedded. These proteins can be divided into at least three families: oleosins, caleosins and steroleosins (Chen et al. 1999; Chapman et al. 2012; Frandsen et al. 2001; Li-Beisson et al. 2013). Caleosins were so-named due to their ability to bind calcium and their oleosin-like structures; caleosins are considered as structural stabilizers of lipid droplets. In addition, reports have shown that caleosins are involved in other biological processes such as dormancy. Some caleosins have been identified as calcium-binding haem-oxygenases with peroxygenase activity induced in stress responses (Hanano et al. 2006; Kim et al. 2011; Murphy 2009; Partridge and Murphy 2009; Poxleitner et al. 2006; Toorop et al. 2005). The model plant *Arabidopsis thaliana* (*A. thaliana*) and many high value Brassica crops belong to the Brassicaceae. These include major oil crops like canola, the world's third largest source of plant oil (Lomascolo et al. 2012), mainly defined by two species: *Brassica napus* (*B. napus*) and *Brassica rapa* (*B. rapa*) as well as vegetable crops like broccoli, cauliflower, Brussels sprouts, cabbage and kale, found in the *Brassica oleracea* (*B. oleracea*) family. The allotetraploid *B. napus* is the product of natural hybridization of the diploids *B. rapa* and *B. oleracea* about 7500 years ago (Chalhoub et al. 2014; Nagaharu 1935). The segregation of Arabidopsis and Brassica plants occurred in 12–20 MYA (Blanc et al. 2003; Town et al. 2006; Yang et al. 1999). *A. thaliana* and its related subspecies *Arabidopsis lyrata* (*A. lyrata*) diverged 10 MYA (Hu et al. 2011). Brassica plants experienced a specific whole-genome triplication process 5–15 MYA (Beilstein et al. 2010; Wang et al. 2011). The separation of *B. rapa* and *B. oleracea* occurred 4.6 MYA (Liu et al. 2014). Recent information provides an opportunity to analyze the evolution of the caleosin gene family between *B. napus* and its progenitors.

The hairpin-like caleosins which are present in higher plants (Chen et al. 1999), fungi (Murphy 2009), and some of single-cell alga (Næsted et al. 2000; Murphy 2009), have three characteristic motifs: The hydrophilic N-terminal domain contains an EF-hand calcium-binding domain; the hydrophobic domain in the centre of the protein contains a proline knot; phosphorylation sites are located in the hydrophilic C-terminal domain (Kim et al. 2011).

Recently, detailed studies of caleosins in the model plant *A. thaliana* have been reported (Chapman et al. 2012; Murphy 2012). Some caleosins in *B. rapa* and *B. napus* have also been identified. A 27-kDa-endoplasmic reticulum (ER)-localized caleosin and a 25-kDa-lipid droplet-localized caleosin have been identified from the embryos of *B. napus* (Hernandez-Pinzon et al. 2001) and seed-specific caleosins have been identified in high and low erucic acid cultivars of *B. napus* (Katavic et al. 2006). According to the sequences of AtCLO1–AtCLO3, nine caleosins had been identified from the ESTs library of *B. napus*, seven of which have been verified by means of proteomics (Jolivet et al. 2009). Hu et al. (2013) have identified 11 caleosins from *B. rapa* using bioinformatics, and performed a phylogenetic analysis. AtCLO3, maybe as a peroxygenase, involves in producing hydroxylated fatty acid derivatives which could have anti-fungal properties (Partridge and Murphy 2009). Previous studies also indicated that a caleosin, designated RD20, participates in stress tolerance in Arabidopsis (Aubert et al. 2010; Blée et al. 2014), while PXG4, a class II caleosin in yeast, is a calcium-dependent membrane-associated hemoprotein catalyzing typical peroxygenase reactions (Blée et al. 2012).

Although some caleosins have been examined extensively in Arabidopsis, little is known on a genome-wide scale regarding these caleosins present in other plants of the Brassicaceae. The current investigation will provide additional information to understand their function and roles in plant development. In addition, many questions still remain: For example, are there any other caleosins in *B. rapa* and *B. napus*? How many caleosins are present in the *A. lyrata* and *B. oleracea* genomes? What can we determine regarding caleosin secondary structures, the evolution process, expression profiles and putative functions in these Brassicaceae?

A wealth of genomic information is available for *A. thaliana* and its relative species *A. lyrata* (Hu et al. 2011; Arabidopsis genome initiative 2000). The sequencing of many plant genomes including *B. napus*, *B. rapa* and *B. oleracea* have been completed in recent years (Chalhoub et al. 2014; Liu et al. 2014; Wang et al. 2011). Collectively, these databases provide a rich resource for the study of caleosins in these plants.

In this study, caleosins are identified from the genomes of *A. lyrata* (7), *B. rapa* (11), *B. oleracea* (11) and *B. napus* (22), using sequences of eight known caleosins from *A. thaliana* as bioinformatic queries. Physical and chemical properties, the evolution process and expression profiles of these caleosins are also analyzed. Functional predictions of four key caleosin genes, BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4 from *B. napus* are reported.

Materials and methods

Identification and naming of the caleosins in *A. lyrata* and three Brassica crop species

In this research, both BLAST and key word searches were carried out in order to determine the number of caleosin genes in *A. lyrata*, *B. napus*, *B. rapa* and *B. oleracea*. Using sequences of eight known caleosin genes from *A. thaliana* asqueries, a BLASTP analysis was carried out to search the target databases for sequences with an *E* value $<10^{-10}$. The databases for *B. rapa* BRAD; (http://brassicadb.org/brad/) (Cheng et al. 2011), *B. oleracea* Bol base v1.0 (http://119.97.203.210/bolbase/index.html) (Yu et al. 2013) and for *A. lyrata* Phytozome 10 (http://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al. 2012) were probed in real time against the eight query sequences from *A. thaliana* obtained from TAIR10.0 (http://www.arabidopsis.org/) (Lamesch et al. 2012). Key words in these searches were "PF05042" or "caleosin" or "IPR007736". After obtaining the sequence matches, redundant information was removed manually. Then reciprocal BLASTX searches back to *A. thaliana* helped to verify the putative match sequences and those not corresponding to any of the eight caleosin genes in *A. thaliana* were discarded. The online tool InterProScan5 (http://www.ebi.ac.uk/interpro/search/sequence-search) (Jones et al. 2014) on EBI was used to analyze whether there were domains belonging to the caleosin gene family (IPR007736) in the translated proteins from the selected sequences; those without any caleosin-like domains were also omitted. This culling process at both the caleosin gene and protein levels resulted in the final list of *B. rapa*, *B. oleracea* and *A. lyrata* sequences addressed in this study.

In addition, caleosin sequences from *B. napus* were identified. Since a BLAST search could not be used to probe the *B. napus* database directly, putative nucleotide sequences were obtained from EMBL-EBI (European nucleotide archive, http://www.ebi.ac.uk/ena/browse) while the corresponding protein sequences were downloaded from the *B. napus* genome browser (BNGB, http://www.genoscope.cns.fr/brassicanapus/data/) (Chalhoub et al. 2014), and the names of the genes were entered manually. Local BLASTN and local BLASTP were used to search for caleosin sequences in a local database of *B. napus* with an *E* value $<10^{-10}$. Redundant data were deleted and reciprocal database challenges were performed as above.

The selected caleosins were named after their homologs in *A. thaliana*. For instance, the CLO1 homolog gene of *A. thaliana* was named as BrCLO1 in *B. rapa*. If there were multiple homologous genes identified, they were given the names BrCLO1-1, BrCLO1-2 etc.

There were both orthologs and paralogs of Arabidopsis caleosin genes among the sets of genes we identified. Orthologs of *B. rapa* caleosin genes from *A. thaliana* are reported in the *B. oleracea* genome (Liu et al. 2014) and those from *B. napus* are reported in the *B. napus* genome (Chalhoub et al. 2014). Gene X from the genome of *B. napus* is the ortholog of gene Y from the genome of *B. rapa,* while the latter is the ortholog of gene Z in the genome of Arabidopsis. By inference therefore, gene X from *B. napus* was regarded as the ortholog of gene Z in Arabidopsis. If there were no *B. napus* ortholog in *B. rapa* or if orthologs were not found in Arabidopsis, CLUSTALX was used to identify the gene paralogs in Arabidopsis. Because of the diversities between the caleosin gene names of *B. oleracea* in the *B. napus* genome, genes of *B. oleracea* as listed in the *B. napus* genome did not necessarily correspond to those in the *B. oleracea* genome. Thus, CLUSTALX was used to judge which *B. napus* genes are the Arabidopsis orthologs/paralogs when there were only orthologs in *B. oleracea* in the absence of *B. rapa* orthologs.

Physicochemical property analysis of caleosins

Isoelectricpoints (pIs) and molecular weights of the putative caleosin proteins were calculated with ExPASy (http://web.expasy.org/compute_pi/) (Gasteiger et al. 2003) and the putative phosphorylation sites (including Tyr, Ser and Thr residues) were predicted with NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) (Blom et al. 1999). MEME (http://meme-suite.org/) (Bailey et al. 2006) was used to analyze the ten motifs of these caleosin sequences. InterProScan5 and multiple alignment were used to elucidate motifs. The hydropathic plot was drawed in ProtScale (http://web.expasy.org/protscale/) with Kyte and Doolittle mehod (1982) and the default setting. To test the possible membrane associations/locations of the caleosins, the hydrophobicity and the folding tendencies of these proteins were predicted by online tool FoldIndex (Prilusky et al. 2005).

Multiple alignment, classification and phylogenetic analysis of caleosins

In order to discover if there were significant signature domains in these 51 caleosin proteins, Clustal 2.1 (Larkin et al. 2007) was used to carry out multiple alignments. Shadow processing and its output were conducted with software Boxshade 3.21 (http://www.ch.embnet.org/software/-B-OX_form.html) from the ExPASy database (Gasteiger et al. 2003).

To acquire more information regarding evolutionary relationships between the various caleosins, eight caleosin

sequences in Arabidopsis were also added to populate the phylogenetic tree. Both Neighbor Joining (NJ) and Maximum Likelihood (ML) trees were built to guarantee the reliability of the trees. MEGA 6.06 (Tamura et al. 2013) was used to build the NJ tree; a Poisson model was chosen, and interspaces were deleted pair-wise, and bootstrapped 1000 times. PhyML 3.0 (http://www.atgcmontpellier.fr/phyml/) (Guindon et al. 2010) was used to build the ML tree; the LG model was chosen, with a substitution rate of 6, alter mode SPR, and bootstrapped 100 times. The output of the assembled trees was obtained using MEGA 6.06 (Tamura et al. 2013).

Chromosomal location and duplication model of caleosin genes

The chromosomal locations and flanking genes of the caleosins from *A. lyrata*, *B. rapa* and *B. oleracea* were obtained from their latest databases cited previously. Locations of *B. napus* genes were obtained from the BNGB website (Chalhoub et al. 2014). The figure of gene locations was drawn using MapChart 2.2 (Voorrips 2002).

Two genes would be regarded as tandem duplications if there were less than five genes between these two genes (Ma and Zhao 2010). The definition of segmental duplication was as reported in a previous study (Maher et al. 2006).

The Codeml of PAML module from PAL2NAL (http://www.bork.embl.de/pal2nal/) (Goldman and Yang 1994; Suyama et al. 2006) was used as tool to calculate the caleosin Ka and Ks values. The formula $T = \mathrm{Ks}/2\lambda$ ($\lambda = 1.5 \times 10^{-8}$ for Brassicaceae) (Blanc and Wolfe 2004) was used to calculate the probable age of segmental duplication.

Analysis of upstream *cis*-elements of *B. napus* caleosin genes

Two-thousand-bp upstream sequences before the start codon were downloaded from BNGB (Chalhoub et al. 2014). These sequences were uploaded to PLACE 30.0 (http://www.dna.affrc.go.jp/PLACE/) (Higo et al. 1999) and plantCARE (Lescot et al. 2002) to make upstream *cis*-element predictions.

Expression pattern analyses

Expression pattern analyses of the caleosins from the three Brassica crop families (*B. rapa*, *B. oleracea* and *B. napus*) were determined using public ESTs databases. RNA-seq analysis was also carried out for *B. rapa* and *B. oleracea* caleosins; because of the deficiency of the data in the *B. napus* database, we were unable to obtain RNA-seq data.

ESTs data were from NCBI. The caleosin genes from the three Brassica species were used as queries to make a BLASTN alignment; the parameters were: identity >95 %, *E* value $<10^{-10}$, length $\geq$200 bp. Only ESTs of highest identity value and most similar in length were retained to prevent repetition. To discover the expression profile of those genes in different tissues, 10 and 12 libraries were summarized for *B. rapa* and *B. napus* (see Supplemental Tables 2, 3, 4 and 5). Expression abundance for each gene was calibrated on the basis of transcripts per million (TPM).

RNA-seq data were obtained from the NCBI SRA database from SRP017757 (SRX213886- SRX213893) for *B. rapa* (Tong et al. 2013) and from SRP017530 (SRX209691–SRX209697) for *B. oleracea*. Gene sequences from these two Brassica species were used as probes to perform a BLASTN analysis in the relevant databases. Parameters: identity = 100 %; lengths were according to the sequenced length in databases. The data were calibrated as fragments per kilobase of exon/per million fragments mapped (FPKM) (see Supplemental Tables 2, 3, 4 and 5). The heatmaps were drawn with R script (R Development Core Team 2014), using the $\log_2$-transformed-calibrated data of expression values.

Gene structure analysis

The gene structure and intron-phase analyses were carried out using the software Gene Structure Display Server 2.0 (GSDS2.0) (Hu et al. 2015).

Results

Classification of caleosins in *A. lyrata* and three Brassica crop species

After the BLASTP and key word search in BRAD, Bol base v1.0 and Phytozome10 databases, 7, 11 and, 12 caleosin genes were obtained from the *A. lyrata*, *B. rapa*, and *B. oleracea* genomes, respectively. Twenty-two caleosin genes were obtained from the *B. napus* genome after local BLASTN and local BLASTP searches. Those sequences were used as queries to search for the most similar genes in Arabidopsis. The gene Bol012293 from *B. oleracea* was deleted as its most similar gene was AT1G23260, which encodes a protein binding/ubiquitin-protein ligase instead of a caleosin. The significant caleosin domain IPR007736 was found in the sequences of the remaining 51 genes using InterProScan5, which indicated all of them are caleosins. In the end, the final number of caleosin genes was confirmed as 7, 11, 11 and 22 in *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus*, respectively (Table 1, Supplemental Table 1).

The names of these caleosins were assigned according to the results of ortholog analyses and the assembled phylogenetic tree (Table 1). Caleosins in *A. lyrata* were named as AlCLO1 to AlCLO6 and AlCLO8, without AlCLO7; caleosins in *B. rapa* as BrCLO1 to BrCLO5 and BrCLO8, without BrCLO6 or BrCLO7; caleosins in *B. oleracea* were named as BoCLO1 to BoCLO4 and BoCLO8, without BoCLO5, BoCLO6 or BoCLO7; caleosins in *B. napus* were named as BnCLO1–BnCLO5 and BnCLO8, without BnCLO6 or BnCLO7.

Multiple alignments were made of the 51 caleosin protein sequences from *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus*, and 8 caleosin protein sequences from *A. thaliana* (see Supplemental Figure 1). In order to demonstrate key motifs more clearly, the alignment of H-form insertions, Ca^{2+} binding motifs and proline knots are shown in Fig. 1. The results indicated that BrCLO2-2 and BrCLO8-2 have lost their EF-hand calcium-binding domain. The proline knot P-3X-PS-3X-P plays a vital role during anchoring of lipid droplets and BrCLO2-2 and BrCLO8-2 have also lost this motif. The first proline was replaced by a threonine residue in the proline knots of BoCLO1-2 and BnCLO1-3; the serine was replaced by a tyrosine in the proline knot of BnCLO2-2; the serine was replaced by a threonine residue in the proline knot of BnCLO8-3. The second proline becomes phenylalanine in all L-caleosins. Additionally serine becomes threonine in BrCLO4-1. The N-terminal of BoCLO3-1 was found to be extremely long (see Supplemental Figure 1) as is reported for caleosins from monocotyledons (Chen et al. 2012). Previous research indicated that histidine plays the most significant role among haem-binding motifs (Allen et al. 2005). Hanano et al. (2006) indicated the haem-oxygenase activity of AtCLO1 relied on two highly conserved histidine sites, H^{70} and H^{138}. Results showed that these two sites were highly conserved among all brassicaceous caleosins except for BrCLO2-2, BoCLO8-2 and AtCLO7 (Supplemental Fig. 1). BrCLO2-2 and BoCLO8-2 lost these two histidine sites because of the severe truncation of their N-termini. AtCLO7 showed changes in these two sites: H^{70} to C and H^{138} to V (Supplemental Fig. 1). The haem-binding motif may be viewed as crucial evidence to characterize whether or not a protein has a peroxygenase function.

Eight caleosins from *A. thaliana* were classified as H-caleosins and L-caleosins in a former study (Shen et al. 2014). According to multiple alignments, the N-termini of H-caleosins share an extra 29 amino acids (characterized as an H-form insertion) compared with L-caleosins (Shen et al. 2014). Among the 51 caleosins identified in this study, there are 30 H-caleosins and 18 L-caleosins, except for BrCLO2-2 and BoCLO8-2, both with entire N-terminal truncations and BoCLO2 with a largely truncated N-terminus. Homologous genes were found to be in the same cluster; for instance, CLOX ($X = 1$–8) were found to be clustered. Results indicated that three caleosins which were hard to classify were clustered with H-caleosins and thus, they probably belongs to this class. Therefore, there are 33 H-caleosins and 18 L-caleosins among these 51 caleosins. The differences of the hydrophobicity and the fold patterns between the monolayer-located caleosins and bilayer-located caleosins (using AtCLO1 and AtCLO3 as examples), were very small (Supplemental Figure 2), which did not provide enough information for definitively identifying the membrane locations of these caleosins.

Phosphorylation sites were predicted by NetPhos 2.0 (see Supplemental Figure 1). Results showed that among H-caleosins (using AlCLO1 as a standard), residues S^{140} and Y^{145} were found to be conserved except for strongly truncated BrCLO2-2 and BoCLO8-2; S^{214} was conserved except in BrCLO2-1 and AlCLO3. Among L-caleosins (using AlCLO4 as a standard), S^{65} and S^{106} were highly conserved; S^{191} was conserved except for in AlCLO6, BrCLO5, BnCLO5-1 and BnCLO5-2, while T^{169} was conserved in all species except in CLO5.

Using AlCLO1 as a standard, residue F^{227} remained unchanged, as did G^{90}, R^{92}, D^{146} and G^{149} among all caleosins found in those brassicaceous plants (except in the case of the highly truncated BrCLO2-2 and BoCLO8-2), which might indicate these sites play important roles in structure/function of these caleosins. Similar sites were also confirmed by another online prediction tool PlantPhos (Lee et al. 2011).

Motif and physicochemical property analyses of caleosins from *A. lyrata* and three Brassica crop species

MEME was used to perform motif analyses of the 51 caleosins. The results (Fig. 2; Table 2) showed all motifs were specific to caleosins. The motif structure of H-caleosins is 7-4-1-2-5-3-variable C-terminal, while the L-caleosins has a 6-1-2-5-3 motif. The highly truncated BrCLO2-2 and BoCLO8-2 have no N-terminus and they have 7-4-1-2 central motifs. According to this result, and the deletion of their proline knot, they have probably lost their calcium-binding capacity and caleosin-anchoring-capability. In addition, BoCLO2 lost the N-terminal of its calcium-binding domain (motifs 7-4) and part of its C-terminal (motif 5). BrCLO2-1 lost its C-terminal entirely (motifs 5-3-10); CLO5 in all species show a deficiency of upstream motif 6 of the calcium-binding domain.

Isoelectric points of the 51 caleosin proteins were also calculated. The result (Table 2) showed that all H-caleosins shared a pI lower than 6.0 in all species except in CLO8, while all L-caleosins shared a pI higher than 7.0. The

Table 1 Gene catalog of caleosin family in *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus*

	Gene name	Locus ID	Alias	Size (ORF, bp)	Size (AA)	pI	Mw	Type
A. lyrata	AlCLO1	329010		738	245	5.81	28,056.07	H
	AlCLO2	495609		732	243	5.49	27,786.48	H
	AlCLO3	482279		720	239	5.44	26,998.51	H
	AlCLO8	894836		717	238	7.06	27,039.66	H
	AlCLO4	926397		588	195	9.23	21,976.98	L
	AlCLO5	472583		630	209	9.47	23,685.99	L
	AlCLO6	926398		591	196	9.05	22,062.02	L
B. rapa	BrCLO1	*Bra026407[a]		738	245	5.81	28,129.11	H
	BrCLO2-1	Bra002921		525	174	6.08	19,715.62	H
	BrCLO2-2	*Bra002920		264	87	5.04	10,370.67	H
	BrCLO3-1	*Bra005501		720	239	5.44	26,933.45	H
	BrCLO3-2	*Bra022936		720	239	5.1	26,952.4	H
	BrCLO3-3	*Bra021847		720	239	5.1	26,960.41	H
	BrCLO8	Bra020623		717	238	7.77	27,113.7	H
	BrCLO4-1	*Bra003948		579	192	8.41	21,500.32	L
	BrCLO4-2	Bra016195		579	192	7.79	21,634.42	L
	BrCLO4-3	*Bra007934		579	192	9.38	21,503.47	L
	BrCLO5	Bra012369		612	203	9.48	22,983.14	L
B. oleracea	BoCLO1-1	Bol006498		738	245	5.65	28,078.05	H
	BoCLO1-2	*Bol039674		738	245	5.81	28,133.1	H
	BoCLO2	*Bol038946		558	185	5.59	21,224.91	H
	BoCLO3-1	*Bol037013		960	319	NA[b]	NA	H
	BoCLO3-2	*Bol037934		720	239	5.11	26,928.35	H
	BoCLO3-3	*Bol027265		720	239	5.3	26,946.4	H
	BoCLO8-1	Bol001379		717	238	7.76	27,085.69	H
	BoCLO8-2	Bol016621		255	84	8.97	9804.19	H
	BoCLO4-1	Bol000440		579	192	7.13	21,635.41	L
	BoCLO4-2	*Bol017417		579	192	8.93	21,499.38	L
	BoCLO4-3	*Bol041428		579	192	9.44	21,517.54	L
B. napus	BnCLO1-1	BnaC08g12280D	CDX97363	738	245	5.65	28,062.05	H
	BnCLO1-2	*BnaA01g15860D	*CDX89307	738	245	5.81	28,129.11	H
	BnCLO1-3	*BnaC01g18950D	*CDX94289	738	245	5.97	28,103.12	H
	BnCLO2-1	BnaA10g09480D	CDY17314	735	244	5.74	27,970.78	H
	BnCLO2-2	BnaC09g31660D	CDY19900	735	244	5.9	28,216.14	H
	BnCLO3-1	*BnaA03g15390D	*CDX84691	720	239	5.19	26,875.32	H
	BnCLO3-2	*BnaC03g18600D	*CDX79524	720	239	5.18	26,822.26	H
	BnCLO3-3	*BnaA04g19410D	*CDY37727	720	239	5.11	26,928.35	H
	BnCLO3-4	*BnaC04g43780D	*CDY23744	720	239	5.11	26,928.35	H
	BnCLO3-5	*BnaC04g11110D	*CDY19488	720	239	5.3	26,946.4	H
	BnCLO3-6	*BnaA05g10200D	*CDY17168	720	239	5.44	26,963.47	H
	BnCLO8-1	BnaA02g30970D	CDY33890	717	238	7.77	27,113.7	H
	BnCLO8-2	BnaC02g39270D	CDY05610	717	238	7.76	27,085.69	H
	BnCLO8-3	BnaC07g27360D	CDX85425	717	238	7.72	27,127.81	H
	BnCLO4-1	BnaC06g31770D	CDY11562	579	192	7.13	21,635.41	L
	BnCLO4-2	BnaA07g28790D	CDX96268	579	192	7.79	21,634.42	L
	BnCLO4-3	*BnaA07g23760D	*CDX68336	696	231	8.92	25,983.68	L
	BnCLO4-4	*BnaC06g24570D	*CDY23488	597	198	9.28	22,401.66	L
	BnCLO4-5	*BnaC02g20100D	*CDY09468	579	192	9.44	21,517.54	L
	BnCLO4-6	*BnaA02g15090D	*CDY03728	579	192	9.38	21,503.47	L
	BnCLO5-1	BnaA07g10050D	CDY25777	612	203	9.32	22,950.02	L
	BnCLO5-2	BnaC07g13170D	CDY12477	612	203	9.32	22,990.08	L

[a] * means the gene of Brassica is the ortholog of corresponding caleosin in *A. thaliana*

[b] NA means no available value

absence of a calculated pI for BoCLO3-1 was due to the un-sequenced 34 amino acids at its N-terminus.

Duplication analysis of the caleosin gene family of *A. lyrata* and three Brassica crop species

The location maps of *B. rapa*, *B. oleracea* and *B. napus* caleosin genes are shown in Fig. 3. A location map of *A. lyrata* caleosins was unattainable for those genes only located on chromosome segments. All of the *B. rapa* caleosins locate to chromosomes and show a distribution of Chr1.2.3.4.5.7.10; seven of the *B. rapa* caleosins locate on chromosomes (Chr1.4.6.7.9), while 4 others locate on chromosome segments. They show a uniform distribution on chromosomes in that most of the chromosomes only have a single (or at most 3) gene(s) on them.

According to the map, two pairs of genes BrCLO2-1 and BrCLO2-2 and AlCLO4 and AlCLO6 are products of tandem duplication.

Nine of eleven *B. rapa* caleosins experienced segmental duplication over times ranging from 56.78 MYA to 6.89 MYA. The earliest segmental duplication took place on BrCLO4-1 and BrCLO5, while the most recent occurred on BrCLO3-2 and BrCLO3-3. Ten of eleven *B. oleracea* caleosinsexperienced segmental duplication, but showed a narrower range of duplication age from 14.34 MYA to 7.35 MYA (Table 3).

The natural hybridization of *B. rapa* and *B. oleracea* produced the allotetraploid *B. napus* (Chalhoub et al. 2014; Nagaharu 1935). Thus, the parents of *B. napus* caleosin genes could be determined by orthologs, phylogenetic trees and the available *B. napus* chromosomal data (Table 4). All parental genes could be verified except for BnCLO5-2.

Phylogenetic analysis of caleosins in *A. thaliana*, *A. lyrata* and three Brassica crop species

The phylogenetic trees (Fig. 4) were built using caleosin sequences from *A. lyrata* and three Brassica crops and 8 sequences from *A. thaliana* and one sequence from a cycad. Both NJ and ML trees shared a bootstrap value higher than 70 %. The similar topological structure of the two trees guaranteed a high fidelity. Two trees showed that the 59 brassicaceous caleosins could be classified into two families, H-caleosins and L-caleosins, with orthologs clustered separately, consistent with multiple alignments. Our phylogenetic analyses also indicated the even older single caleosin from a cycad presented a closer relationship to H-caleosins, which conforms to the result that the cycad caleosin is an H-caleosin (Shen et al. 2014). This indicates L-caleosins emerged after the speciation of cycads and that the L-caleosins might therefore have arisen from H-caleosins.

Expression analysis of Brassica caleosin genes

Due to insufficient data for *A. lyrata*, we were unable to perform expression analyses. Data for *B. rapa*, *B. oleracea* and *B. napus* were obtained from public expression databases. EST data were accessible for all three species, while RNA-seq data of *B. rapa* and *B. oleracea* were also analyzed (Fig. 5, Supplemental Tables 2, 3, 4 and 5). There were limited EST data for *B. oleracea*. At least one expression datum was collected for each, except for BoCLO2, BnCLO2-1, BnCLO3-1, BnCLO4-5, BnCLO5-1, BnCLO5-2 and BnCLO8-3, which might indicate the lack of expression of these genes.

EST and RNA-seq data of *B. rapa* were collected, with an absence of data for buds in RNA-seq. The expression patterns obtained by EST and RNA-seq databases were consistent except for BrCLO4-3. BrCLO4-3 EST data showed the highest expression level in roots and no expression in floral tissues, yet the RNA-seq data indicated it presented its highest expression level in floral tissues, with far lower levels in roots (see Supplemental Tables 2, 3, 4 and 5). BrCLO3-3 was the most highly expressed of the *B. rapa* caleosins and BrCLO3-1 ranked second. Their expression patterns were both floral-specific; no RNA-seq data was obtained for BrCLO5, but the EST data indicated its expression in buds; no EST data was obtained for BrCLO2-1 and BrCLO2-2, but their RNA-seq data showed an extremely low expression level; for BrCLO8, the low expression level in RNA-seq data and the deficiency of its EST might be evidence for a loss of function (Fig. 5a).

The EST data for *B. oleracea* was not comprehensive enough for us to draw a detailed map; it was only used as a reference for expression. The heatmap of gene expression was constructed primarily using RNA-seq data. BoCLO3-2 showed the highest expression level while BoCLO3-3 was second. There was not any expression data for BoCLO2. No EST data were collected for BoCLO1-1 and BoCLO8-2, while only a few RNA-seq data were collected (Fig. 5b).

ESTs data for *B. napus* was collected for heatmap (Fig. 5c). Leaf, root, stem and whole-plant were not included in the heatmap because there are no data available for these tissues among any of the *B. napus* caleosin genes. The lack of expression data for BnCLO2-1, BnCLO3-1, BnCLO4-5, BnCLO5-1, BnCLO5-2 and BnCLO8-3 might indicate that they have extremely low expression levels. BnCLO3-3 and BnCLO3-4 exhibited the highest expression level followed by BnCLO1-2 and BnCLO1-3. We also discovered a meristem-specific expression for BnCLO3-3 and BnCLO3-4, and a seed-specific expression for BnCLO1-2 and BnCLO1-3.

To determine which of the orthologs or paralogs make a greater contribution to functional diversity among caleosin genes, the triplicated gene CLO3 was used as a typical

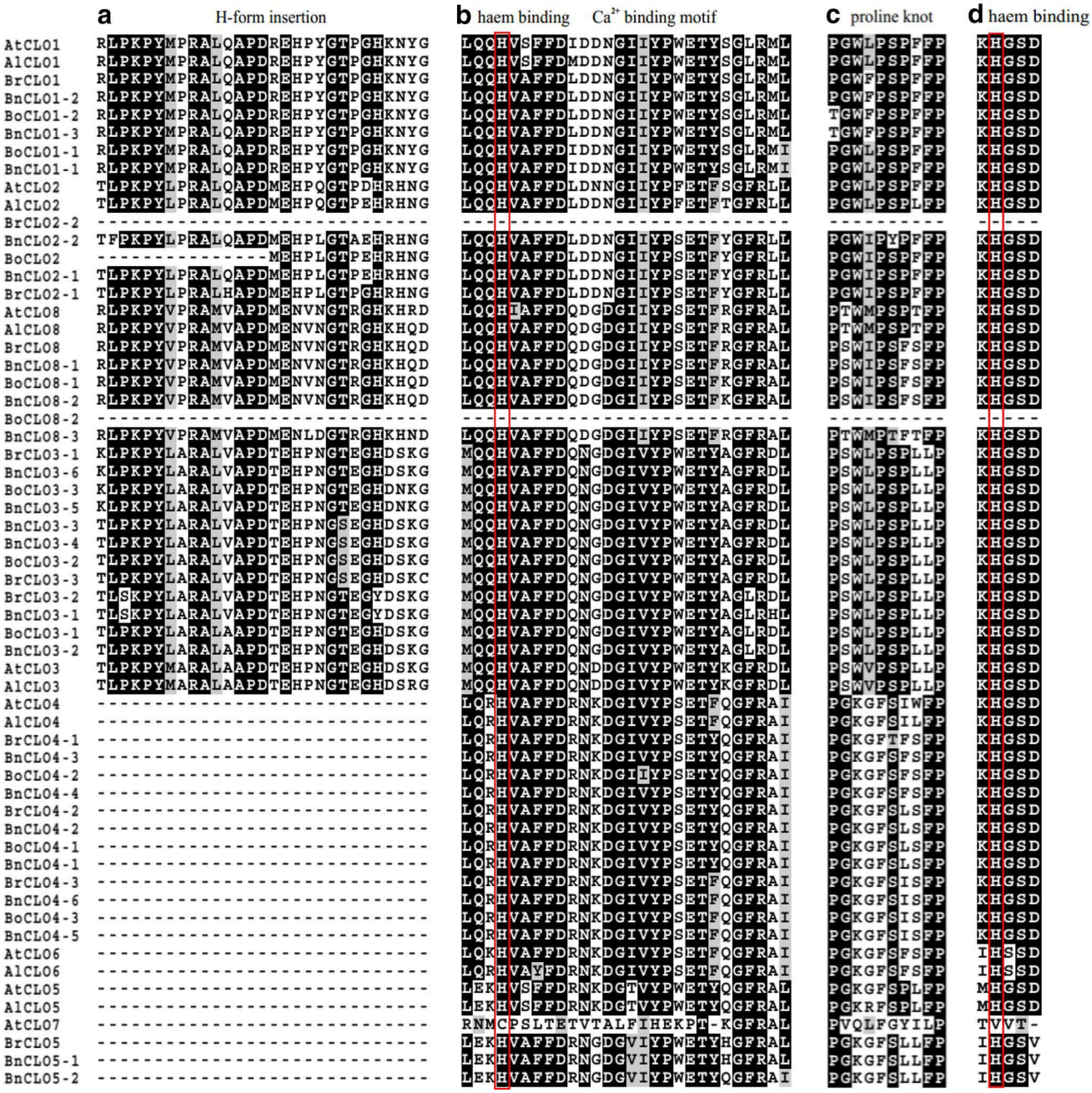

Fig. 1 Alignment of H-form insertion (**a**), Ca^{2+} binding motif (**b**), proline knot (**c**) and haem-binding motif (**b**, **d**) from caleosins in *A. thaliana*, *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus*. *Dashes* represent gaps introduced in the sequences for optimal alignment, *red boxes* represent haem-binding motif (color figure online)

example to perform analysis. Synteny Block showed that BrCLO3-2 and BoCLO3-1, BrCLO3-3 and BoCLO3-2, as well as BrCLO3-1 and BoCLO3-3 are orthologs. An expression map showed that orthologs had more similar expression levels than do paralogs, which means that paralogs made greater contributions to functional diversity. Comparisons were made between *B. napus* genes and their progenitor parents. *B. napus* caleosin genes showed distinct expression diversities with their parents except BnCLO2 and its corresponding parents, BrCLO2 and BoCLO2.

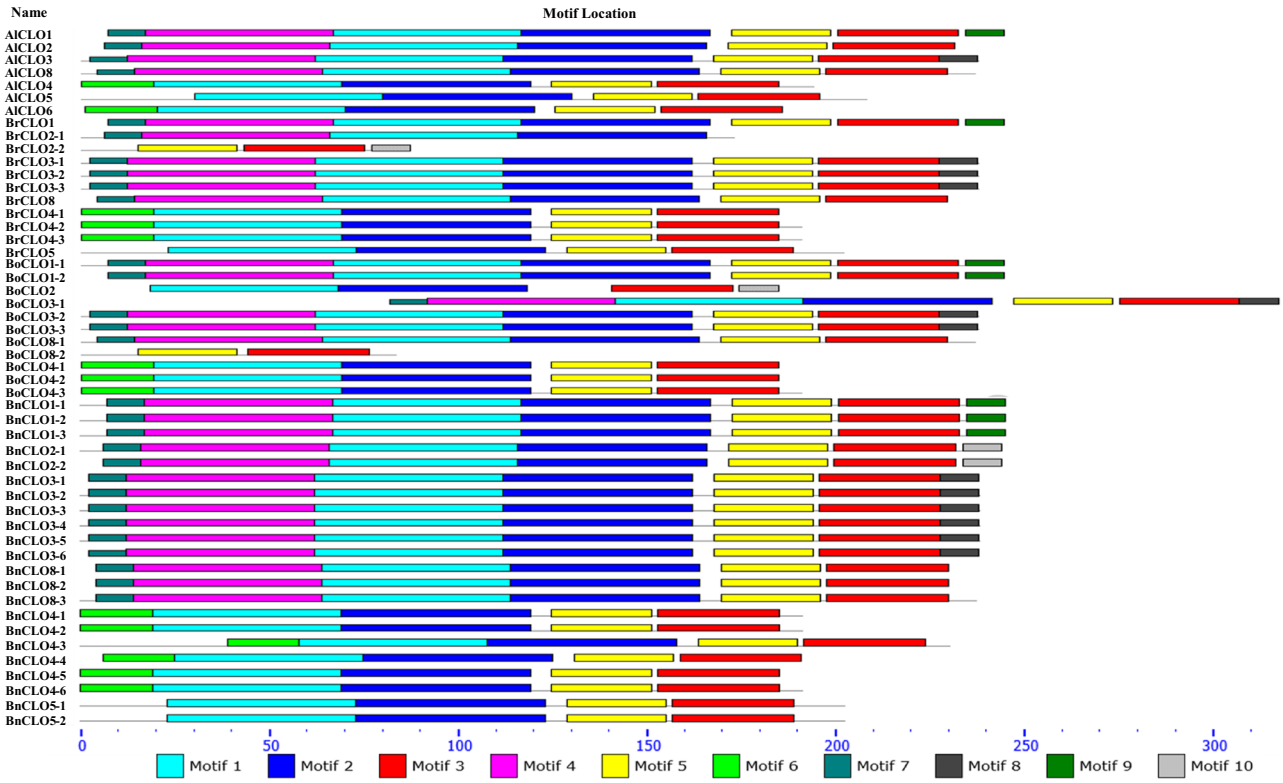

Fig. 2 Distribution of motifs in caleosins of *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus*. Sequences and annotations of motifs are provided in Table 2

Table 2 Motif sequences and annotations

Motif	Width	Best possible match	Annotation by multiple alignment	Annotation by InterProScan
1[a]	50	QQHVAFFDQNGDGIVYPWETYQGFRMIGCNYIS-SAFWSLFINMALSYVTR	Ca^{2+} binding site	Caleosin (IPR007736)
2	50	PGWGPSPSFPIYIKNIHKAKHGSDSSTYDTEGRFVPVN-LENIFSKYARTH	Proline knot	Caleosin (IPR007736)
3	32	LYYLCKDENGFLHKEAVRRCFDGSLFEYCAKQ	C-terminal-Pi	Caleosin (IPR007736)
4	50	PCAPVTGHRRVRNDLDDRLPKPYMPRALQAPD-MEHPNGTEGHDHQGMSVM	H-caleosin insertion	Caleosin (IPR007736)
5	26	KELWQMTEGNRMAWDRFGWIANKVEW	NH[b]	NH
6	19	MASSASTGVKFVPEEDNFL	L-caleosin N-terminal	NH
7	10	MEREAMATVA	H-caleosin N-terminal	NH
8	10	NKEKANSRKQ	CLO3 C-terminal	NH
9	10	GINEDKTAYY	CLO1 C-terminal	NH
10	10	EMSEYRSYYY	CLO2 C-terminal	NH

[a] Motif number refer to the motifs identified in Fig. 2

[b] NH means no hits found

Cis-element analysis of crucial caleosin genes from *B. napus*

The promoter sequences of BnCLO1-2 and BnCLO1-3 consisting of 333 and 335 bp, respectively, were used as they are both shorter than 2000 bp (the selection criterion).

According to the results obtained using PLACE, an ABRERATCAL element (S000507) was found in BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4. An RY-repeat element was found in BnCLO1-2, BnCLO1-3

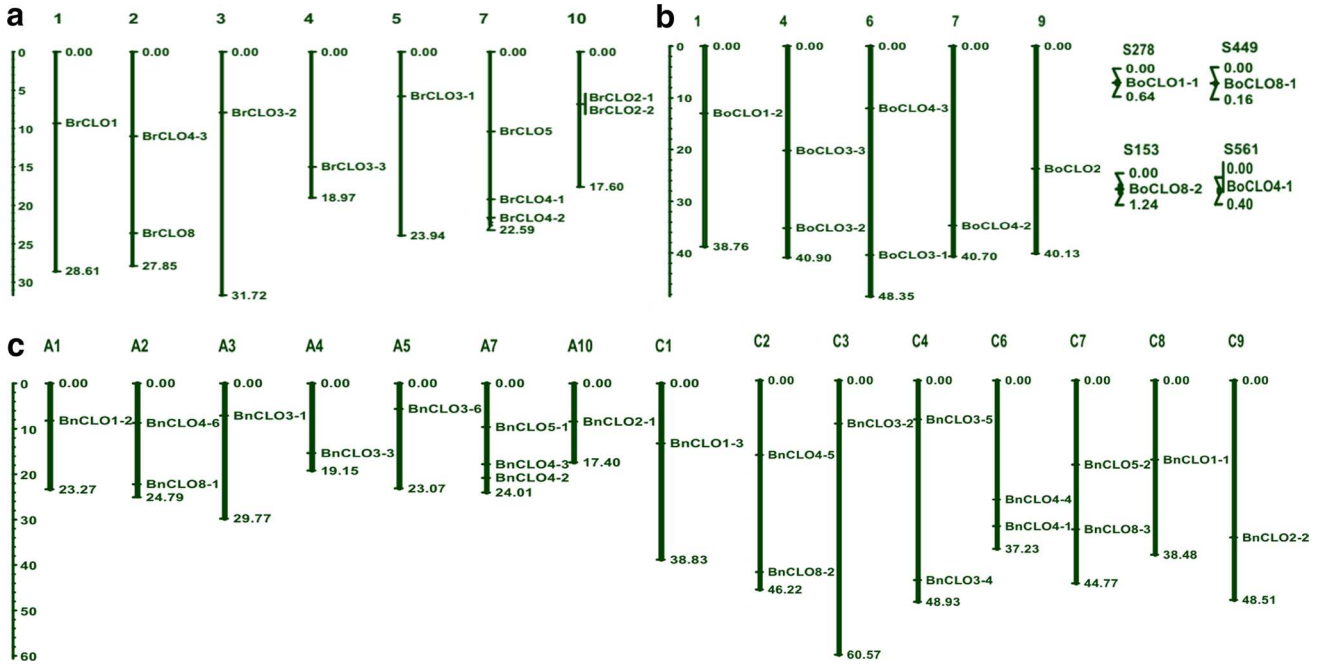

Fig. 3 Chromosomal architecture of caleosin genes in *B. rapa* (**a**), *B. oleracea* (**b**), and *B. napus* (**c**). The *number/word* at the *top* designates the chromosome number. S means scaffold. *Scale bar* on the *left* designates 30 Mb (**a**), 40 Mb (**b**) and 60 Mb (**c**) chromosomal distance. Chromosome sizes (Mb) are *marked* at the *bottom* end of each chromosome

Table 3 Segmentally duplicated caleosin genes in *A. lyrata*, *B. rapa* and *B. oleracea*

Species	Duplicated caleosin 1	Caleosin 1 locus	Duplicated caleosin 2	Caleosin 2 locus	Flanking genes	Ks	Age (MYA)
A. lyrata	AlCLO5	472583	AlCLO4	926397	5	1.4700	49.00
	AlCLO1	329010	AlCLO2	495609	10	0.8696	28.99
B. rapa	BrCLO2-1	Bra002921	BrCLO1	Bra026407	4	1.2297	40.99
	BrCLO3-1	Bra005501	BrCLO3-3	Bra021847	12	0.3203	10.68
	BrCLO3-3	Bra021847	BrCLO3-2	Bra022936	11	0.2066	6.89
	BrCLO4-1	Bra003948	BrCLO4-3	Bra007934	8	0.4034	13.45
	BrCLO4-3	Bra007934	BrCLO4-2	Bra016195	9	0.3584	11.95
	BrCLO4-1	Bra003948	BrCLO5	Bra012369	6	1.7034	56.78
B. oleracea	BoCLO1-1	Bol006498	BoCLO1-2	Bol039674	7	0.2592	8.64
	BoCLO3-3	Bol027265	BoCLO3-2	Bol037934	12	0.3096	10.32
	BoCLO3-1	Bol037013	BoCLO3-3	Bol027265	7	0.3144	10.48
	BoCLO4-2	Bol017417	BoCLO4-3	Bol041428	2	0.4303	14.34
	BoCLO4-1	Bol000440	BoCLO4-3	Bol041428	2	0.3506	11.69
	BoCLO8-1	Bol001379	BoCLO8-2	Bol016621	9	0.2205	7.35

and BnCLO3-4 (see Supplemental Table 6). PlantCARE analysis showed that an RY-element is present in BnCLO3-4, while Skn-1 motifs were found in BnCLO1-2 and BnCLO1-3. Several phytohormone-related elements were found in BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4. Additionally, some stress-related elements were found in BnCLO3-3 and BnCLO3-4, which was supported by their reported expressions patterns under stress (Supplemental Tables 2, 3, 4 and 5).

Gene structure analysis of caleosins

The intron/exon structures of each caleosin gene from Brassica species were analyzed (Fig. 6). The results showed a highly conserved intron/exon structure among these caleosins. Eighty-four percent (38/44) of caleosins contained 6 exons and 5 introns, except for BnCLO4-3 (seven exons and six introns), BnCLO4-4, BoCLO2 and BrCLO2-1 (five exons and four introns), BoCLO8-2 and BrCLO2-2 (two exons and one intron).

Table 4 Caleosins in *B. napus* and their diploid parents

	B. napus locus in GGD	*B. napus* locus in ENA	Gene name of *B. napus* gene	Parent locus	Gene name of parent
CLO1	BnaC08g12280D	CDX97363	BnCLO1-1	Bol006498	BoCLO1-1
	*BnaA01g15860D	*CDX89307	BnCLO1-2	*Bra026407	BrCLO1
	*BnaC01g18950D	*CDX94289	BnCLO1-3	*Bol039674	BoCLO1-2
CLO2	BnaA10g09480D	CDY17314	BnCLO2-1	*Bra002920	BrCLO2-2
	BnaC09g31660D	CDY19900	BnCLO2-2	*Bol038946	BoCLO2
CLO3	*BnaA03g15390D	*CDX84691	BnCLO3-1	*Bra022936	BrCLO3-2
	*BnaC03g18600D	*CDX79524	BnCLO3-2	*Bol037013	BoCLO3-1
	*BnaA04g19410D	*CDY37727	BnCLO3-3	*Bra021847	BrCLO3-3
	*BnaC04g43780D	*CDY23744	BnCLO3-4	*Bol037934	BoCLO3-2
	*BnaC04g11110D	*CDY19488	BnCLO3-5	*Bra005501	BrCLO3-1
	*BnaA05g10200D	*CDY17168	BnCLO3-6	*Bol027265	BoCLO3-3
CLO4	BnaC06g31770D	CDY11562	BnCLO4-1	Bra016195	BrCLO4-2
	BnaA07g28790D	CDX96268	BnCLO4-2	Bol000440	BoCLO4-1
	*BnaA07g23760D	*CDX68336	BnCLO4-3	*Bra003948	BrCLO4-1
	*BnaC06g24570D	*CDY23488	BnCLO4-4	*Bol017417	BoCLO4-2
	*BnaC02g20100D	*CDY09468	BnCLO4-5	*Bol041428	BoCLO4-3
	*BnaA02g15090D	*CDY03728	BnCLO4-6	*Bra007934	BrCLO4-3
CLO5	BnaA07g10050D	CDY25777	BnCLO5-1	Bra012369	BrCLO5
	BnaC07g13170D	CDY12477	BnCLO5-2	NA[a]	NA
CLO6	NA	NA	NA	NA	NA
CLO7	NA	NA	NA	NA	NA
CLO8	BnaA02g30970D	CDY33890	BnCLO8-1	Bra020623	BrCLO8
	BnaC02g39270D	CDY05610	BnCLO8-2	Bol001379	BoCLO8-1
	BnaC07g27360D	CDX85425	BnCLO8-3	Bol016621	BoCLO8-2

[a] NA means no available value

The intron phases were also analyzed (Fig. 6). Phase 0 means the intron locates between two codons, while phase 1 means the intron locates after the first nucleotide of the codon and phase 2 means the intron locates after the second nucleotide of the codon. Intron phases of the brassicaceous caleosins were highly conserved. Those genes which showed the same structure also shared the same phase of 1, 1, 0, 2 and 2. The intron phase of BrCLO4-3 was 0, 1, 1, 0, 2 and 2; BnCLO4-4 and BoCLO2 were 1, 0, 2 and 2; BrCLO2-1 was 1, 1, 0 and 2, while BoCLO8-2 and BrCLO2-2 were 2. For those genes which had no less than 4 exons, their last four exons were very similar in size.

Discussion

Phylogenetic analysis of caleosin genes in Brassicaceae plants

There are eight caleosin proteins in *A. thaliana*, named as AtCLO1–AtCLO8. Four of these (AtCLO4–AtCLO7) were determined to be L-caleosins. Among them, AtCLO4 and AtCLO5 are segmentally duplicated, while AtCLO4 and AtCLO6, AtCLO5 and AtCLO7 are tandem duplications. Among other 4 H-caleosins, AtCLO1 and AtCLO2 are segmental duplications (Shen et al. 2014). Neither *A. lyrata* nor the three Brassica species had homologous genes of AtCLO7 (Table 1). Arabidopsis and Brassica plants experienced an archaic speciation (Blanc et al. 2003; Town et al. 2006; Yang et al. 1999), which may indicate the tandem duplication of AtCLO5 and AtCLO7 occurred after the speciation between *A. thaliana* and it relative *A. lyrata* approximately 10 MYA (Hu et al. 2011). While a homologous gene of AtCLO6 was found in *A. lyrata*, no homologs were found in the three Brassica species (Table 1). This may suggest that the tandem duplication of AtCLO4 and AtCLO6 happened after the speciation of Arabidopsis and Brassica plants, but before the speciation of *A. thaliana* and *A. lyrata* approximately10–20 MYA (Blanc et al. 2003; Hu et al. 2011; Town et al. 2006; Yang et al. 1999). It is probable that a series of tandem duplications occurred in the *A. thaliana* genome, which is consistent with previous conclusions drawn by others (Cannon et al. 2004).

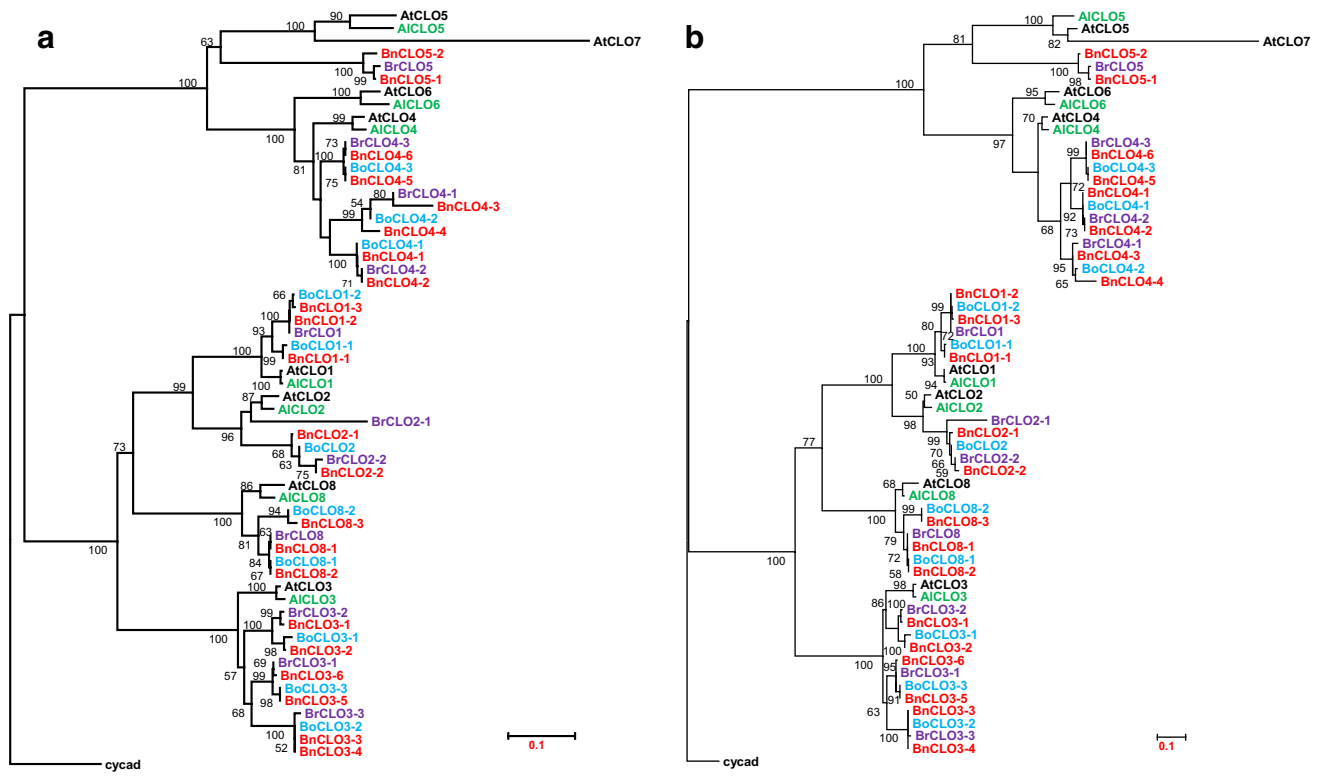

Fig. 4 Phylogenetic trees of caleosins in *A. thaliana*, *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus* by the NJ method in MEGA 6.06 (**a**) and the ML method in PhyML 3.0 (**b**). *Black*, *green*, *purple*, *blue* and *red* represents caleosin genes from *A. thaliana*, *A. lyrata*, *B. rapa*, *B. oleracea* and *B. napus*, respectively. *Numbers* next to a node indicate percent bootstrap values (>50 %). The *bars* at the *bottom* indicate the average number of amino acid substitutions per site (color figure online)

Homologous genes of AtCLO1 and AtCLO2 are found in all of four species (Table 1), indicating that segmental duplication of CLO1 and CLO2 happened before speciation of these species. It probably happened in common ancestors of *A. thaliana* and Brassica plants; this is consistent with a previous study (Shen et al. 2014). The time of segmental duplications of AlCLO1 and AlCLO2, BrCLO1 and BrCLO2-1 also provides plausible evidence of this. However, no segmental duplication is found between BoCLO1 and BoCLO2 (Table 1), which indicates a rearrangement in the *B. oleracea* genome after the formation of this species, approximately 4.6 MYA. This conclusion is consistent with other studies (Liu et al. 2014).

Homologous genes of AtCLO4 and AtCLO5 are identified in three species except *B. oleracea.* BnCLO5-2 evolved from the C-genome of *B. oleracea* (Table 1), indicating the loss of BoCLO5 after the formation of *B. napus.* The segmental duplication of CLO4 and CLO5 in *A. thaliana*, *A. lyrata* and *B. rapa* happened between 56.78–40.33 MYA (Shen et al. 2014; this paper), earlier than formations of brassicaceous plants approximately 24–40 MYA (Blanc et al. 2003). This confirms the segmental duplication of CLO4 and CLO5 in common ancestors of *A. thaliana* and Brassica plants (Shen et al. 2014). BoCLO5 was lost in the last 7500 years, after the formation of *B. napus*. These results may indicate that while the majority of gene elimination in *B. oleracea* occurred before 4.6 MYA (Liu et al. 2014), some elimination happened within the last 7500 years.

Nine *B. rapa* genes were segmentally duplicated between 7 and 13.67 MYA, except for BrCLO4 and BrCLO5, BrCLO1 and BrCLO2. All ten segmental duplications of *B. oleracea* genes occurred between 14.34–7.35 MYA (Table 1). A specific process of whole-genome triplication occurred in Brassica plants between 5 and 16 MYA (Liu et al. 2014; Song et al. 1995; Wang et al. 2011), which may indicate that these segmental duplications are the results of whole-genome triplication.

There are two tandem duplications in CLO2 genes in *B. rapa* (BrCLO2-1 and BrCLO2-2), while there is only one CLO2 found in *B. oleracea.* In the genome of *B. napus* no gene has been found to have evolved from BrCLO2-1. This indicates that *B. rapa* may have experienced tandem duplication after its speciation with *B. oleracea* approximately 4.6 MYA. This result is consistent with a previous study which reported that tandem duplications occurred throughout the evolution of *B. rapa* (Wang et al. 2011). Among these caleosins, BnCLO8-3 was evolved from BoCLO8-2.

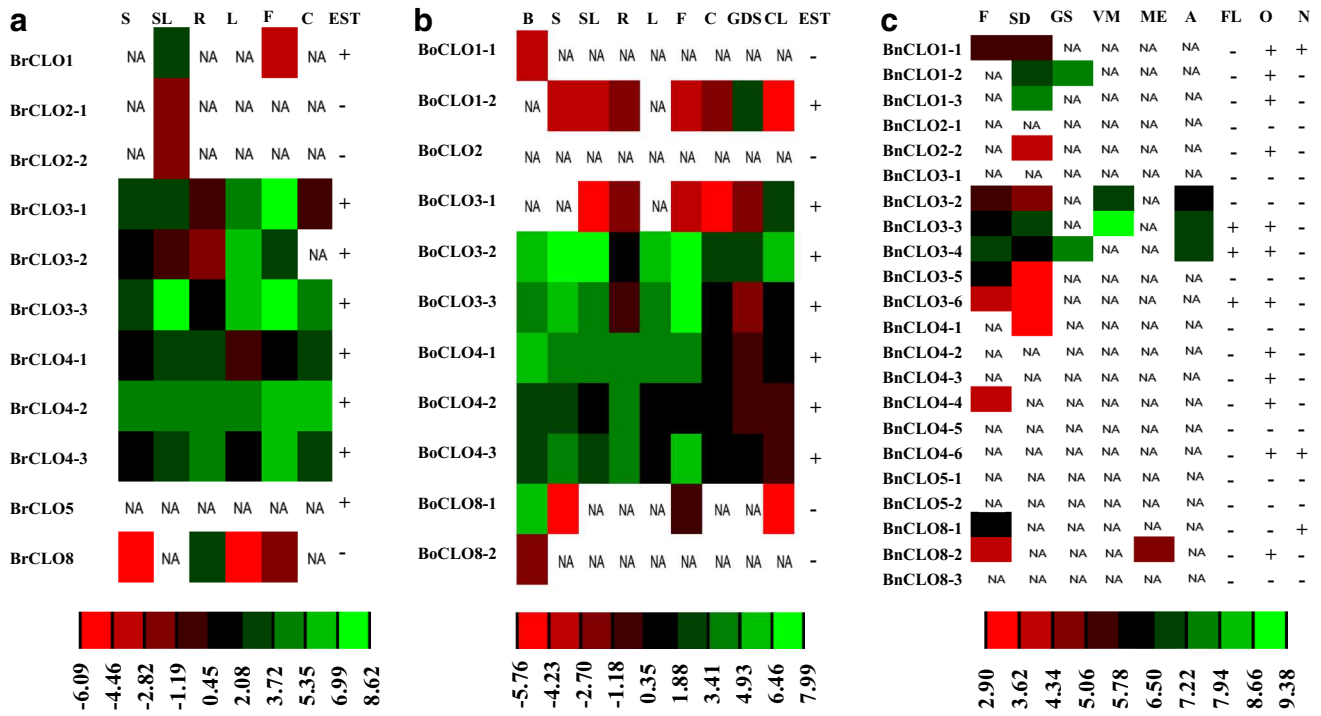

Fig. 5 Expression pattern of caleosin genes in *B. rapa* (**a**), *B. oleracea* (**b**), and *B. napus* (**c**). The heat map were generated by log2 normalized values of caleosin genes. The *color scale* representing log2 values is shown at the *bottom*. The *plus sign* means expression data exist in the corresponding library. The *minus sign* means no expression data exist in the corresponding library. *S* stem, *SL* silique, *R* root, *L* leaf, *F* flower, *C* callus, *B* bud, *GDS* germinated seeds, *CL* cotyledons, *SD* seed, *GS* germinating seed, *VM* vegetative meristem, *ME* microspore-derived embryo, *A* abnormal, *FL* full-length, *O* other, *N* none, *NA* not available

The structure of BnCLO8-3 was conserved but BoCLO8-2 was severely truncated (see Supplemental Figure 1) and contained only two exons and one intron (phase 2). This indicates that the truncation happened after the speciation of *B. napus*. BnCLO4-3 evolved from BrCLO4-1, while BnCLO4-4 evolved from BoCLO4-2. BrCLO4-1 and BoCLO4-2 shared a conserved gene structure and intron phase. In contrast, the gene structures and intron phases of BnCLO4-3 and BnCLO4-4 were different, which appears to be unique within the *B. napus* genome. Gene structures of BoCLO2 and BrCLO2-2 were different, while no change was found in caleosins from their diploid parents. This indicated these changes/differences arose after the emergence of *B. napus*.

According to results of this study, there were six caleosins (CLO1, CLO2, CLO3, CLO4, CLO5, CLO8) in the genome of the common ancestor of Arabidopsis and Brassica plants. A whole genome triplication happened among Brassica plants after their speciation from Arabidopsis (Wang et al. 2011). However, three homologs of CLO3 and CLO4 were found in the *B. napus* genome, while there were only 0–2 homologs of others (Table 1 and S1). This indicated only CLO3/4 homologies were retained, while others lost some (or even all) members of their homologous genes after the triplication. According to the 'gene dosage hypothesis' (Birchler and Veitia 2007), CLO3/4 might play more important roles.

Sub-functionalization, neo-functionalization and non-functionalization may also occur in a multigene family. These processes may be caused by gene mutation or structural transformation (Lynch and Conery 2000; Lynch 2002). For instance, using AlCLO4 as a standard, in the L-caleosin CLO5, a conserved phosphorylation site S^{191} is substituted by H in BrCLO5, BnCLO5-1 and BnCLO5-2, and T^{169} is replaced by A in these three caleosins (see Supplemental Figure 1). Motif analyses showed that part of the N-terminus of CLO5 had been lost (Fig. 2). Expression analyses indicated its extremely low expression level (Fig. 5). Phylogenetic analysis shows its loss from the *B. oleracea* genome after the formation of this species (Fig. 4). Collectively, these data indicate the caleosin CLO5 engendered its non-functionalization through gene mutation and structural transformation, and would be lost upon subsequent evolution.

Caleosin gene families and their expansion patterns in Brassicaceae

In this study, 7, 11, 11 and 22 caleosin genes were identified from the genomes of *A. lyrata*, *B. rapa*, *B. oleracea*

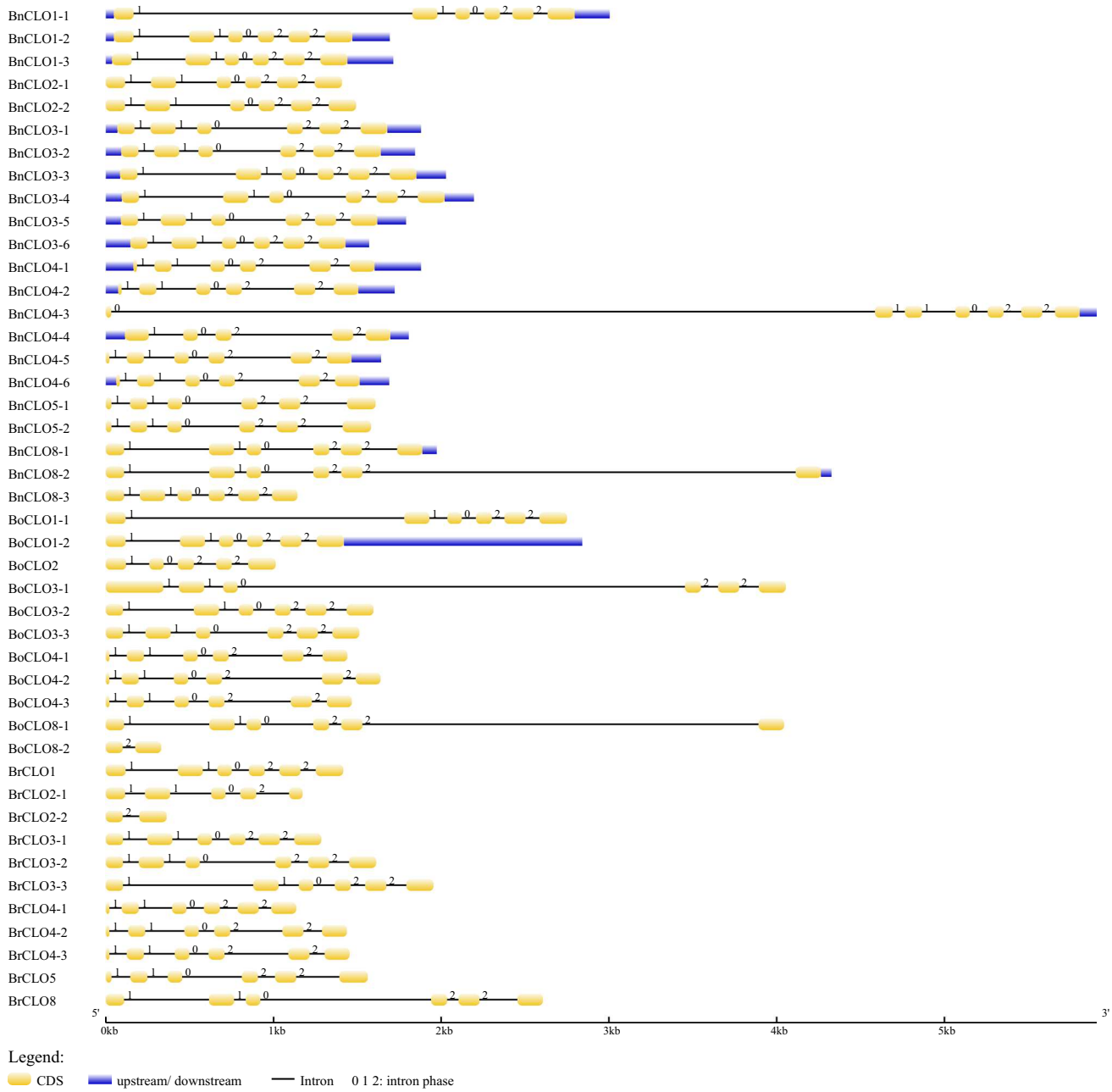

Fig. 6 Gene structure and intron phase of Cruciferous caleosins predicted by GSDS 2.0. *Yellow rectangles* represent CDS; *blue rectangles* represent upstream/downstream regions; the *thin lines* represent the intron and the figures represent the phase of corresponding introns (color figure online)

and *B. napus*, respectively. Caleosins from *A. lyrata* and *B. oleracea* are newly-discovered in this study. Information regarding *B. rapa* caleosins is consistent with previous research (Hu et al. 2013). A 27-kDa-caleosin and a 25-kDa-caleosin have been identified in *B. napus* embryos (Hernandez-Pinzon et al. 2001); a seed-specific caleosin has also been identified in seed of oilseed rape by Katavic et al. (2006). Nine caleosins (including seven BnCLO1/2 and two BnCLO3) had been assembled from the EST-library of *B. napus* using *A. thaliana* sequences as probes, seven of which (including five BnCLO1/2 and two BnCLO3) have been confirmed by means of proteomics (Jolivet et al. 2009). The newly-identified BnCLO1/2 sequences do not fit the number of caleosin genes assembled from the *B. napus* EST database. While this may be due to a splicing error, results in this study fit the proteomics well. Six BnCLO3s are identified (Table 1). This number is larger than results from EST databases, which may

be due to incomplete EST data. This study also shows a larger number of caleosins than predicted by proteomics which might be due to the similarities among BnCLO3-X genes.

The caleosin gene family is a multigene-family. The numbers of *B. rapa* and *B. oleracea* caleosins are nearly 1.4 times that in *A. thaliana*, which is lower than the ratio of gene numbers between *B. rapa* or *B. oleracea* with *A. thaliana* (41,174/25,498 = 1.6, 45,758/25,498 = 1.8) (Arabidopsis genome initiative 2000; Liu et al. 2014; Wang et al. 2011). The probable reason is that both two-tandem duplications in *A. thaliana* occurred after the segregation of Arabidopsis and Brassica plants.

Among seven *A. lyrata* caleosin genes, four of them experienced segmental duplication, two experienced tandem duplication, and only one experienced both processes. Nine of the eleven *B. rapa* caleosins experienced segmental duplication, two experienced tandem duplication and one of them experienced both processes. Ten of eleven *B. oleracea* caleosins experienced segmental duplication (Table 3). We can therefore infer that segmental duplication is the main expansion form for those genes. The combined actions of tandem duplication and segmental duplication are the main expansion form of *A. thaliana* caleosins.

The number of *B. napus* caleosins is the sum of *B. rapa* and *B. oleracea* caleosins. This may indicate that the expansion of *B. napus* caleosins was due to natural hybridization and the subsequent process of genome doubling.

According to the results, segmental duplication is the main gene expansion form for caleosins in those brassicaceous plants.

Crucial motifs of caleosins in the tested brassicaceous plants

According to this study, caleosins can be characterized by a single calcium binding domain called an EF-hand (motif 1), proline knot (motif 2), and several potential phosphorylation and haem-binding sites. Our motif analysis indicate that these putative caleosins are able to bind calcium, phosphate and haem, which enable them to play roles during lipid droplet formation or in response to stress. Loss of these motifs in BrCLO2 and BoCLO8-2 (Fig. 2), when combined with expression analyses (Fig. 5), suggests that they may have lost their function. The changes of H^{70} to C and H^{138} to V for AtCLO7 may indicate the loss of its peroxygenase activity.

The N-terminal upstream sequence (motifs 7-4) of the calcium-binding domain of BoCLO2 has been lost. All CLO5 genes lost their upstream N-terminal sequences (motif 6) of the calcium-binding domain (Fig. 2). Combined with expression data, these N-terminal upstream sequences (motif 7-4 for H-caleosins and motif 6 for L-caleosins), are probably critical for caleosin function.

Expression analysis of caleosins in the tested Brassicaceae plants

Functional differentiation of caleosin genes can be explained using differentially expressed gene (DEG) analysis of their orthologs and paralogs. For instance, for the CLO3 genes, BrCLO3-1, BrCLO3-2, BrCLO3-3 and BoCLO3-1, BoCLO3-2, BoCLO3-3, are paralogs, while BrCLO3-1 and BoCLO3-1, BrCLO3-2 and BoCLO3-2, and BoCLO3-3 and BrCLO3-3 are three pairs of orthologs. Among them, differences in ortholog expression were smaller than in paralog expression (Fig. 5). This may be due to the sub-genome dominance effect, which can conserve functions of a single gene while promoting the sub-functionalization or neo-functionalization of its paralogs (Wang et al. 2011). According to the current study, paralogs are the main cause for functional diversities among caleosins.

Expression patterns of *B. rapa* and *B. oleracea* orthologs are similar (Fig. 5). Obvious differences are detected among *B. napus* caleosin genes and their progenitor parents, except for BnCLO2. These distinctions may due to the necessity of gene expression changes during the process of environmental adaptations of the allopolyploid formed by natural polyploidization. These changes are results of many short-lived occurrences of genetic diversity happening in allopolyploid genomes (Song et al. 1995). According to the "additivity hypothesis", expression patterns of allopolyploids are the average of the expression in their parents (Higgins et al. 2012). However, the expression in *B. napus* caleosin genes and that in their parents did not fit the "additivity hypothesis", which is consistent with the fact that commonly expressed genes in diploid parents do not fit the "additivity hypothesis" in the allotetraploid (Higgins et al. 2012; Jiang et al. 2015). Thus the expression patterns of *B. napus* caleosin genes cannot be directly inferred from that of their parents.

Among several caleosin paralogs in Brassica plants, genes with higher expression levels are likely to play more important roles (Cheng et al. 2014). Thus, BrCLO3-3, BrCLO4-2, BoCLO3-2, BoCLO4-1, BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4 are worthy of further research, while the extremely low expression levels of BrCLO2-1, BrCLO2-2, BrCLO5, BnCLO2-1, BnCLO3-1, BnCLO4-5, BnCLO5-1, BnCLO5-2, and BnCLO8-3 might indicate that under normal growth conditions, they may play roles of less importance.

Functional analysis of important *B. napus* caleosins

The ABRERATCAL element (S000507) is detected in BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4 by PLACE analysis (see Supplemental Table 6). This element can respond to calcium signaling (Kaplan et al. 2006). PlantCARE analysis indicates there are several phytohormone repsonse-related elements in BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4. These genes contain integrated EF-hand domains, which show a consistent capacity for calcium-binding. Taken together, this suggests that calcium is involved in signal transduction processes enabled by these caleosins and subject to regulation by phytohormones. In addition, BnCLO1-2 and BnCLO1-3 contain an "RY-repeat" element, which can bind transcription factors within the B3-domain (Ezcurra et al. 2000). Transcription factors ABI3, FUS3 and LEC2, containing B3 domains, are associated with lipid metabolism (McGlew et al. 2015). The PlantCARE analysis shows that the RY-element is only found in BnCLO3-4, while the Skn-1 motif, which is related to lipid metabolism (Pang et al. 2014), is found in BnCLO1-2 and BnCLO1-3. This may be evidence that these caleosins have a relationship with triacylglycerol accumulation and decomposition/turnover.

BnCLO1-2 shows specific expression in seed or germinating seed and BnCLO1-3 exhibits seed-specific expression (Fig. 5). The seed is the most important tissue for triacylglycerol accumulation in *B. napus*; the decomposition of lipid droplets by proteases and lipases and the subsequent β oxidation of the fatty acids released, can provide energy for the germinating seed. Therefore, these two caleosins may play roles in triacylglycerol accumulation or turnover, or calcium-meidated signal transduction in seeds, while BnCLO1-2 may also play a role in calcium-mediated signal transduction during seed germination. BnCLO3-3 shows meristem-specific expression while BnCLO3-4 shows a uniform expression in different organs and their expression is also augmented under stress, suggesting a role for calcium-based signal transduction in stress responses in relevant tissues.

In summary, analyses of phylogeny, structural motifs, genome expansion, gene expression and functional analysis were studied. All 51 Brassica and *A. lyrata* caleosins, showing different motif structures and physicochemical properties, can be divided into two distinct families: H-family or L-family. Structural analysis indicated the intron phases are conserved among the majority of the Brassica caleosins and segmental duplication was identified as the main gene expansion form for them. Among them, BrCLO2 and BoCLO8-2 have lost their calcium binding domain, proline knot, potential phosphorylation sites and haem-binding site motifs, which, when combined with their low expression levels, suggests that these caleosins have likely lost their related functions. Others like BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4 may involve phytohormone-related responses because associated signature elements are found in their sequences. In addition, the 'RY-repeat' elements, which can be bound by transcription factors possessing the B3-domain, such as ABI3, FUS3 or LEC2, were discovered in BnCLO1-2 and BnCLO1-3. The expression data indicated that caleosins of high expression level, such as BrCLO3-3, BrCLO4-2, BoCLO3-2, BoCLO4-1, BnCLO1-2, BnCLO1-3, BnCLO3-3 and BnCLO3-4, are likely to play an important role as structural stabilizers of lipid droplets or are invoked in response to stress, both of which will require further study.

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

References

Allen J, Leach N, Ferguson S (2005) The histidine of the c-type cytochrome CXXCH haem-binding motif is essential for haem attachment by the *Escherichia coli* cytochrome *c* maturation (Ccm) apparatus. Biochem J 389:587–592

Arabidopsis genome initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796

Aubert Y, Vile D, Pervent M, Aldon D, Ranty B, Simonneau T, Vavasseur A, Galaud J-P (2010) RD20, a stress-inducible caleosin, participates in stomatal control, transpiration and drought tolerance in *Arabidopsis thaliana*. Plant Cell Physiol 51:1975–1987

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34:W369–W373

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci 107:18724–18728

Birchler JA, Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell 19:395–402

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res 13:137–144

Blée E, Flenet M, Boachon B, Fauconnier ML (2012) A non-canonical caleosin from Arabidopsis efficiently epoxidizes physiological unsaturated fatty acids with complete stereoselectivity. FEBS J 279:3981–3995

Blée E, Boachon B, Burcklen M, Le Guédard M, Hanano A, Heintz D, Ehlting J, Herrfurth C, Feussner I, Bessoule J-J (2014) The reductase activity of the Arabidopsis caleosin RESPONSIVE TO DESSICATION20 mediates gibberellin-dependent flowering time, abscisic acid sensitivity, and tolerance to oxidative stress. Plant Physiol 166:109–124

Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294:1351–1362

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4:10

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345:950–953

Chapman KD, Dyer JM, Mullen RT (2012) Biogenesis and functions of lipid droplets in plants thematic review series: lipid droplet synthesis and metabolism: from yeast to man. J Lipid Res 53:215–226

Chen JCF, Tsai CCY, Tzen JTC (1999) Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds. Plant Cell Physiol 40:1079–1086

Chen D, Chyan C, Jiang P, Chen C, Tzen JTC (2012) The same oleosin isoforms are present in oil bodies of rice embryo and aleurone layer while caleosin exists only in those of the embryo. Plant Physiol Biochem 60:18–24

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X (2011) BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol 11:136

Cheng F, Wu J, Wang X (2014) Genome triplication drove the diversification of Brassica plants. Hortic Res 1:14024

Ezcurra I, Wycliffe P, Nehlin L, Ellerström M, Rask L (2000) Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different *cis*-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. Plant J 24:57–66

Frandsen GI, Mundy J, Tzen JTC (2001) Oil bodies and their associated proteins, oleosin and caleosin. Physiol Plant 112:301–307

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Hanano A, Burcklen M, Flenet M, Ivancich A, Louwagie M, Garin J, Blée E (2006) Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif. J Biol Chem 281:33140–33151

Hernandez-Pinzon I, Patel K, Murphy DJ (2001) The *Brassica napus* calcium-binding protein, caleosin, has distinct endoplasmic reticulum-and lipid body-associated isoforms. Plant Physiol Biochem 39:615–622

Higgins J, Magusin A, Trick M, Fraser F, Bancroft I (2012) Use of mRNA-seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. BMC Genomics 13:247

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

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481

Hu L, Li S, Gao W (2013) Expression, divergence and evolution of the caleosin gene family in Brassica rapa. Arch Biol Sci 65:863–876

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31:1296–1297

Jiang J, Wang Y, Zhu B, Fang T, Fang Y, Wang Y (2015) Digital gene expression analysis of gene expression differences within Brassica diploids and allopolyploids. BMC Plant Biol 15:22

Jolivet P, Boulard C, Bellamy A, Larré C, Barre M, Rogniaux H, d'Andréa S, Chardot T, Nesi N (2009) Protein composition of oil bodies from mature *Brassica napus* seeds. Proteomics 9:3268–3284

Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240

Kaplan B, Davydov O, Knight H, Galon Y, Knight MR, Fluhr R, Fromm H (2006) Rapid transcriptome changes induced by cytosolic Ca^{2+} transients reveal ABRE-related sequences as Ca^{2+}-responsive cis elements in Arabidopsis. Plant Cell 18:2733–2748

Katavic V, Agrawal GK, Hajduch M, Harris SL, Thelen JJ (2006) Protein and lipid composition analysis of oil bodies from two *Brassica napus* cultivars. Proteomics 6:4586–4598

Kim YY, Jung KW, Yoo KS, Jeung JU, Shin JS (2011) A stress-responsive caleosin-like protein, AtCLO4, acts as a negative regulator of ABA responses in Arabidopsis. Plant Cell Physiol 52:874–884

Kyte J, Doolittle RF (1982) A simple mehod for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M (2012) The Arabidopsis information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40:D1202–D1210

Larkin MA, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Lee T, Bretaña NA, Lu C (2011) PlantPhos: using maximal dependence decomposition to identify plant phosphorylation sites with substrate site specificity. BMC Bioinform 12:261

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, DeBono A, Durrett TP (2013) Acyl-lipid metabolism. The Arabidopsis book/American Society of Plant Biologists 11

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930

Lomascolo A, Uzan-Boukhris E, Sigoillot J-C, Fine F (2012) Rapeseed and sunflower meal: a review on biotechnology status and challenges. Appl Microbiol Biotechnol 95:1105–1114

Lynch M (2002) Gene duplication and evolution. Science 297:945–947

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Ma H, Zhao J (2010) Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). J Exp Bot 61:2647–2668

Maher C, Stein L, Ware D (2006) Evolution of Arabidopsis microRNA families through duplication events. Genome Res 16:510–519

McGlew K, Shaw V, Zhang M, Kim RJ, Yang W, Shorrosh B, Suh MC, Ohlrogge J (2015) An annotated database of Arabidopsis mutants of acyl lipid metabolism. Plant Cell Rep 34:519–532

Murphy DJ (2009) Plant lipids: biology, utilisation and manipulation. Wiley, New York

Murphy DJ (2012) The dynamic roles of intracellular lipid droplets: from archaea to mammals. Protoplasma 249:541–585

Næsted H, Frandsen GI, Jauh G-Y, Hernandez-Pinzon I, Nielsen HB, Murphy DJ, Rogers JC, Mundy J (2000) Caleosins: Ca^{2+}-binding proteins associated with lipid bodies. Plant Mol Biol 44:463–476

Nagaharu U (1935) Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn J Bot 7:389–452

Pang S, Lynn DA, Lo JY, Paek J, Curran SP (2014) SKN-1 and Nrf2 couples proline catabolism with lipid metabolism during nutrient deprivation. Nat Commun 5:5048

Partridge M, Murphy DJ (2009) Roles of a membrane-bound caleosin and putative peroxygenase in biotic and abiotic stress responses in Arabidopsis. Plant Physiol Biochem 47:796–806

Poxleitner M, Rogers SW, Lacey Samuels A, Browse J, Rogers JC (2006) A role for caleosin in degradation of oil-body storage lipid during seed germination. Plant J 47:917–933

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435–3438

R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Shen Y, Xie J, Liu R, Ni X, Wang X, Li Z, Zhang M (2014) Genomic analysis and expression investigation of caleosin gene family in Arabidopsis. Biochem Biophys Res Commun 448:365–371

Song K, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci 92:7719–7723

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Tong C, Wang X, Yu J, Wu J, Li W, Huang J, Dong C, Hua W, Liu S (2013) Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. BMC Genomics 14:689

Toorop PE, Barroco RM, Engler G, Groot SP, Hilhorst HW (2005) Differentially expressed genes associated with dormancy or germination of *Arabidopsis thaliana* seeds. Planta 221:637–647

Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell 18:1348–1359

Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Yang Y, Lai K, Tai P, Li W (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. J Mol Evol 48:597–604

Yu J, Zhao M, Wang X, Tong C, Huang S, Tehrim S, Liu Y, Hua W, Liu S (2013) Bolbase: a comprehensive genomics database for *Brassica oleracea*. BMC Genomics 14:664