Insights into the caleosin family in Cyperus esculentus, an oil-rich tuber plant in Cyperaceae

Zhi Zou
zouzhi2008@126.com

Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences

Yuliang Chen
Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences

Huijian Huang
Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences

Chunqiang Li
Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences

Xiangdong Yang
Jilin Provincial Key Laboratory of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences

Jiaquan Huang
School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication) and College of Tropical Agriculture and Forestry, Hainan University

Yongguo Zhao
College of Biology and Food Engineering, Guangdong University of Petrochemical Technology

Research Article

Keywords: Vegetative tissue, Underground tuber, Phylogenomics, Orthologous analysis, Synteny analysis, Gene expansion, Expression divergence

Posted Date: February 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8567486/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Caleosins (CLOs), a class of structural proteins of lipid droplets (LDs), widely function in LD formation, stabilization, and degradation as well as in plant development and stress responses. However, their characterization in tigernut (*Cyperus esculentus* L., Cyperaceae), a rare example accumulating significant amounts of oil in underground tubers, is still in the infancy.

Results

In this study, we present a first genome-wide analysis of the *caleosin* family in tigernut. A number of eight members, which represent two previously defined clades (i.e., H and L), were identified from the tigernut genome, in stark contrast to only two present in the basal angiosperm *Amborella trichopoda*. Comparison of 193 *caleosin* genes from 31 representative plant species reveals lineage-specific expansion and functional diversification. Eight *CeCLO* genes belong to eight out of 11 orthogroups identified in this study, and extensive expansion in this species was contributed by whole-genome duplication (WGD), tandem, and dispersed duplications. 1:1 orthologous relationships observed between tigernut and its close relative *Cyperus rotundus* suggest that gene copies are not the contributor of high tuber oil accumulation in tigernut. Instead, most *CeCLO* genes was shown to express more than their orthologs in *C. rotundus*, implying species-specific activation in oil-bearing tigernut tubers. Correspondingly, expression of *CeCLO1*, *-2*, *-3*, and *− 7* was shown to positively correlate with oil accumulation during tuber development. Structure and expression divergence of paralogous pairs were also observed, and good examples are 1) *CeCLO7* and *− 8* that have gained one additional intron and 2) *CeCLO1* that has become the predominant isoform in oil-rich tubers.

Conclusions

These findings provide insights into the evolution, expression, and structural variation of *CeCLO* genes, which improve our knowledge on the mechanism of high oil accumulation in tigernut tubers.

Background

Plant seeds store triacylglycerols (TAGs) in discrete spherical organelles called oil bodies or lipid droplets (LDs), which are coated by a monolayer of phospholipids embedded with abundant oleosins (OLEs) and several minor structural proteins [1]. Caleosins (CLOs) are a class of minor LD proteins that are capable of binding to calcium ion (Ca^{2+}) [2, 3, 4]. Since the first *caleosin* gene (i.e., *responsive to desiccation 20* (*RD20*)) that was identified in arabidopsis (*Arabidopsis thaliana*) [5, 6], its homologs have been widely found in land plants as well as in some fungi and green algae [2, 7, 8, 9, 10, 11, 12]. Caleosins exhibit some similarities to oleosins, which feature a long central hydrophobic hairpin of about 72 residues ($X_{30}PX_5SPX_3PX_{30}$) [1, 13, 14]. Like oleosins, caleosins are composed of three characteristic structural domains, i.e., an N-terminal hydrophilic EF-hand calcium binding domain, a central hydrophobic anchoring domain, and a C-terminal hydrophilic phosphorylation domain [6, 7]. The central hydrophobic domain, which is responsible for anchorage in LDs, comprises an amphipathic α-helix followed by a pair of anti-parallel β-strands that are connected by a proline-knot motif (PX_3PSX_3P) [7, 15, 16]. Both N- and C-terminals are more likely to expose to the cytoplasm and are involved in environmental adaptation [7, 8]. Though caleosins were first recognized as minor proteins functioning in LD stabilization and degradation in seeds, they were later proven to have catalytic functions as peroxygenases (PXGs, EC 1.11.2.3), whose activity is dependent on a heme group coordinated by the two invariant His residues located in N- and C-terminal regions [3, 4, 7, 17, 18, 19]. Hence, they are also called peroxygenases [4, 11, 20, 21]. Unlike oleosins, caleosins have additional intracellular locations apart from LDs, e.g., the envelope of chloroplasts, endoplasmic reticulum (ER), plasmalemma, vacuole, and nucleus, and they were also found to be abundant in a range of non-

seed plant tissues such as callus, shoot, hypocotyl, root, cotyledon, leaf, megagametophyte, male flower, pollen, stamen, ovule, silique, fruit, and phloem sap [8, 20, 22, 23, 24, 25, 26, 27, 28, 29].

In land plants, caleosins comprise a small family with up to 10 members [9, 10, 11, 15, 16, 29, 30, 31, 32]. Generally, there are two major types present in each species, i.e., high (H) and low (L) molecular weights (MW), where the H-forms contain an extra N-terminal insertion of about 30 residues relative to L-forms [15, 16]. In contrast to only two members present in castor bean (*Ricinus communis*, *RcCLO-1* and − *2*) and flax (*Linum usitatissimum*, *LuCLO-1* and − *2*) [30], a number of eight *caleosin* genes were described in arabidopsis, including four H-forms (i.e., *AtCLO1*, *-2*, *-3*, and − *8*) and four L-forms (i.e., *AtCLO4*, *-5*, *-6*, and − *7*), and the family expansion was shown to be contributed by both local and segmental duplications [15, 29]. Expression analyses showed that *RcCLO-2* and *LuCLO-2* are preferentially expressed in seeds, whereas *RcCLO-1* and *LuCLO-1* are constitutively expressed, including seeds, cotyledons, and young leaves [29, 30]. Similar expression profiles were also observed for *AtCLO* genes, where *AtCLO3* and − *4* are expressed in various organs and others seem to be organ-specific, e.g., *AtCLO1* and − *2* in seeds and *AtCLO5* in buds [15]. Functional analyses revealed that caleosins are not only involved in LD formation, stabilization, and degradation, but also are associated with various plant development and stress responses [2, 7, 8, 12, 15, 25, 30, 31, 32]. In rice (*Oryza sativa*), *OsCLO5* was initially named *OsEFA27* (EF-hand, ABA responsive, 27 kilodalton (kDa)), which could be induced by ABA treatment, salt and drought stresses [2]. In arabidopsis, four seed-expressed *caleosin* genes (i.e., *AtCLO1*, *-2*, *-4*, and − *6*) were proven to redundantly function in oil accumulation and embryo development [34]. Interestingly, a high number of *caleosin* genes were shown to be regulated by various abiotic and biotic stresses, e.g., *OsCLO5*/*EFA27* in rice [2], *AtCLO3*/*PXG3*/*RD20* in arabidopsis [8], *Sopl* in sesame (*Sesamum indicum*) [7], *HvCLO1* and − *2* in barley (*Hordeum vulgare*) [33], *BnCLO1-1* in rapeseed (*Brassica napus*) [35], and *OeCLO3* in olive (*Olea europaea*) [25].

Tigernut (*Cyperus esculentus* L. *var. sativus* Boeckeler), which is also known as chufa and yellow nutsedge, is an unconventional oil crop uniquely accumulating significant amounts of TAGs in its underground tubers [36, 37]. It resides in the Cyperaceae family, which is native to Africa [38]. Tigernut was initially domesticated in ancient Egypt for the nutrient-rich tubers, e.g., 25–45% starch, 20–35% oil, 15–20% sugars, 5–10% protein, 8–10% dietary fiber, 0.08–0.14 g/kg vitamin C and E, high levels of minerals (e.g., calcium (Ca), phosphorus (P), potassium (K), and magnesium (Mg)), and various active substances (e.g., flavone, alkaloid, tannin, terpenoid, saponin, and phytosterol) [39, 40]. Recently, much attention has been attracted for its unique feature of high oil accumulation in the non-seed storage organ [41, 42, 43, 44, 45, 46, 47, 48, 49], and the disclosure of the mechanisms behind shall facilitate oil production in plant vegetative tissues via genetic engineering [50]. According to a previous proteomic analysis, like oilseeds, the LDs of tigernut tubers are mainly coated by oleosins and caleosins [51]. Thus far, genome-wide characterization of the *oleosin* family was reported, and six members present in tigernut are equal to those found in other Cyperaceae plants such as purple nutsedge (*C. rotundus*) and *Rhynchospora breviuscula* [45]. Interestingly, the majority of *CeOLE* genes were shown to exhibit a tuber-preferential expression pattern with mRNA/protein abundances positively correlating with oil accumulation during tuber development [45]. Transcriptome-based identification of *caleosin* genes was also reported [52], however, 21 homologs described in that study are considerably more than eight detected from our full-length transcriptome, which was constructed via single-molecule real-time (SMRT) sequencing of various tissues such as shoot apex, leaf, sheath, root, rhizome, tuber, flower, and floral stem [43]. Considering the inherent drawback of incorrect *de novo* assembly from short reads, a genome-scale identification may provide a more precise view of the *caleosin* family in this special species.

In this study, a genome-wide analysis of the *CeCLO* family was conducted, which include sequence characteristics, gene structures, chromosome localizations, evolutionary relationships, and expression profiles. Significantly, comparison of 193 *caleosin* genes identified from 31 representative plant species reveals lineage-specific expansion and functional diversification in tigernut. In contrast to only two members present in the basal angiosperm *Amborella trichopoda*, a high number of eight *CeCLO* family genes were shown to arise from whole-genome duplication (WGD), tandem, and dispersed duplications, whereas tuber-specific activation of certain members appears to be species-specific. These findings highlight the evolution patterns of the *CeCLO* family, which provide valuable information for further functional analysis.

Results

Identification, gene localization, and duplication event analysis of eight caleosin genes in tigernut

As shown in Table 1, HMMER-based homologue search resulted in eight *caleosin* genes from the tigernut genome [53]. Though the overall sequence similarity of their deduced peptides varies from 44.32% to 78.57% (**Additional file 1**), all of them were shown to possess a single caleosin domain (under the Pfam accession number of PF05042) that is specific to this family (Table 1). Notably, the amounts of the *CeCLO* family are somewhat bigger than six members as described for the *CeOLE* family [45]. Compared with CeOLEs with an average of approximately 150 residues [45], the peptide length of CeCLOs is relatively longer, varying from 198 (i.e., CeCLO6) to 261 (i.e., CeCLO5) residues with an average of about 234 residues, which result in the higher MW of 21.82–29.93 kDa, i.e., 26.42 vs 15.52 kDa (Table 1). Interestingly, different patterns of isoelectric point (pI) and grand average of hydropathicity (GRAVY) were also observed between CeCLOs and CeOLEs as described in papaya (*Carica papaya*) [29]. Whereas only 62.50% of CeCLOs (i.e., CeCLO4–8) exhibit the alkaline feature like CeOLEs, others (i.e., CeCLO1–3) harbor the pI values of less than 7, varying from 5.91 (i.e., CeCLO2) to 6.63 (i.e., CeCLO3) (Table 1). In contrast to the highly hydrophobic feature of CeOLEs [45], all CeCLOs have the GRAVY value of less than 0, varying from − 0.593 to -0.117 (Table 1). The overall hydrophilic feature of CeCLOs was also supported by the ProtScale analysis. As shown in Fig. 1a, similar hydropathicity scales were observed for all eight CeCLOs. Whereas the central domain is hydrophobic, N- and C-terminal regions are usually hydrophilic (Fig. 1a). Except for CeCLO6, one to two transmembrane helixes (TMH) were identified (Table 1).

Table 1
Caleosin genes identified in *C. esculentus*. (*AA* amino acid, *Ce C. esculentus*, *CLO* caleosin, *GRAVY* grand average of hydropathicity, *kDa* kilodalton, *MW* molecular weight, *pI* isoelectric point, *Scf* scaffold, *TMH* transmembrane helix)

Gene name	Locus ID	Position	AA	MW (kDa)	pI	GRAVY	TMH	Caleosin location
CeCLO1	CESC_00080	Scf30:3622141..3623943(-)	251	28.96	6.50	-0.366	104..123	74..240
CeCLO2	CESC_16572	Scf47:433600..435126(-)	242	27.87	5.91	-0.410	94..117	63..231
CeCLO3	CESC_16571	Scf47:436090..438271(-)	239	27.38	6.63	-0.228	88..107,184–201	59..225
CeCLO4	CESC_18483	Scf13:1911499..1913058(-)	244	27.63	9.04	-0.293	88..110,184..201	59..226
CeCLO5	CESC_18484	Scf13:1915073..1916809(-)	261	29.93	8.78	-0.324	97..116	68..235
CeCLO6	CESC_02525	Scf21:760234..762404(-)	198	21.82	7.81	-0.593	-	21..188
CeCLO7	CESC_02526	Scf21:770229..771732(-)	217	23.57	8.79	-0.363	7..24,70..92	38..206
CeCLO8	CESC_02527	Scf21:775539..780932(-)	220	24.24	9.03	-0.117	6..25	43..210

The amino acid (AA) composition of CeCLOs was further calculated and compared with that of CeOLEs. Except for the absence of the Cys in CeCLO6–8, other members contain all 20 residues (Fig. 1b), usually rich in Leu, Gly, and Ala as observed in CeOLEs (**Additional file 2a**). Interestingly, the cluster analysis revealed distinct patterns of AA composition between CeCLOs and CeOLEs, where proteins were grouped in family. Compared with CeOLEs, CeCLOs contain more hydrophilic but less hydrophobic residues (**Additional file 2a**), which contribute to their hydrophilic feature. The secondary structure of CeCLOs was also shown to differ from that of CeOLEs, where CeCLOs harbor more random coils but less alpha helixes and were clustered together (Fig. 1c and **Additional file 2b**).

Sequence alignment of CeCLOs is shown in Fig. 1d. Except for the conserved caleosin domain, N- and C-terminals are highly diverse. Though the usual proline-knot pattern PX_3PSX_3P was observed in CeCLO2–6, the PX_4S/QSX_3P variant was identified in CeCLO7 and − 8. Moreover, the PX_3LNX_3P variant was found in CeCLO1. Prior to the proline-knot are the amphipathic α-helix and the Ca^{2+} binding motif. Two His residues, which are essential for the PXG activity [4], were found in all CeCLOs, residing in front of the Ca^{2+} binding motif and somewhat behind the proline-knot motif, respectively (Fig. 1d). Compared with CeCLO6–8, CeCLO1–5 harbor an H-form insertion as observed in arabidopsis [15], which contributes to the longer peptide

size and higher MW (Table 1 and Fig. 1d). As shown in **Additional file 3**, a high number of phosphorylation sites were identified for each sequence. Among them, four are shared by most sequences, i.e., one for Tyr kinase and three for casein kinase II (CK II) (Fig. 1d), implying their posttranslational regulation. Interestingly, despite few Cys residues present in CeCLOs, one is shared by CeCLO1–5 (Fig. 1d), which was also described for SopI in sesame [7].

In contrast to the distribution of six *CeOLE* genes over six scaffolds (Scfs) [45], gene localization showed that eight *CeCLO* genes are unevenly spread across four scaffolds, and three hotspots for tandem duplication were detected, i.e., *CeCLO2*/*-3* on Scf47, *CeCLO4*/*-5* on Scf13, and *CeCLO6*/*-7*/*-8* on Scf21, respectively. Further synteny analysis revealed that *CeCLO1* and − *2* are located within syntenic blocks, implying their derivation from WGD. On the contrary, *CeCLO2*/*-4* and *CeCLO4*/*-6* were characterized as dispersed repeats (Fig. 1e).

Phylogenetic and comparative analyses revealed species-specific expansion of the caleosin family among tigernut, rice, and arabidopsis

To uncover the evolutionary relationships of *CeCLO* genes, homologs were also identified from arabidopsis and rice, two model plants for dicots and monocots, respectively. In accordance with previous studies [15, 27], eight family members were obtained from updated genomes of both species (**Additional file 4**). Interestingly, despite harboring the same family amounts, phylogenetic analysis of caleosins in tigernut, rice, and arabidopsis revealed species-specific expansion. As shown in Fig. 2a, these proteins were clustered into two main clades, corresponding to H and L as described in arabidopsis [15]. Moreover, these two clades could be divided into two and three groups, respectively. Among them, Group Ia includes CeCLO1–3, OsCLO4, OsCLO5, AtCLO1–3, and AtCLO8; Group Ib includes CeCLO4, CeCLO5 and OsCLO3; Group IIa is specific to arabidopsis, including AtCLO4–7; Group IIb is specific to tigernut, including CeCLO6–8; and Group IIc is specific to rice, including OsCLO1–2, and OsCLO6–8. The results are largely in accordance with sequence similarities and duplication event analyses. As shown in **Additional files 1** and **5**, *OsCLO1*/*-2*, *OsCLO4*/*-5*, and *OsCLO6*/*-7*, which share 72.65%, 67.48%, and 95.56% sequence similarities at the protein level, were defined as proximal repeats; *OsCLO7*/*-8* (53.41%), *AtCLO4*/*-6* (85.71%), and *AtCLO5*/*-7* (62.66%) were characterized as tandem repeats; *OsCLO1*/*-6* (57.33%), *AtCLO1*/*-2* (87.76%), and *AtCLO5*/*-6* (66.67%) were characterized as WGD repeats, respectively; *OsCLO3*/*-4* (54.10%), *OsCLO4*/*-6* (39.74%), *AtCLO1*/*-3* (72.47%), and *AtCLO1*/*-4* (47.35%) were characterized as dispersed repeats.

To learn more about structural variation, exon-intron structures and conserved motifs were further compared. As shown in Fig. 2b, the most majority of *caleosin* genes present in three species feature five introns with the conserved phase pattern of 1, 1, 0, 2, and 2. Nevertheless, gain or loss of certain introns was observed in *CeCLO7*, *CeCLO8*, *AtCLO7*, and *AtCLO8*. In contrast to the loss of the fifth intron in *AtCLO8*, *CeCLO7* and *CeCLO8* were shown to have gained one more intron (in phase 2) at the 5' end, which is different from the phases of 2, 0, 0, 0, 2, and 2 as observed in *AtCLO7*, implying independent origin. Ten motifs identified using MEME are shown in Fig. 2c. Whereas Motifs 1–5 are widely present, others appear to be sequence-specific. Motifs 1, 2, and 4 were characterized as the conserved caleosin domain; Motif 3 was characterized as the H-form insertion; and Motif 5 was characterized as the C-terminal phosphorylation sites (Fig. 2d). Notably, Motifs 2 and 4 are absent from AtCLO7, and Motif 1 is absent from AtCLO7 and − 8, implying their divergence. Whereas Motif 3 was found in all H-forms, Motif 10, which was also characterized as the C-terminal of the H-form insertion, are limited to CeCLO1, CeCLO2, OsCLO4, and AtCLO2. Motifs 6 and 7, which are confined to OsCLO6 and − 7, are located at the N- and C-terminals, respectively. Motif 8, which is confined to AtCLO1 and − 2, is located at the C-terminal. Motif 9, which is confined to AtCLO5 and − 7, is located at the N-terminal (Fig. 2c).

Characterization of caleosin genes from representative plant species and insights into lineage-specific family evolution in Poales

Above phylogenetic analysis suggests the divergence of Clades H and L sometime before monocot-eudicot split. However, several issues need to be further resolved: 1) the tandem duplications identified in tigernut are species or lineage-specific; 2) the exact time of the WGD that gave birth to *CeCLO2*; 3) the exact origin of *CeCLO4* and − *6*; 4) the exact time of gain of the additional intron as observed in *CeCLO7* and − *8*. For the purposes, homologs were also identified from representative plant

species, which include the basal angiosperm *A. trichopoda* (an Amborellaceae plant in Amborellales) and three early diverged monocots that didn't experience the τ WGD shared by core monocots, i.e., American sweet flag (*Acorus americanus*, an Acoraceae plant in Acorales), eelgrass (*Zostera marina*, a Zosteraceae plant in Alismatales), and duckweed (*Spirodela polyrhiza*, an Araceae plant in Alismatales). After discarding pseudogenes and gene fragments, as shown in **Additional file 4**, a total of 169 members were further identified. Notably, a single member for each clade was not only found in two out of three tested core eudicots, i.e., grapevine (*Vitis vinifera*) and castor bean, but also in *A. trichopoda* and eelgrass, supporting early divergence of Clades H and L in the angiosperm ancestor. By contrast, equal amounts of eight members were identified in purple nutsedge, a close relative of tigernut (**Additional file 4**).

To infer lineage-specific gene expansion and contraction, species-specific duplication events were further examined and orthologs across different species were identified using Orthofinder. As shown in Fig. 3, a total of 11 orthogroups were obtained and eight *CeCLO* genes belong to H1, H2, H3, H4, H5, L1, L2, and L3. Among them, H1 and L1 are shared by all species examined in this study, whereas others appear to be lineage-specific: H4 is widely present in core monocots, including garden asparagus (*Asparagus officinalis*), apostasia (*Dendrobium catenatum*), and greater yam (*Dioscorea alata*), implying its early origin that may arise from the τ WGD; H2 is present in most Poales plants and is usually located within syntenic blocks with H1, implying its possible origin along with the Poales-specific σ WGD; H3, which is limited to Cyperaceae and Juncaceae species, was characterized as the tandem repeat of H2; H5 is confined to three out of four examined *Cyperus* species, i.e., tigernut, purple nutsedge, and *C. iria*, where further expansion via WGD was also observed in the latter; L2, which is widely present in Cyperaceae plants with the exception of *C. littledalei*, was characterized as the tandem repeat of L1; L3, which is also Cyperaceae-specific, was characterized as the tandem repeat of L2 in tigernut, purple nutsedge, *C. iria*, *C. fuscus*, and *R. breviuscula*, but the tandem repeat of L1 in *C. littledalei*; L4 and L5, which is Poaceae-specific, was characterized as the tandem and WGD (ρ) repeats of L1, respectively; L6, which is also Poaceae-specific, was characterized as the tandem repeat of L5 (Fig. 3 and **Additional file 4**).

Inter-specific syntenic analyses were further conducted and results showed that *CeCLO1*, *CeCLO2/-3*, *CeCLO4/-5*, *CeCLO6/-7/-8* have syntelogs in at least one of examined species. Significantly, as shown in Fig. 4a, 1:1 and 2:2 relationships were observed between tigernut and purple nutsedge/*R. breviuscula*/*C. littledalei*, implying that the WGD event (i.e., *CeCLO1/-2*) detected in tigernut is shared by these species. Moreover, 1:1, 1:2, and 2:4 relationships were observed between tigernut and *C. iria*, supporting species-specific expansion via WGD in *C. iria*, followed by group-specific gene loss in H3 and L3. Actually, their gene fragments could be detected in the *C. iria* genome. In accordance with orthologous analysis, no syntelog was identified for *CeCLO4/-5* in *C. littledalei*, implying species-specific contraction. By contrast, their syntelogs were not only identified in pineapple (*Ananas comosus*), *Joinvillea ascendens* (Fig. 4b), but also in oil palm (*Elaeis guineensis*), greater yam, duckweed (Fig. 4c), and rice (Fig. 4d). Notably, *CeCLO4/-5*, *CeCLO6/-7/-8*, *EgCLO2*, and *EgCLO3/-4/-5/-6* were shown to locate within syntenic blocks (Fig. 4c). Since *EgCLO3* was characterized as a recent repeat of *EgCLO2* from the Arecaceae-specific p WGD (**Additional file 4**), WGD-derivation of *CeCLO4/-6* followed by lineage-specific transposition could be speculated. Collectively, these results imply early diversification of the *caleosin* family before the radiation of angiosperms followed by lineage-specific expansion even in early diverged angiosperms.

The sequence identities and evolutionary rates of *Ce/CrCLO* genes as well as duplicate pairs identified in tigernut are summarized in Table 2. As expected, high sequence identities of 89.30–98.50% and small Ks (synonymous substitution rate) values of 0.1048–0.1810 were observed between orthologs of tigernut and purple nutsedge. By contrast, only 43.90–68.40% sequence identities were found between seven duplicate pairs identified in tigernut. Moreover, except for high sequence divergence between *CeCLO4* and − *6*, the Ks values of other pairs were shown to vary from 0.9012 (*CeCLO4* and − *5*) to 3.7504 (*CeCLO6* and − *7*) (Table 2), implying distinct time of their birth. Given that the Ka (nonsynonymous substitution rate)/Ks values of all duplicate pairs identified in tigernut are less than one (Table 2), a role of purifying selection during their evolution could be speculated.

Table 2
Sequence identities and evolutionary rates of duplicate pairs identified in tigernut as well as orthologs between *C. esculentus* and *C. rotundus*. (*Ce C. esculentus*, *CLO* caleosin, *Cr C. rotundus*, *Ka* nonsynonymous substitution rate, *Ks* synonymous substitution rate)

Species	Gene 1	Gene 2	Identity (%)	Ka	Ks	Ka/Ks
C. esculentus-*C. esculentus*	*CeCLO1*	*CeCLO2*	66.50	0.2346	2.4592	0.0944
	CeCLO2	*CeCLO3*	62.90	0.3019	2.4057	0.1255
	CeCLO2	*CeCLO5*	54.10	0.5105	2.1566	0.2367
	CeCLO4	*CeCLO5*	68.40	0.2376	0.9012	0.2637
	CeCLO4	*CeCLO6*	43.90	0.5190	-	-
	CeCLO6	*CeCLO7*	58.50	0.3064	3.7504	0.0817
	CeCLO7	*CeCLO8*	62.30	0.3924	1.1637	0.3372
C. esculentus-*C. rotundus*	*CeCLO1*	*CrCLO1*	94.70	0.0226	0.1810	0.1246
	CeCLO2	*CrCLO2*	98.50	0.0127	0.1723	0.0736
	CeCLO3	*CrCLO3*	96.80	0.0073	0.1225	0.0598
	CeCLO4	*CrCLO4*	93.70	0.0419	0.1522	0.2751
	CeCLO5	*CrCLO5*	89.30	0.0169	0.1696	0.0995
	CeCLO6	*CrCLO6*	95.80	0.0132	0.1407	0.0938
	CeCLO7	*CrCLO7*	95.30	0.0266	0.1048	0.2536
	CeCLO8	*CrCLO8*	95.80	0.0160	0.1203	0.1333

Caleosin genes in tigernut exhibited distinct expression patterns

To provide a global view of expression profiles of *CeCOL* genes, RNA-seq data of various tissues were first examined, i.e., young leaf, mature leaf, sheath, shoot apex, root, rhizome, young tuber (collected at 40 days after sowing (DAS)), middle tuber (collected at 85 DAS), and mature tuber (collected at 120 DAS). As shown in Fig. 5, FPKM (fragments per kilobase of exon per million fragments mapped) values exceeding 1 in at least one tested samples were observed for five out of eight *CeCOL* genes, i.e., *CeCLO1*–*3*, *-6*, and *− 7*. Among them, *CeCLO3*, *-6*, and *− 7* were constitutively expressed, whereas *CeCLO1* and *− 2* were predominantly expressed in tubers, especially in middle and mature tubers. In accordance with the high oil accumulation, total *CeCLO* transcripts were most abundant in middle and mature tubers, which were clustered together and whose transcripts were 12.41 and 13.90 folds more than those in young tubers. Followed were roots, young tubers, shoot apexes, sheaths, and rhizomes, whose transcripts were 2.20 to 5.98 folds more than those in mature and young leaves (Fig. 5a). When *CeOLE* genes were included for comparison, three tuber stages were clustered together; two most abundant genes *CeOLE2* and *CeOLE5* formed one group; *CeCLO1* was clustered with *CeOLE4*, *CeOLE6*, and *CeOLE1*; *CeCLO2* was clustered with *CeOLE3*; and *CeCLO3*/*-6*/*-7* and *CeCLO4*/*-5*/*-8* formed two independent groups (**Additional file 6a**). Correspondingly, correlation analysis showed that *CeCLO1* transcripts were positively related with those of *CeOLE* genes, with the *r* value varying from 0.458 (i.e., *CeOLE3*) to 0.913 (i.e., *CeOLE2*). Expression of *CeCLO1* was also positively correlated with *CeCLO3* ($r = 0.917$), which exhibited the *r* values of 0.390 (i.e., *CeOLE1*)–0.843 (i.e., *CeOLE2*) (**Additional file 6b**).

Comparison revealed apparent expression divergence of caleosin genes between tigernut and purple nutsedge

Our previous study showed that *oleosin* genes between tigernut and purple nutsedge underwent apparent expression divergence during tuber development, and species-specific activation is in accordance with high oil accumulation in tigernut tubers [45]. To address whether similar cases are also present for *CeCLO* genes, their expression profiles over three

representative tuber stages were compared, i.e., 20, 50, and 90 days after tuber initiation (DAI). As shown in Fig. 5b, expression was detected for five *CeCLO* and four *CrCLO* genes, i.e., *CeCLO1*, *CeCLO2*, *CeCLO3*, *CeCLO6*, *CeCLO7*, *CrCLO1*, *CrCLO3*, *CrCLO6*, and *CrCLO7*. Among them, *CeCLO1* was mostly expressed, followed by *CeCLO3* and *CrCLO6*, whereas *CrCLO3* was lowly expressed. Notably, though most *CeCLO* genes were expressed more than their orthologs in purple nutsedge, *CrCLO6* represents the single one that was expressed more than its ortholog *CeCLO6* (Fig. 5b). The result implies that *CeCLO1*, *-2*, *-3*, and *− 7* may contribute to tuber oil accumulation in tigernut.

CeCLO transcripts were positively correlated with TAG accumulation during tuber development

According to our previous study, the TAG content of Reyan3 gradually increased along with tuber development, from 5.09% at 5 DAI to 34.29% at 35 DAI when tubers have completely matured [45]. To uncover the correlation between gene expression and TAG accumulation during tuber development, five tuber-expressed *CeCLO* genes (i.e., *CeCLO1*, *-2*, *-3*, *-6*, and *− 7*) and five representative stages were selected for qRT-PCR analysis, i.e., S1 (5 DAI), S2 (10 DAI), S3 (20 DAI), S4 (25 DAI), and S5 (35 DAI). As shown in Fig. 6, gradual promotion of transcripts during tuber development was observed for *CeCLO1*, *-2*, *-3*, and *− 7*, whereas *CeCLO6* was highly expressed in three early stages, followed by significant downregulation at two latter stages. Further correlation analysis showed that the transcripts of *CeCLO1*, *-2*, *-3*, and *− 7* were positively correlated with TAG accumulation, with the *r* value varying from 0.837 (i.e., *CeCLO1*) to 0.964 (i.e., *CeCLO2*), whereas a negative correlation with *r* = 0.933 was observed for *CeCLO6* (Fig. 6), supporting the result obtained from above transcriptional profiling.

Discussion

As a rare example accumulating high levels of oil in the nutrient storage tubers, tigernut is emerging as an ideal model to study lipid regulation mechanisms beyond oilseeds [36, 42, 45, 46, 47, 48, 49]. Like oleosins, caleosins also feature the 12-residue proline-knot motif, a key component of the family-specific caleosin domain [11, 29, 54]. A crucial role of caleosins in LD formation and oil accumulation prompted us to study this special gene family in tigernut.

In the current study, the first genome-wide identification and characterization of the *CeCLO* gene family was reported. A number of eight family members identified from the tigernut genome are considerably less than 21 transcripts obtained from a *de novo* transcriptome assembly [52], implying the necessity of a genome-based identification. Interestingly, the family amounts present in this species are equal to those described in two model plants arabidopsis and rice [15, 27], but considerably more than two found in *A. trichopoda*, castor bean, flax, and cucumber (*Cucumis sativus*) [11, 29, 30], implying lineage and/or species-specific family expansion during angiosperm radiation.

Further phylogenetic analysis assigned eight *CeCLO* genes into two evolutionary groups as defined in arabidopsis [15], i.e., H (5) and L (3), where the H clade features a high MW with the so-called H insertion at the N-terminal (Fig. 1d). Interestingly, the group composition observed in tigernut is highly distinct from rice (H, 3; L, 5) and arabidopsis (H, 4; L, 4), in contrast to the basal angiosperm *A. trichopoda* harboring a single member for each group [29]. Extensive expansion in tigernut was shown to be resulted from WGD, tandem, and dispersed duplications (Fig. 1e). By contrast, expansion in rice was contributed by WGD, tandem, proximal, and dispersed duplications, whereas in arabidopsis, expansion was contributed by WGD, tandem, transposed, and dispersed duplications (**Additional file 5**). The pattern is also different from the *CeOLE* family, whose expansion was shown to be contributed by WGD and dispersed duplication [45]. The results reflect the occurrence of more than one WGD in these species after monocot-eudicot divergence. According to comparative genomics analyses, arabidopsis in the eudicot clade experienced three additional WGDs known as γ, β, and α in sequence [55]; similarly, rice in the monocot clade experienced three additional WGDs known as τ, σ, and ρ, where τ and σ WGDs are shared by tigernut [48, 56, 57, 58]. Correspondingly, four WGD repeats identified in arabidopsis, i.e., *AtCLO1*/*-2* and *AtCLO5*/*-6*, were shown to arise from the α WGD that is specific to the Brassicaceae family [15, 29], whereas *OsCLO1* and *− 6* were derived from the ρ WGD that is specific to the Poaceae family (**Additional file 5**). By contrast, considering the absence of a recent WGD in tigernut, two identified WGD repeats, i.e., *CeCLO1* and *− 2*, are more likely to arise from the σ WGD that is shared by all Poales plants [56].

To confirm the hypothesis and gain insights into the origin and evolution of *CeCLO* genes, approximately 200 homologs from 31 representative plant species were included for comparison, which resulted in 11 orthogroups named H1, H2, H3, H4, H5, L1, L2, L3, L4, L5, and L6. These species cover a wide range of 17 plant families, i.e., Amborellaceae (*A. trichopoda*), Vitaceae (grapevine), Euphorbiaceae (castor bean), Brassicaceae (arabidopsis), Acoraceae (American sweet flag), Zosteraceae (eelgrass), Araceae (duckweed), Asparagaceae (garden asparagus), Orchidaceae (apostasia), Dioscoreaceae (greater yam), Arecaceae (oil palm), Bromeliaceae (pineapple), Typhaceae (*Sparganium stoloniferum*), Cyperaceae (tigernut, purple nutsedge, *C. iria*, *C. fuscus*, *Bolboschoenus planiculmis*, *Schoenoplectus tabernaemontani*, *C. littledalei*, *R. breviuscula*, and *Eleocharis parvula*), Juncaceae (*Juncus effusus* and *J. inflexus*), Joinvilleaceae (*J. ascendens*), and Poaceae (*Pharus latifolius*, rice, barley, *Brachypodium distachyon*, *Sorghum bicolor*, and *Setaria italica*). In accordance with previous studies [11, 29], our data suggest that the *caleosin* family has diverged into Clades H and L in the last common ancestor of angiosperms, which is reserved as a single copy for each clade in some eudicots such as grapevine and castor bean. By contrast, extensive expansion via both local and large-scale duplications was frequently observed in Poales, and *CeCLO* genes belong to eight orthogroups, i.e., H1–5 and L1–3. Based on the results from orthologous, syntenic, and evolutionary rate analyses, a possible evolution route in tigernut is proposed as follows: the wide present H4 (e.g., *CeCLO4*) first gave birth to H2 (e.g., *CeCLO2*) via dispersed duplication sometime before monocot radiation, which may be along with the τ WGD followed by gene transposition; then, H2 produced H1 (e.g., *CeCLO1*) via the σ WGD in Poales, though H1 is most conserved; the last common ancestor of Cyperaceae and Juncaceae generated H3 (e.g., *CeCLO3*) via tandem duplication (unequal crossover) sometime after the split with other families within Poales; in Cyperaceae, L1 (e.g., *CeCLO6*) first brought forth L2 (e.g., *CeCLO7*) via tandem duplication sometime after the split with Juncaceae, which further gained an additional intron; subsequently, L2 produced L3 (e.g., *CeCLO8*) via tandem duplication; in *Cyperus*, the last common ancestor of tigernut, purple nutsedge, and *C. iria* generated H5 (e.g., *CeCLO5*) via tandem duplication sometime after the split with *C. fuscus*.

Interestingly, despite a very close genetic relationship of tigernut and purple nutsedge, purple nutsedge only accumulates less than 2.5% of oil in its tubers, in contrast to up to 35% in tigernut [36, 42, 45]. 1:1 orthologous relationships of *caleosin* genes observed between these two species suggest that all duplicate pairs identified in tigernut are lineage-specific and gene copies are not the contributor of high oil accumulation in tigernut tubers. Similar results were also reported for the *oleosin* family, which is composed of six members in both tigernut and purple nutsedge [45]. Compared with tuber-preferential expression of most *CeOLE* genes [45], eight *CeCLO* genes was shown to exhibit distinct expression patterns with constitutive expression of *CeCLO3*, *-6*, and *− 7*, tuber-predominant expression of *CeCLO1* and *− 2*, and rare expression of *CeCLO4*, *-5*, and *− 8* (Fig. 5a). Moreover, though the *CeCLO* family contains two more members than the *CeOLE* family, their total transcripts in tubers were shown to be considerably less abundant (**Additional file 6a**). Nevertheless, the transcript level of *CeCLO1* ranks only after *CeOLE5* and *CeOLE2*, two most abundant *CeOLE* genes in tigernut tubers (**Additional file 6a**). A putative role of *CeCLO1* and *− 3* in tuber oil accumulation was first inferred from Pearson's correlation analysis, which was shown to exhibit the high *r* value up to 0.9 with *CeOLE2* (**Additional file 6b**). Correspondingly, our qRT-PCR analysis showed that transcription of *CeCLO1*, *-2*, *-3*, and *− 7* in five representative tuber stages was positively correlated with oil accumulation during tuber development (Fig. 6). Last but most important, our comparative transcriptome analysis revealed that *CeCLO1*, *-2*, *-3*, and *− 7* were expressed considerably more than their orthologs in purple nutsedge (Fig. 5b), implying species but not tuber-specific activation for oil accumulation.

As discussed above, unlike the *oleosin* family, tandem duplication was shown to play a crucial in the expansion of the *caleosin* family [15, 27, 29]. Though this duplication mode has been proven to play a key role in environmental adaptation of plants [59–61], an amazing result is that most tandem repeats identified in tigernut, e.g., *CeCLO4*, *-5*, and *− 8*, were not or rarely expressed in all tissues examined in this study. Similar results were also reported for plasma membrane intrinsic protein (PIP), stearoyl-acyl carrier protein desaturase (SAD), and fatty acid desaturase 2 (FAD2) (sub)families [47, 48, 62]. Thereby, further clarifying their biological significance in tigernut is of particular interest.

Conclusions

This study presents the first genome-wide analysis of the *caleosin* family in tigernut, a unique oil-rich tuber plant in Cyperaceae. Eight family members representing two clades identified from the tigernut genome are considerably more than two present in the basal angiosperm *A. trichopoda*. Comparison of 193 members from 31 representative plant species reveals lineage-specific family expansion and functional diversification in tigernut, whose expansion was contributed by WGD, tandem, and dispersed duplications. Comparison between tigernut and purple nutsedge suggests that species-specific activation of certain members but not gene copies is the main contributor of high oil accumulation in tigernut tubers. These findings provide valuable information for further functional analysis and genetic improvement in tigernut and species beyond.

Materials and methods

Plant materials

Tigernut plants of *Cyperus esculentus var. sativus* Boeckeler (NCBI:txid1504802) variety Reyan3, which was mainly bred by Zhi Zou, were cultivated in sandy soil as described before [57]. Tubers of various developmental stages were collected at 5, 10, 20, 25, and 35 DAI, and three biological replicates were harvested for each stage. All samples were first freezed with liquid nitrogen and then stored at 80°Cfor further uses.

Datasets and identification of caleosin family genes

Genomic and transcriptome data of *C. esculentus* and representative plant species were accessed from CNGBdb (https://db.cngb.org/), RGAP7 (http://rice.plantbiology.msu.edu/), TAIR11 (https://www.arabidopsis.org/), Phytozome v13 (https://phytozome.jgi.doe.gov/pz/portal.html), and NCBI (https://www.ncbi.nlm.nih.gov/): *C. esculentus* (v1; Cyperaceae, Poales), *C. rotundus* (v1; Cyperaceae, Poales), *C. iria* (v1; Cyperaceae, Poales), *C. fuscus* (v1; Cyperaceae, Poales), *B. planiculmis* (v1; Cyperaceae, Poales),, *S. tabernaemontani* (v1; Cyperaceae, Poales), *C. littledalei* (v1; Cyperaceae, Poales), *R. breviuscula* (v1; Cyperaceae, Poales), *E. parvula* (v1; Cyperaceae, Poales), *J. effusus* (v1; Juncaceae, Poales), *J. inflexus* (v1; Juncaceae, Poales), *Sparganium stoloniferum* (v1; Typhaceae, Poales), *A. comosus* (v3; Bromeliaceae, Poales), *J. ascendens* (v1.1; Joinvilleaceae, Poales), *P. latifolius* (v1; Poaceae, Poales), *O. sativa* (RGAP7; Poaceae, Poales), *B. distachyon* (v3.2; Poaceae, Poales), *H. vulgare* (v1; Poaceae, Poales), *S. bicolor* (5.1; Poaceae, Poales), *S. italica* (v2.2; Poaceae, Poales), *E. guineensis* (EG5; Arecaceae, Arecales), *D. alata* (v2.1; Dioscoreaceae, Dioscoreales), *A. officinalis* (v1.1; Asparagaceae, Asparagales), *D. catenatum* (v1; Orchidaceae, Asparagales), *Spirodela polyrhiza* (v2; Lemnaceae, Alismatales), *Z. marina* (v3.1; Zosteraceae, Alismatales), *A. americanus* (v1; Acoraceae, Acorales), *A. thaliana* (Araport11; Brassicaceae, Brassicales), *R. communis* (WT05; Euphorbiaceae, Malpighiales), *V. vinifera* (T2T; *Vitaceae*, *Vitales*), and *A. trichopoda* (v2.1; Amborellaceae, Amborellales). Caleosin homologs were identified via HMMER (v3.3, http://hmmer.janelia.org/) searching with the Pfam profile PF05042 (v35.0, https://pfam.xfam.org/). Gene models of all candidates were curated with available mRNAs as previously described [63], and gene structure display was performed using GSDS2.0 [64]. Presence of the conserved caleosin domain in deduced proteins was confirmed using Pfam Search.

Sequence alignment and phylogenetic analysis

Multiple sequence alignment was conducted using Muscle (v5) [65], and sequence alignment display was carried out using SMS (v1, http://www.bioinformatics.org/sms/). Phylogenetic tree construction was performed using RAxML (v8) [65] with the maximum likelihood method and bootstrap of 1000 replicates. Systematic names of *caleosin* family genes were assigned with two italic letters denoting the source organism and a progressive number based on sequence similarity. To distinguish with *caleosin* genes in arabidopsis, three italic letters were used in *A. trichopoda*, i.e., *AtrCLO1* and − *2*.

Protein properties and conserved motif analysis

Physical and chemical parameters (e.g., MW, pI, and GRAVY) of deduced proteins were calculated using ProtParam (v1, http://web.expasy.org/protparam/), whereas hydropathicity scales were determined using ProtScale (v1,

https://web.expasy.org/protscale/). Protein secondary structures were determined using SOPMA (v1, https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), while conserved motifs were identified using MEME (v5.4.1, https://meme-suite.org/tools/meme) with the parameters as follows: any number of repetitions; the maximum number of motifs, 10; and, the optimum width of each motif, between 5 and 200 residues.

Chromosomal localization, synteny analysis, definition of orthogroups, and calculation of evolutionary rate

Synteny analysis was conducted as previously described [67], where duplicate pairs were identified using the all-to-all BLASTp method with the *E*-value cutoff of 1e-10 and syntenic blocks were inferred using MCScanX [68] with at least five BLAST hits. Orthologs between different species were identified using OrthoFinder (v2.3.8) [69], whereas different modes of gene duplication were identified using the DupGen_finder pipeline [61]. Chromosomal localization and calculation of Ks and Ka values were performed using TBtools-II (v2.390) [70].

Gene expression analysis based on RNA-seq

Global expression profiles of *caleosin* and *oleosin* genes in tigernut and yellow nutsedge were investigated by using transcriptome datasets that are under NCBI accession numbers of PRJNA703731 and PRJNA671562. The data are 150 bp paired-end reads with three biological replicates, which were derived from shoot apexes, young leaves, mature leaves, leaf sheaths, roots, rhizomes, and tubers of 40 DAS, 85 DAS, 120 DAS, 20 DAI, 50 DAI, and 90 DAI. Raw reads in the FASTQ format were obtained using fastq-dump (v1) [71], and quality control was carried out using Trimmomatic (v1) [72]. Read mapping was performed using HISAT (v2) [73], and relative gene expression level was presented as FPKM [74].

Gene expression analysis based on qRT-PCR

Total RNA extraction, the integrity and concentration detection, synthesis of the first-strand cDNA, and qRT-PCR analysis were conducted as previously described [57, 63]. Primers used are shown in **Additional file 7**, where *CeTIP41* and *CeUCE2* are two reference genes as described before [37]. All qRT-PCR assays were conducted in triplicate for each biological sample, and relative gene abundance was estimated with the $2^{-\Delta\Delta Ct}$ method and statistical analysis was performed using SPSS Statistics 20 as described before [63].

Declarations

Acknowledgements

The authors appreciate those contributors who make the related genome and transcriptome data accessible in public databases.

Authors' contributions

The study was conceived and directed by ZZ. All the experiments and analyses were directed by ZZ and carried out by ZZ, XC, HH, CL, XY, JH, and ZY. ZZ, JH, and ZY wrote the paper. All the authors read and approved the final manuscript.

Funding

This work was supported by the Hainan Province Science and Technology Special Fund (ZDYF2024XDNY171 and ZDYF2024XDNY156) and the National Natural Science Foundation of China (32460342 and 31971688). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets analyzed during the current study are available in the NCBI SRA repository (https://www.ncbi.nlm.nih.gov/sra/) under accession numbers of PRJNA703731 and PRJNA671562.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Huang AHC. Plant lipid droplets and their associated proteins: Potential for rapid advances. Plant Physiol. 2018;176(3):1894–918. doi:10.1104/pp.17.01677.
2. Frandsen G, Müller-Uri F, Nielsen M, Mundy J, Skriver K. Novel plant Ca^{2+}-binding protein expressed in response to abscisic acid and osmotic stress. J Biol Chem. 1996;271(1):343–8. doi:10.1074/jbc.271.1.343.
3. Chen EC, Tai SS, Peng CC, Tzen JT. Identification of three novel unique proteins in seed oil bodies of sesame. Plant Cell Physiol. 1998;39(9):935–41. doi: 10.1093/oxfordjournals.pcp.a029457.
4. Hanano A, Burcklen M, Flenet M, Ivancich A, Louwagie M, Garin J, Blee E. Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif. J Biol Chem. 2006;281(44):33140–51. doi:10.1074/jbc.M605395200.
5. Yamaguchi-Shinozaki K, Koizumi M, Urao S, Shinozaki K. Molecular cloning and characterization of 9 cDNAs for genes that are responsive to desiccation in *Arabidopsis thaliana*: sequence analysis of one cDNA clone that encodes a putative transmembrane channel protein. Plant Cell Physiol. 1992;33:217–24.
6. Takahashi S, Katagiri T, Yamaguchi-Shinozaki K, Shinozaki K. An *Arabidopsis* gene encoding a Ca^{2+}-binding protein is induced by abscisic acid during dehydration. Plant Cell Physiol. 2000;41(7):898–903. doi: 10.1093/pcp/pcd010.
7. Chen JC, Tsai CC, Tzen JT. Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds. Plant Cell Physiol. 1999;40(10):1079–86. doi: 10.1093/oxfordjournals.pcp.a029490.
8. Naested H, Frandsen GI, Jauh GY, Hernandez-Pinzon I, Nielsen HB, Murphy DJ, Rogers JC, Mundy J. Caleosins: Ca^{2+}-binding proteins associated with lipid bodies. Plant Mol Biol. 2000;44(4):463–76. doi:10.1023/a:1026564411918.
9. Rahman F, Hassan M, Hanano A, Fitzpatrick DA, McCarthy CGP, Murphy DJ. Evolutionary, structural and functional analysis of the caleosin/peroxygenase gene family in the Fungi. BMC Genom. 2018;19(1):976. doi:10.1186/s12864-018-5334-1.
10. Rahman F, Hassan M, Rosli R, Almousally I, Hanano A, Murphy DJ. Evolutionary and genomic analysis of the caleosin/peroxygenase (CLO/PXG) gene/protein families in the Viridiplantae. PLoS One. 2018;13(5):e0196669. doi:10.1371/journal.pone.0196669.
11. Hanano A, Blée E, Murphy DJ. Caleosin/peroxygenases: Multifunctional proteins in plants. Ann Bot. 2023;131(3):387–409. doi:10.1093/aob/mcad001.
12. Fu X, Liu K, Du Y, Li B, Yuan H, Yang J. ApCaleosin resists cadmium stress by regulating lipid droplet formation and intracellular oxidative stress levels in microalgae. New Phytol. 2025;248(3):1321–35. doi: 10.1111/nph.70507.
13. Zou Z, Zhao Y, Zhang L. Genomic insights into lineage-specific evolution of the *oleosin* family in Euphorbiaceae. BMC Genomics. 2022;23(1):178. doi: 10.1186/s12864-022-08412-z.
14. Zou Z, Zhang L, Zhao Y. Integrative analysis of *oleosin* genes provides insights into lineage-specific family evolution in Brassicales. Plants (Basel). 2024;13(2):280. doi: 10.3390/plants13020280.

15. Shen Y, Xie J, Liu RD, Ni XF, Wang XH, Li ZX, Zhang M. Genomic analysis and expression investigation of *caleosin* gene family in *Arabidopsis*. Biochem Biophys Res Commun. 2014;448(4):365–71. doi:10.1016/j.bbrc.2014.04.115.
16. Shen Y, Liu M, Wang L, Li Z, Taylor DC, Li Z, Zhang M. Identification, duplication, evolution and expression analyses of caleosins in Brassica plants and *Arabidopsis* subspecies. Mol Genet Genomics. 2016;291(2):971–88. doi:10.1007/s00438-015-1156-x.
17. Chen JC, Tzen JT. An *in vitro* system to examine the effective phospholipids and structural domain for protein targeting to seed oil bodies. Plant Cell Physiol. 2001;42(11):1245–52. doi:10.1093/pcp/pce160.
18. Blée E, Flenet M, Boachon B, Fauconnier ML. A non-canonical caleosin from *Arabidopsis* efficiently epoxidizes physiological unsaturated fatty acids with complete stereoselectivity. FEBS J. 2012;279(20):3981–95. doi:10.1111/j.1742-4658.2012.08757.x.
19. Charuchinda P, Waditee-Sirisattha R, Kageyama H, Yamada D, Sirisattha S, Tanaka Y, Mahakhant A, Takabe T. Caleosin from *Chlorella vulgaris* TISTR 8580 is salt-induced and heme-containing protein. Biosci Biotechnol Biochem. 2015;79(7):1119–24. doi:10.1080/09168451.2015.1010480.
20. Partridge M, Murphy DJ. Roles of a membrane-bound caleosin and putative peroxygenase in biotic and abiotic stress responses in *Arabidopsis*. Plant Physiol Biochem. 2009;47(9):796–806. doi:10.1016/j.plaphy.2009.04.005.
21. Kim YY, Jung KW, Yoo KS, Jeung JU, Shin JS. A stress-responsive caleosin-like protein, AtCLO4, acts as a negative regulator of ABA responses in *Arabidopsis*. Plant Cell Physiol. 2011;52(5):874–84. doi:10.1093/pcp/pcr039.
22. Hernandez-Pinzon I, Patel K, Murphy DJ. The *Brassica napus* calcium-binding protein, caleosin, has distinct endoplasmic reticulum- and lipid body-associated isoforms. Plant Physiol Biochem. 2001;39(7-8):615–22. doi: 10.1016/S0981-9428(01)01274-8
23. Jiang PL, Jauh GY, Wang CS, Tzen JT. A unique caleosin in oil bodies of lily pollen. Plant Cell Physiol. 2008;49(9):1390–5. doi: 10.1093/pcp/pcn103.
24. Jiang PL, Chen JC, Chiu ST, Tzen JT. Stable oil bodies sheltered by a unique caleosin in cycad megagametophytes. Plant Physiol Biochem. 2009;47(11-12):1009–16. doi:10.1016/j.plaphy.2009.07.004.
25. Zienkiewicz K, Castro AJ, Alché Jde D, Zienkiewicz A, Suárez C, Rodríguez-García MI. Identification and localization of a caleosin in olive (*Olea europaea* L.) pollen during *in vitro* germination. J Exp Bot. 2010;61(5):1537–46. doi: 10.1093/jxb/erq022.
26. Lu JY, Xiong SX, Yin W, Teng XD, Lou Y, Zhu J, Zhang C, Gu JN, Wilson ZA, Yang ZN. MS1, a direct target of MS188, regulates the expression of key sporophytic pollen coat protein genes in *Arabidopsis*. J Exp Bot. 2020;71(16):4877–89. doi: 10.1093/jxb/eraa219.
27. Jing P, Kong D, Ji L, Kong L, Wang Y, Peng L, Xie G. OsClo5 functions as a transcriptional co-repressor by interacting with OsDi19-5 to negatively affect salt stress tolerance in rice seedlings. Plant J. 2021;105(3):800–15. doi: 10.1111/tpj.15074.
28. Miklaszewska M, Zienkiewicz K, Klugier-Borowska E, Rygielski M, Feussner I, Zienkiewicz A. CALEOSIN 1 interaction with AUTOPHAGY-RELATED PROTEIN 8 facilitates lipid droplet microautophagy in seedlings. Plant Physiol. 2023;193(4):2361–80. doi: 10.1093/plphys/kiad471.
29. Zou Z, Fu XW, Yi XP, Li CQ, Zhao YG. Characterization of *caleosin* genes in *Carica papaya* and insights into lineage-specific family evolution in Brassicales. Plants. 2025;14(21):3296. doi: 10.3390/plants14213296.
30. Hyun TK, Kumar D, Cho YY, Hyun HN, Kim JS. Computational identifcation and phylogenetic analysis of the oil-body structural proteins, oleosin and caleosin, in castor bean and fax. Gene. 2013;515(2):454–60. doi:10.1016/j.gene.2012.11.065.
31. Khalil HB, Brunetti SC, Pham UM, Maret D, Laroche A, Gulick PJ. Characterization of the *caleosin* gene family in the Triticeae. BMC Genom. 2014;15(1):239. doi: 10.1186/1471-2164-15-239.
32. Song W, Qin Y, Zhu Y, Yin G, Wu N, Li Y, Hu Y. Delineation of plant caleosin residues critical for functional divergence, positive selection and coevolution. BMC Evol Biol. 2014;14:124. doi: 10.1186/1471-2148-14-124.

33. Liu H, Hedley P, Cardle L, Wright KM, Hein I, Marshall D, Waugh R. Characterisation and functional analysis of two barley caleosins expressed during barley caryopsis development. Planta. 2005;221(4):513–22. doi: 10.1007/s00425-004-1465-5.
34. Liu X, Yang Z, Wang Y, Shen Y, Jia Q, Zhao C, Zhang M. Multiple caleosins have overlapping functions in oil accumulation and embryo development. J Exp Bot. 2022;73(12):3946–62. doi: 10.1093/jxb/erac153.
35. Jolivet P, Boulard C, Bellamy A, Valot B, d'Andréa S, Zivy M, Nesi N, Chardot T. Oil body proteins sequentially accumulate throughout seed development in *Brassica napus*. J Plant Physiol. 2011;168(17):2015–20. doi: 10.1016/j.jplph.2011.06.007.
36. Turesson H, Marttila S, Gustavsson KE, Hofvander P, Olsson ME, Bülow L, Stymne S, Carlsson AS. Characterization of oil and starch accumulation in tubers of *Cyperus esculentus* var. *sativus* (Cyperaceae): A novel model system to study oil reserves in nonseed tissues. Am J Bot. 2010;97(11):1884–93. doi:10.3732/ajb.1000200.
37. Zou Z, Xiao Y, Zhang L, Zhao Y. Analysis of *Lhc* family genes reveals development regulation and diurnal fluctuation expression patterns in *Cyperus esculentus*, a Cyperaceae plant. Planta. 2023;257(3):59. doi: 10.1007/s00425-023-04092-5.
38. Pereira-Silva L, Trevisan R, Escudero M, Alcantara S, de Oliveira Ribeiro AR, Larridon I. Biogeography and diversification dynamics of the megadiverse plant genus *Cyperus* (Cyperaceae, Poales). Ann Bot. 2025;4:mcaf117. doi: 10.1093/aob/mcaf117. Epub ahead of print.
39. Yang X, Niu L, Zhang Y, Ren W, Yang C, Yang J, Xing G, Zhong X, Zhang J, Slaski J, Zhang J. Morpho-agronomic and biochemical characterization of accessions of tiger nut (*Cyperus esculentus*) grown in the north temperate zone of China. Plants (Basel). 2022;11(7):923. doi: 10.3390/plants11070923.
40. Zhang S, Li P, Wei Z, Cheng Y, Liu J, Yang Y, Wang Y, Mu Z. *Cyperus* (*Cyperus esculentus* L.): A review of its compositions, medical efficacy, antibacterial activity and allelopathic potentials. Plants (Basel). 2022;11(9):1127. doi: 10.3390/plants11091127.
41. Gao Y, Sun Y, Gao H, Chen Y, Wang X, Xue J, Jia X, Li R. Ectopic overexpression of a type-II DGAT (CeDGAT2-2) derived from oil-rich tuber of *Cyperus esculentus* enhances accumulation of oil and oleic acid in tobacco leaves. Biotechnol Biofuels. 2021;14(1):76. doi: 10.1186/s13068-021-01928-8.
42. Ji H, Liu D, Yang Z. High oil accumulation in tuber of yellow nutsedge compared to purple nutsedge is associated with more abundant expression of genes involved in fatty acid synthesis and triacylglycerol storage. Biotechnol Biofuels. 2021;14(1):54. doi: 10.1186/s13068-021-01909-x.
43. Zou, Z., Zhao, Y.G., Zhang, L., Kong, H., Guo, Y.L., Guo, A.P., 2021. Single-molecule real-time (SMRT)-based full-length transcriptome analysis of tigernut (*Cyperus esculentus* L.). Chin. J. Oil Crop Sci. 43, 229–235. doi:10.19802/j.issn.1007-9084.2020230.
44. Xu, S., Zou, Z., Xiao, Y.H., Zhang, L., Kong, H., Guo, J.Y., Guo, A.P., 2022. Cloning and functional characterization of *CeWRI1*, a gene involved in oil accumulation from tigernut (*Cyperus esculentus* L.) tubers. Chin. J. Trop. Crops 43, 923–929. doi:46.1019.S.20220225.1223.002.
45. Zou Z, Zheng Y, Zhang Z, Xiao Y, Xie Z, Chang L, Zhang L, Zhao Y. Molecular characterization *oleosin* genes in *Cyperus esculentus*, a Cyperaceae plant producing oil in underground tubers. Plant Cell Rep. 2023;42(11):1791–808. doi:10.1007/s00299-023-03066-x.
46. Zou Z, Fu X, Huang J, Zhao Y. Molecular characterization of *CeOLE6*, a diverged SH *oleosin* gene, preferentially expressed in *Cyperus esculentus* tubers. Planta. 2024;260(6):122. doi: 10.1007/s00425-024-04553-5.
47. Zou Z, Fu X, Li C, Yi X, Huang J, Zhao Y. Insights into the stearoyl-acyl carrier protein desaturase (SAD) family in tigernut (*Cyperus esculentus* L.), an oil-bearing tuber plant. Plants (Basel). 2025;14(4):584. doi: 10.3390/plants14040584.
48. Zou Z, Fu X, Li C, Huang J, Zhao Y. Insights into membrane-bound fatty acid desaturase genes in tigernut (*Cyperus esculentus* L.), an oil-rich tuber plant in Cyperaceae. BMC Plant Biol. 2025;25(1):382. doi: 10.1186/s12870-025-06398-w.

49. Zou Z, Fu X, Li N, Zheng P, Wang R, Liu H, Huang J, Zhao Y. The transcription factor CeWRI3 functions in TAG accumulation through activating *CeOLE2* in *Cyperus esculentus*. Plant Cell Rep. 2025;44(7):146. doi: 10.1007/s00299-025-03534-6.
50. Xu C, Shanklin J. Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. Annu Rev Plant Biol. 2016;67:179–206. doi: 10.1146/annurev-arplant-043015-111641.
51. Niemeyer PW, Irisarri I, Scholz P, Schmitt K, Valerius O, Braus GH, Herrfurth C, Feussner I, Sharma S, Carlsson AS, de Vries J, Hofvander P, Ischebeck T. A seed-like proteome in oil-rich tubers. Plant J. 2022;112(2):518–34. doi: 10.1111/tpj.15964.
52. Zhu Y, Wang Y, Wei Z, Zhang X, Jiao B, Tian Y, Yan F, Li J, Liu Y, Yang X, Zhang J, Wang X, Mu Z, Wang Q. Analysis of oil synthesis pathway in *Cyperus esculentus* tubers and identification of *oleosin* and *caleosin* genes. J Plant Physiol. 2023;284:153961. doi: 10.1016/j.jplph.2023.153961.
53. Zhao X, Yi L, Ren Y, Li J, Ren W, Hou Z, Su S, Wang J, Zhang Y, Dong Q, Yang X, Cheng Y, Lu Z. Chromosome-scale genome assembly of the yellow nutsedge (*Cyperus esculentus*). Genome Biol Evol. 2023;15(3):evad027. doi: 10.1093/gbe/evad027.
54. Huang CY, Huang AHC. Unique motifs and length of hairpin in oleosin target the cytosolic side of endoplasmic reticulum and budding lipid droplet. Plant Physiol. 2017;174(4):2248–60. doi: 10.1104/pp.17.00366.
55. Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 2003;422(6930):433–8. doi: 10.1038/nature01521.
56. Jiao Y, Li J, Tang H, Paterson AH. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell. 2014;26(7):2792–802. doi: 10.1105/tpc.114.127597.
57. Zou Z, Zhao YG, Zhang L, Xiao YH, Guo AP. Analysis of *Cyperus esculentus SMP* family genes reveals lineage-specific evolution and seed desiccation-like transcript accumulation during tuber maturation. Ind Crop Prod. 2022;187:115382. doi: 10.1016/j.indcrop.2022.115382.
58. Zhao Y, Fu X, Zou Z. Insights into genes encoding LEA_1 domain-containing proteins in *Cyperus esculentus*, a desiccation-tolerant tuber plant. 2024;13(20):2933. doi: 10.3390/plants13202933.
59. Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, Difazio SP. Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. Genome Res. 2012;22(1):95–105. doi: 10.1101/gr.125146.111.
60. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171(4):2294–316. doi: 10.1104/pp.16.00523.
61. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol. 2019;20(1):38. doi: 10.1186/s13059-019-1650-2.
62. Zou Z, Zheng YJ, Xiao YH, Liu HY, Huang JQ, Zhao YG. Molecular insights into PIP aquaporins in tigernut (*Cyperus esculentus* L.), a Cyperaceae tuber plant. Tropical Plants. 2024;3:e027. doi: 10.48130/tp-0024-0030.
63. Zou Z, Gong J, An F, Xie G, Wang J, Mo Y, Yang L. Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. BMC Genomics. 2015;16:1001. doi: 10.1186/s12864-015-2152-6.
64. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7. doi: 10.1093/bioinformatics/btu817.
65. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun. 2022;13(1):6968. doi: 10.1038/s41467-022-34630-w.
66. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3. doi: 10.1093/bioinformatics/btu033.
67. Zou Z, Yang JH. Genomic analysis of Dof transcription factors in *Hevea brasiliensis*, a rubber-producing tree. Ind Crop Prod. 2019;134:271–83. doi: 10.1016/j.indcrop.2019.04.013.

68. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. doi: 10.1093/nar/gkr1293.
69. Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. doi: 10.1186/s13059-019-1832-y.
70. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant. 2023;16(11):1733–42. doi: 10.1016/j.molp.2023.09.010.
71. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. doi: 10.1093/bioinformatics/bty560.
72. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170.
73. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60. doi: 10.1038/nmeth.3317.
74. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods. 2008;5(7):621–8. doi: 10.1038/nmeth.1226.

Figures

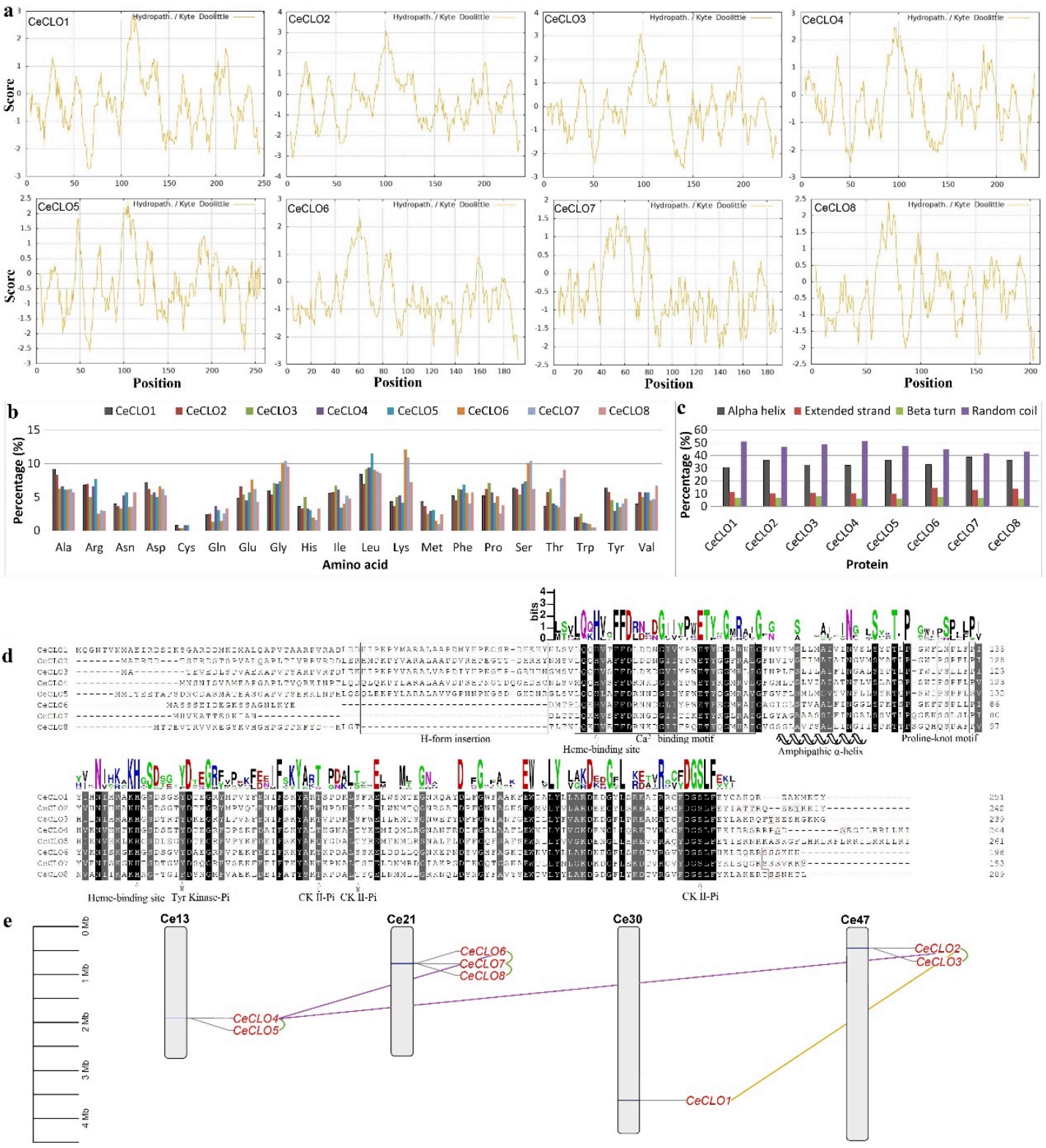
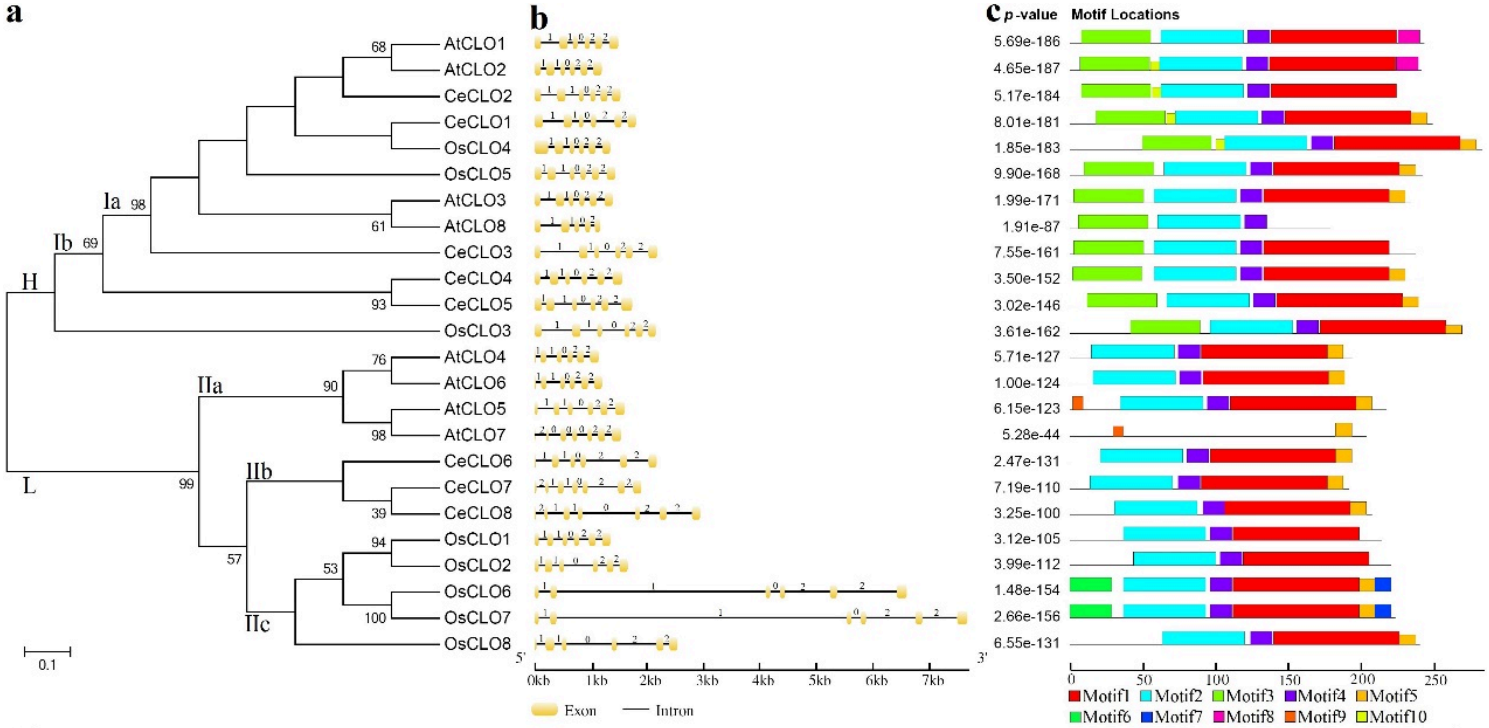


Figure 1

Sequence and duplication event analysis of *caleosin* family genes in *C. esculentus*. **a** Kyte-Doolittle hydrophobicity plots of CeCLOs analyzed using ProtScale (v1). **b** Amino acid composition of CeCLOs calculated using ProtParam (v1). **c** Predicted secondary structure of CeCLOs using the SOPMA method. **d** Multiple sequence alignment of CeCLOs using MUSCLE (v5.1). Potential phosphorylation sites were determined using NetPhos (v3.1). Broken lines in the sequences represent gaps introduced for best alignment, whereas identical and similar amino acids are highlighted in black or dark grey, respectively. The SeqLogo of the caleosin domain generated using WebLogo (v3) is shown above the alignment, and the positions of the

H-form insertion, an EF-hand calcium binding motif, the amphipathic α-helix, and the proline knot-like motif ($PX_{3-4}PSX_3P$) are boxed in black and are indicated at the bottom of the sequences. Two invariable His residues for heme-binding and four phosphorylation sites (one for Tyr kinase and three for CK II) are underlined and pointed by arrows. Potential phosphorylated residues in the C-terminal regions are boxed in red, whereas a conserved Cys residue prior to the last CK II phosphorylation site is boxed in purple. **e** Gene localization and duplication events detected in *C. esculentus*. Gene localization was conducted using TBtools-II (v2.210), whereas different modes of gene duplication were identified using the DupGen_finder pipeline. Serial numbers are indicated at the top of each scaffold, and the scale is in Mb. Duplicate pairs identified in this study are connected using lines in different colors, i.e., tandem (green), dispersed (purple), and WGD (gold). (*Ce C. esculentus*, *CK II* casein kinase II, *CLO* caleosin, *Mb* megabase, *WGD* Whole-genome duplication)

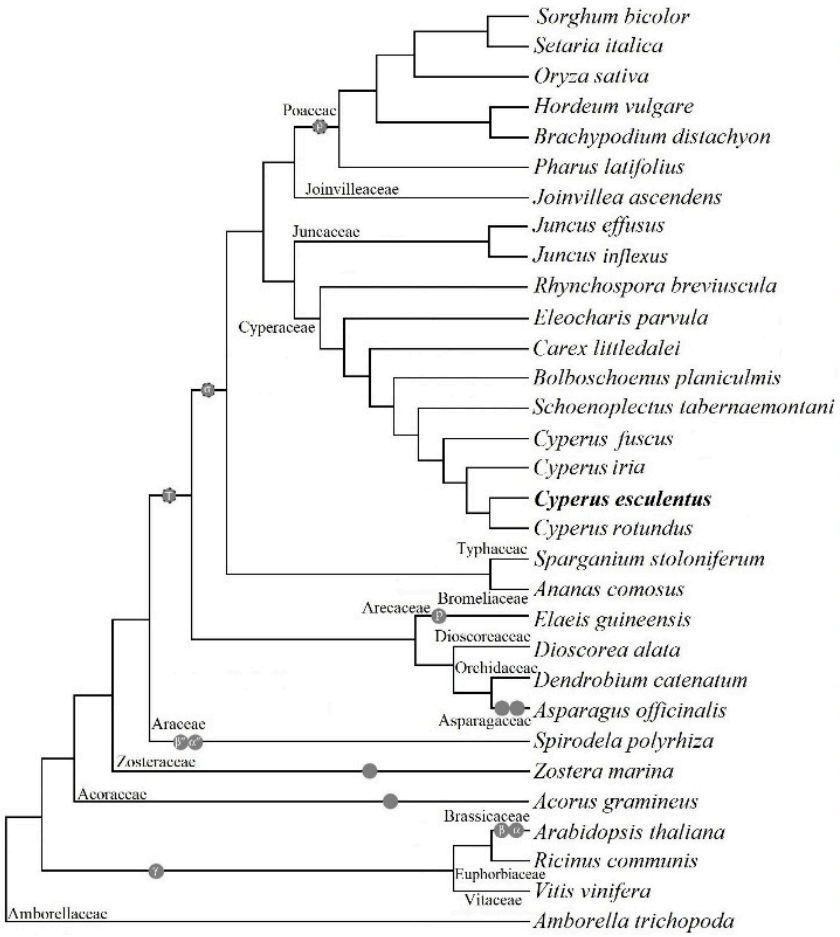


d

Motif	Sites	Width	Best possible match	Functions
1	22	87	SDSGTYDTEGRFVPVKFEEIFSKYARTHPDALTYKELWEMLEGN REPYDFFGWLAAKGEWKLLYVLAKDKDGFLSKEAVRGCYDGSL	C-terminal of the caleosin domain that includes the second heme-binding site essential for the PXG activity and four phosphorylation sites for posttranslational regulation
2	23	57	MSVLQQHVAFFDRBKDGIIYPWETYQGFRAJGFGVLLSLFLAIFIN GGLSYKTLPGW	N-terminal of the caleosin domain that includes the first heme-binding site essential for the PXG activity, the Ca^{2+} binding motif for calcium signaling, and the amphipathic α-helix for LD anchoring
3	12	48	ERDSLATVAPQAPVTSERPVRPDLEDHLPKPYLARALAAPDMEH PNGT	The H-form insertion with unkown functions
4	23	15	PLFPIYIKNIHKAKH	The proline-knot motif necessary for budding-LD entering cytosol
5	17	11	FEKLEKZRSSS	C-terminal phosphorylation sites
6	2	29	MRAQRPSLAAAGNPVLALLFLWVLGWGHV	Unkown
7	2	11	QSTVCRSLSFR	Unkown
8	2	15	EYCAKNYAEIKEDKT	Unkown
9	2	7	FCFCFCF	Unkown
10	4	6	DHRHYH	The H-form insertion with unkown functions

Figure 2

Structural and phylogenetic analysis of *caleosin* family genes in *C. esculentus*, *O. sativa*, and *A. thaliana*. **a** Shown is an unrooted phylogenetic tree resulting from full-length caleosins with RAxML (maximum likelihood method and bootstrap of 1000 replicates), where the distance scale denotes the number of amino acid substitutions per site and the name of each clade (i.e., H and L) is indicated next to the corresponding group. **b** Shown are the exon-intron structures. “0”, “1”, and “2” indicate that an intron is located between codons, the first and second bases of a codon, and the second and third bases of a codon, respectively. **c** Shown is the distribution of conserved motifs identified using MEME, where different motifs are represented by different color blocks as indicated and the same color block in different proteins indicates a certain motif. **d** Shown is the detailed information of ten conserved motifs identified in this study. (*At A. thaliana*, *Ce C. esculentus*, *CLO* caleosin, *LD* lipid droplet, *PXG* peroxygenase, *Os O. sativa*)



	H1	H2	H3	H4	H5	L1	L2	L3	L4	L5	L6	Total
Sorghum bicolor	1	0	0	2	0	1	0	0	1	3	2	10
Setaria italica	1	0	0	1	0	1	0	0	1	1	3	8
Oryza sativa	1	1	0	1	0	1	0	0	1	2	1	8
Hordeum vulgare	1	3	0	1	0	1	0	0	1	1	2	10
Brachypodium distachyon	1	2	0	1	0	1	0	0	1	2	1	9
Pharus latifolius	1	1	0	1	0	1	0	0	0	1	0	5
Joinvillea ascendens	1	1	0	1	0	3	0	0	0	0	0	6
Juncus effusus	1	1	0	1	0	2	0	0	0	0	0	5
Juncus inflexus	1	1	1	0	0	1	0	0	0	0	0	4
Rhynchospora breviuscula	1	1	1	1	0	1	2	1	0	0	0	8
Eleocharis parvula	1	1	1	1	0	2	1	0	0	0	0	7
Carex littledalei	1	1	1	0	0	2	1	1	0	0	0	7
Bolboschoenus planiculmis	1	1	1	2	0	1	1	1	0	0	0	8
Schoenoplectus tabernaemontani	1	1	1	1	0	2	1	0	0	0	0	7
Cyperus fuscus	1	1	1	1	0	1	1	1	0	0	0	7
Cyperus iria	2	2	1	2	2	2	2	1	0	0	0	14
Cyperus esculentus	1	1	1	1	1	1	1	1	0	0	0	8
Cyperus rotundus	1	1	1	1	1	1	1	1	0	0	0	8
Sparganium stoloniferum	2	1	0	0	0	3	0	0	0	0	0	6
Ananas comosus	1	1	0	2	0	1	0	0	0	0	0	5
Elaeis guineensis	1	0	0	0	0	5	0	0	0	0	0	6
Dioscorea alata	3	0	0	1	0	1	0	0	0	0	0	5
Dendrobium catenatum	1	0	0	1	0	1	0	0	0	0	0	3
Asparagus officinalis	3	0	0	1	0	1	0	0	0	0	0	5
Spirodela polyrhiza	1	0	0	0	0	3	0	0	0	0	0	4
Zostera marina	1	0	0	0	0	1	0	0	0	0	0	2
Acorus gramineus	3	0	0	0	0	1	0	0	0	0	0	4
Arabidopsis thaliana	4	0	0	0	0	4	0	0	0	0	0	8
Ricinus communis	1	0	0	0	0	1	0	0	0	0	0	2
Vitis vinifera	1	0	0	0	0	1	0	0	0	0	0	8
Amborella trichopoda	1	0	0	0	0	1	0	0	0	0	0	2

Figure 3

Species-specific distribution of 11 caleosinorthogroups in 31 representative plant species. The species tree is referred to NCBI Taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy) and well-established recent WGDs are marked, which include the γ WGT shared by all core eudicots, the β WGD shared by the majority of Brassicales plants, the α WGD specific to Brassicaceae plants, the τ WGD shared by all core monocots, the σ WGD specific to Poales plants, the ρ WGD specific to Poaceae plants, and the p WGD specific to Arecaceae plants. Names of tested plant families are indicated next to the corresponding branches. (*H*high molecular weight, *L* low molecular weight, *WGD* whole-genome duplication, *WGT* whole-genome triplication)

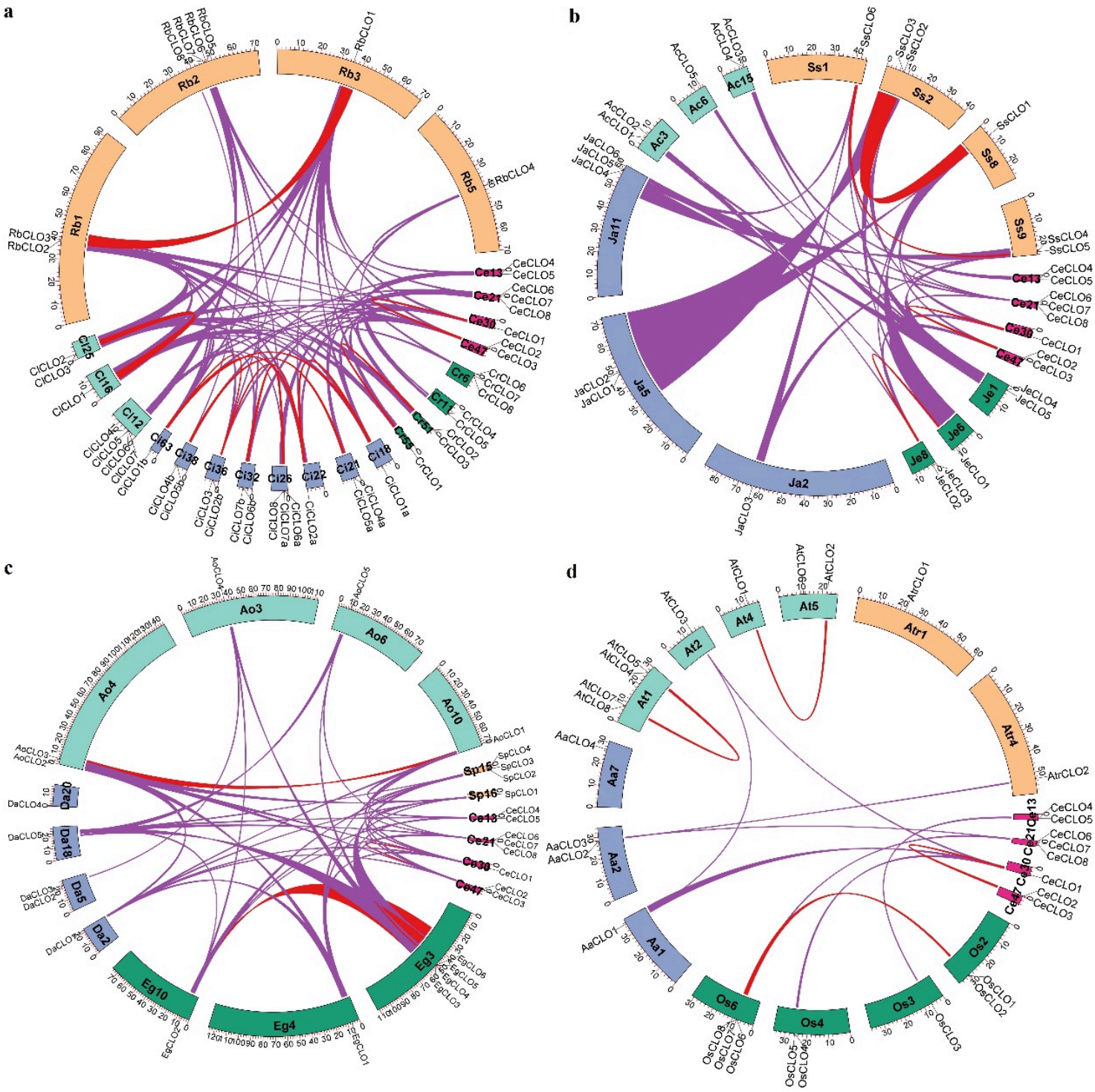


Figure 4

Synteny analyses within and between *C. esculentus* and representative plant species. **a** Synteny analysis within and between *C. esculentus*, *C. rotundus*, *C. iria*, *C. littledalei*, and *R. breviuscula*. **b** Synteny analysis within and between *C. esculentus*, *J. effusus*, *J. ascendens*, *A. comosus*, and *S. stoloniferum*. **c** Synteny analysis within and between *C. esculentus*, *E. guineensis*, *D. alata*, and *A. officinalis*. **d** Synteny analysis within and between *C. esculentus*, *O. sativa*, *A. americanus*, *A. thaliana*, and *A. trichopoda*. Syntenic blocks were inferred using MCScanX (*E*-value≤1e−10; BLAST hits≥5). Shown are *caleosin* gene-encoding chromosomes/scaffolds and only syntenic blocks containing *caleosin* genes are marked, where red and purple lines indicate intra- and inter-species, respectively. The scale is in Mb. (*Aa A. americanus*, *Ac A. comosus*, *Ao A. officinalis*, *At A. thaliana*, *Atr A. trichopoda*, *Ce C. esculentus*, *Ci C. iria*, *Cl C. littledalei*, *CLO* caleosin, *Cr C. rotundus*, *Da D. alata*, *Eg E. guineensis*, *Ja J. ascendens*, *Je J. effusus*, *Os O. sativa*, *R. breviuscula*, *Ss S. stoloniferum*)

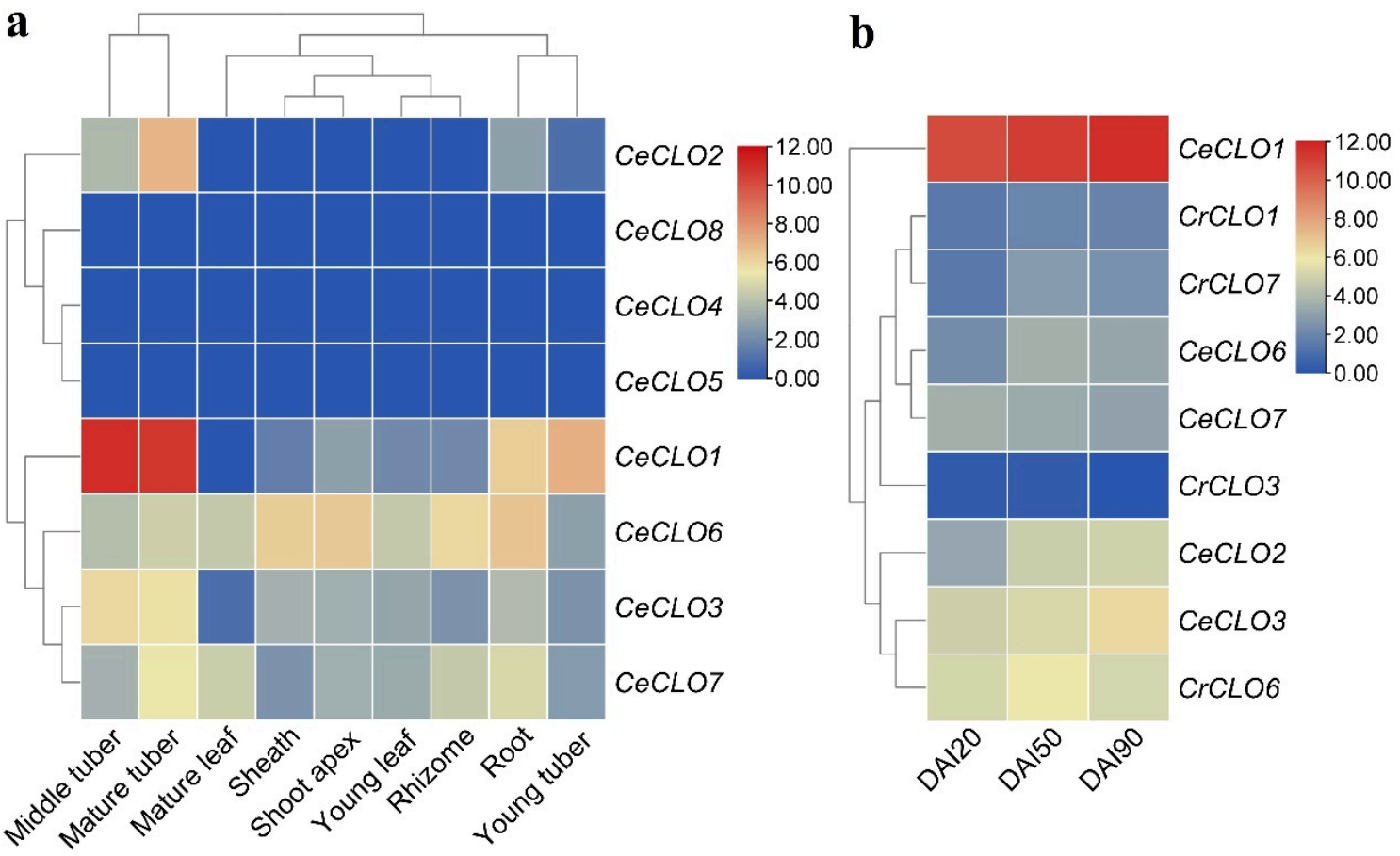


Figure 5

Expression profiles of *Ce/CrCLO* genes. **a** Tissue-specific expression profiles of *CeCLO* genes. Young, middle, and mature represent tubers collected at 40, 85, and 120 DAS. **b** Expression profiles of *Ce/CrOLE* genes in three representative stages of tuber development. The heatmaps were generated using the R package (v3.5.0), where color scale represents FPKM normalized $\log_2$ transformed counts. Blue and red indicate low and high expression, respectively. (*Ce C. esculentus*, *CLO* caleosin, *Cr C. rotundus*, *DAS* days after sowing, *FPKM* fragments per kilobase of exon per million fragments mapped)

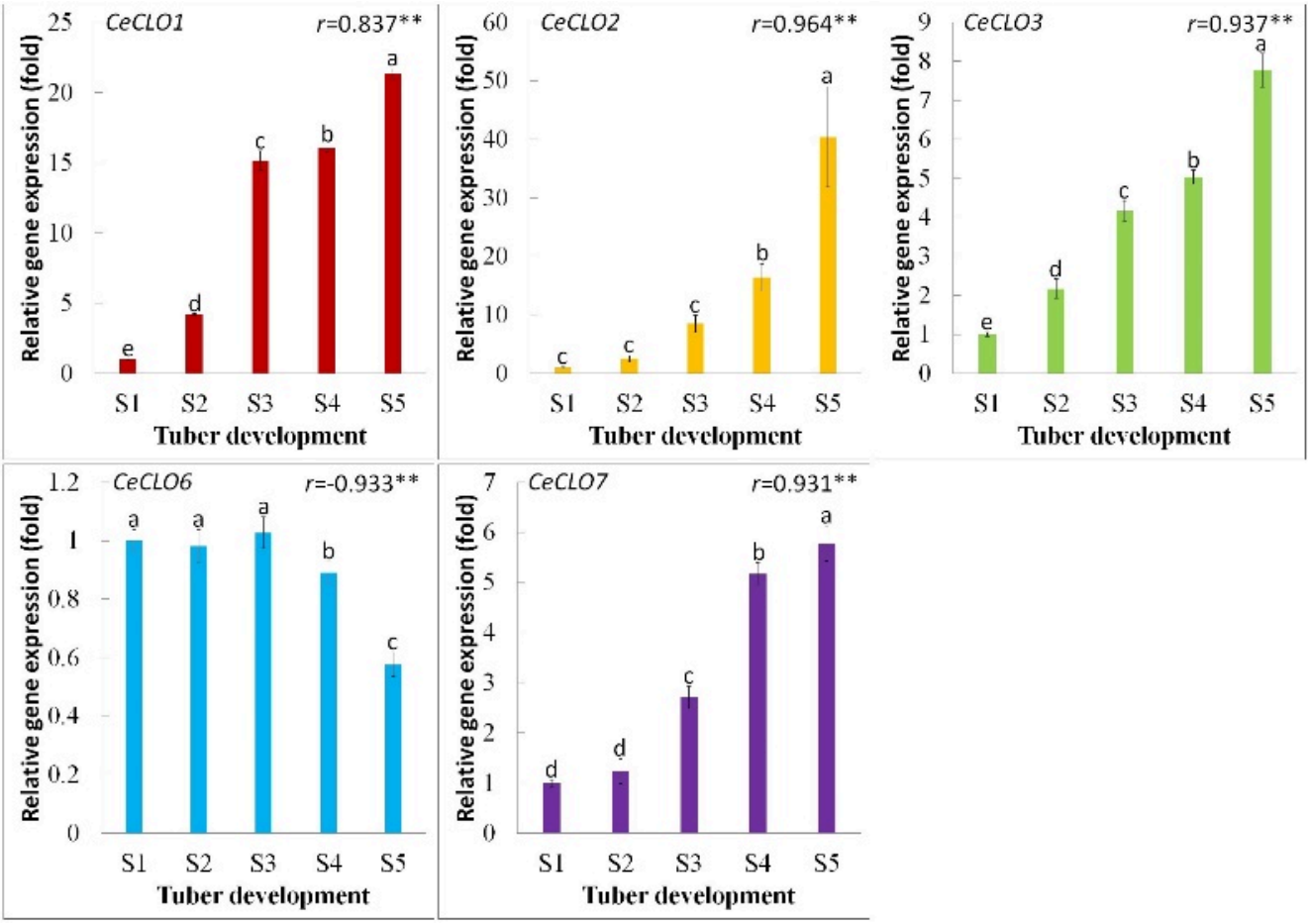


Figure 6

Expression profiles of five *CeCLO* genes during tuber development. S1–5 represent tubers collected at 5, 10, 20, 25, and 35 DAI. Bars indicate SD (N=3) and lowercase letters indicate difference significance tested following Duncan's one-way multiple-range post hoc ANOVA ($P<0.05$). *r* represents the Pearson's correlation coefficient. (*Ce C. esculentus*, *CLO* caleosin, *DAI* days after tuber initiation)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.zip