Planta (2020) 252:83
https://doi.org/10.1007/s00425-020-03490-3

ORIGINAL ARTICLE

CnMADS1, a MADS transcription factor, positively modulates cell proliferation and lipid metabolism in the endosperm of coconut (*Cocos nucifera* L.)

RuHao Sun[1] · Li Gao[1] · Zhiqi Mi[1] · Yusheng Zheng[1] · Dongdong Li[1]

Received: 24 September 2020 / Accepted: 4 October 2020 / Published online: 10 October 2020

Abstract

Main conclusion **The function of the first MADS-box transcription factor from endosperm of coconut, CnMADS1, was characterized via seed-specific overexpression in Arabidopsis seeds and further confirmed in protoplasts of coconut.**

Abstract Coconut (*Cocos nucifera* L.), which belongs to the palm family (*Arecaceae*), is one of the world's most useful economical tropical crops. However, few genes related to coconut endosperm development have been studied. In previous research, an AGAMOUS-like (*AGL*) MADS-box transcription factor, named *CnMADS1,* was identified in the endosperm of coconut through the SSH cDNA library. In this paper, functional characterization of the *CnMADS1* gene was carried out by seed-specific overexpression in *A. thaliana* seeds and protoplasts of coconut. The results indicated that in the twelve independent T_2 transgenic *Arabidopsis* lines with high overexpression of *CnMADS1*, the size of the mature seeds of transgenic plants was increased significantly (19.64% increase in the long axis and 8.6% increase in the short axis) compared to that of the wild-type seeds. Moreover, the total lipid content also increased significantly in mature seeds of transgenic plants. After comparing the expression of related genes in wild-type and transgenic plants and confirmation by EMSA, *AtOSR1*, a regulatory gene related to seed size, was proven to be significantly up-regulated by *CnMADS1* in transgenic plants. Moreover, the transient transformation of protoplasts of coconut also proved that *CnLECRK3* (the homologous gene of *AtOSR1* in coconut) is up-regulated by the *CnMADS1* gene in the same way. All these results indicated that a similar regulation mode existed in *Arabidopsis* and the endosperm of coconut and ultimately affected the yield and quality of coconut copra.

Keywords Coconut · Endosperm · MADS-box · Transcription factor · Cell expansion

Introduction

Coconut (*Cocos nucifera* L.), which belongs to the palm family (*Arecaceae*), is one of the only accepted species in the Cocos genus. As one of the world's most useful economical tropical crops, coconut is well known for its industrial and commercial applications (Dasanayaka et al. 2009). This species continues to have hundreds of uses as a source of food, drink, fiber, construction material, charcoal, and oil (used in cooking, pharmaceuticals, industrial applications, and biofuels) (Gunn et al. 2011). Recently, it was shown that over 12 million hectares of coconut are currently planted across 89 tropical countries (Rivera et al. 1999). The mature fruit of coconut takes 12–13 months to develop and accumulates approximately 70–75% oil in its kernel (endosperm). Although most plant lipids contain predominantly C_{16} and C_{18} fatty acids, coconut oil contains over 50%-medium chain fatty acids (MCFAs), including C12 and C14 fatty acids (Laureles et al. 2002; Liang et al. 2014). Due to special components, coconut oil has unique features such as a pleasant odor, high resistance to rancidity, narrow melting temperature range and superior foam retention capacity

Communicated by Anastasios Melis.

RuHao Sun and Li Gao contributed equally to the article.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00425-020-03490-3) contains supplementary material, which is available to authorized users.

✉ Dongdong Li
lidd@hainanu.edu.cn

[1] College of Tropical Crops, Hainan University, Hainan 570228, China

for use in whipped toppings (Harris et al. 2015). However, metabolic pathways and regulatory mechanisms related to fruit development and material metabolism have remained largely unknown due to limited knowledge about related molecular biology research (Zheng et al. 2019). Endosperm development in coconut is a unique and poorly characterized process. Only a few genes and related metabolic processes involved in coconut endosperm development have been investigated (Sun et al. 2017a; Knutzon et al. 1995).

The morphological development and nuclear behavior of the growth-promoting activity of the liquid syncytial endosperm (milk) of coconut is very unique in this remarkable tissue (Fig. 1). More than half a century ago, Cutter et al. (1955, 1952) presented a generalized account of the development of major tissues in ripening coconuts (Dutt 1953). After fertilization and the enlargement of the embryo sac of coconut, the sac cavity is filled with a clear fluid, in which numerous free nuclei of various sizes are suspended. Following this stage, many free spherical cells develop in this fluid and serve as the centers around which these cells are organized. Shortly thereafter, these cells and free nuclei migrate toward the periphery of the sac and adhere to the endothelial surface. Development of the solid endosperm by coalescence of cells and nuclei is initiated at the antipodal end of the sac and proceeds upward toward the micropyle. However, nearly half a century after the completion of this study on the morphology of coconut fruit, little progress has been made toward understanding the molecular biological mechanism of coconut endosperm developments.

MADS-box transcription factors contain a series of the best-studied gene families, and members of this family play essential roles in plant development (Becker and Theissen 2003). Plant MADS-box genes were first identified as regulators of floral organ identity and have since been reported to control additional developmental processes, such as the meristem identification of vegetative, inflorescence and floral meristems; root growth; ovule and female gametophyte development, flowering time; fruit ripening; and dehiscence (Smaczniak et al. 2012). These genes are major components in the well-known "ABCDE model" that depicts their roles in floral organ development (van der Linden et al. 2002; Shan et al. 2009). Based on sequence similarities, the known MADS family genes are classified into separate subfamilies, including the SQUAMOSA (class A), DEFICIENS (class B), GLOBOSA (class B), AGAMOUS (classes C and D), and AGL2-like (class E) gene subfamilies (Dreni and Kater 2014; Li et al. 2019). Recently, different studies revealed functions of MADS-box genes in the morphogenesis of almost all organs and

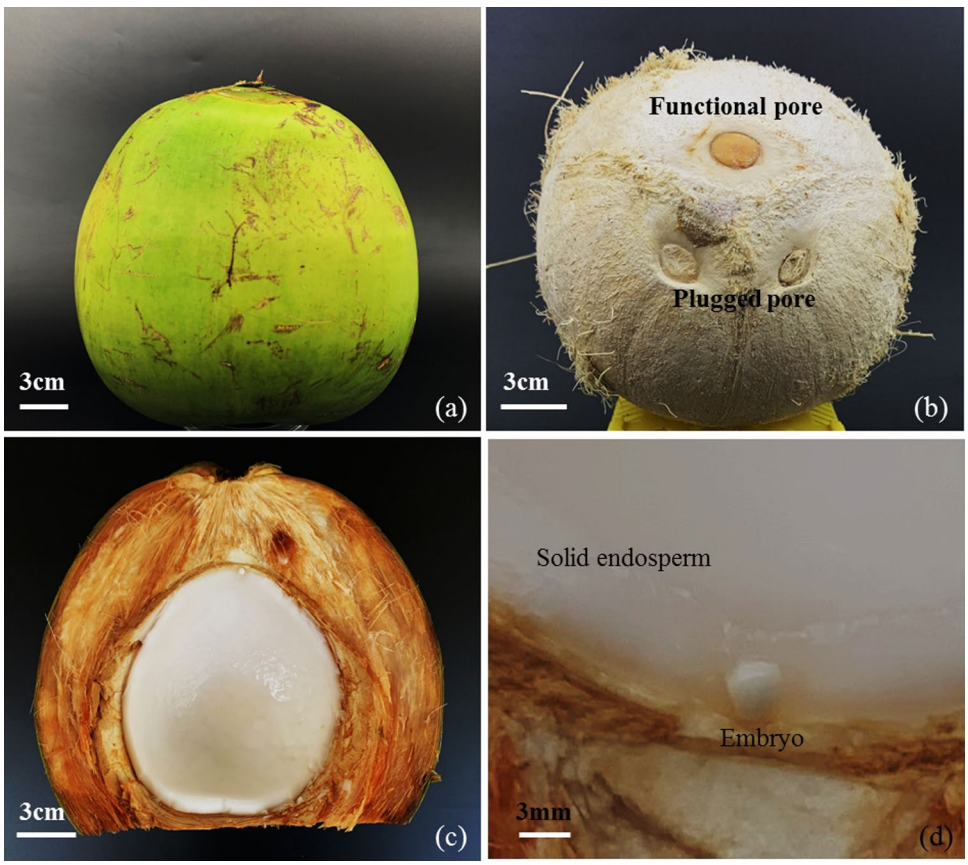

Fig. 1 Integral and dissected mature fruit of oil coconut (*Cocos nucifera* L.). **a** Integral coconut fruit. **b** Fruit remove the exocarp (out layer) and mesocarp (fibrous husk). **c** The longitudinal section of coconut fruit. **d** Embryo and solid endosperm of coconut fruit

throughout the plant life cycle, from embryo to gametophyte development (Yamaguchi and Hirano 2006).

In plants, most of the MADS box genes identified to date (more than 24 in *Arabidopsis* alone) are expressed in flowers; many of these genes affect floral organ identity (Weigel 1995). AGAMOUS (AG) subfamily proteins are encoded by MADS-box family genes, and they have been shown to play key roles in the determination of reproductive floral organs such as stamens, carpels and ovules (Dreni and Kater 2014). In rice, *OsMADS3* appears to be important for stamen identity and stamen development (Yamaguchi et al. 2006); OsMADS29 was presumed to be a key regulator of rice seed development (Nayar et al. 2014), and OsMADS87 is correlated with precocious endosperm cellularization (Chen et al. 2016). In tomato, the AGL15 subfamily gene SlMBP11 is an important regulator involved in vegetative and reproductive developmental processes, and overexpression of SlMBP11 1 promotes bud outgrowth and delays senescence of the perianth organ of the tomato plant (Guo et al. 2017). Recently, the identification of diverse MADS expressed during banana (*Musa acuminata*) ripening raises the questions of whether the function of some MADS box proteins is common to monocots and whether divergence occurs after the separation of dicots and monocots (Liu et al. 2017; Elitzur et al. 2010). In the *Arecaceae* family, the AG-like and GLO-like MADS box genes have been found to be involved in the maturation and ripening of the oil palm mesocarp (Tranbarger et al. 2011). However, due to the lack of data, it is unclear whether similar or diverse regulatory mechanisms function during maturation and ripening in palm fruit in general and oil-accumulating fruit tissues, especially in coconut endosperm.

In this paper, an *AGL* MADS-box transcription factor, named *CnMADS1* (GenBank: KF574389), which was identified from an SSH cDNA library of the coconut endosperm in previous research (Liang et al. 2014), was cloned and overexpressed in transgenic *Arabidopsis* plants under the control of a seed-specific promoter (*napin*). Phenotypic analysis found that transgenic plants showed normal growth and development, but the size and oil content of mature seeds were significantly and simultaneously increased compared to those of wild-type *Arabidopsis*. Moreover, the transient transformation of coconut protoplasts also proved that the homologous genes of *OSR1* in coconut are also up-regulated by the *CnMADS1* gene in the same way. To the best of our knowledge, this is the first report about the functional characterization of the MADS-box transcriptional factor from coconut. As a result, this study provides new insight into the regulatory mechanism of MADS-box genes and their role in coconut endosperm development and will lay the groundwork for genetic improvement in the future.

Materials and methods

Plant materials

The coconut pulp used in this research was obtained from plants grown in the field at the Coconut Research Institute, Chinese Agricultural Academy of Tropical Crops, Hainan, P.R. China, and was collected approximately 7 months after fertilization (MAF), 9 MAF and 11 MAF. The pulp was immediately frozen in nitrogen and stored at − 70 °C after collection. *Arabidopsis thaliana* and transgenic plants were grown under normal greenhouse conditions (22 ± 1 °C, 16 h of light/8 h of dark and 50–70% relative humidity). Plants were illuminated with cool-white fluorescent lights. Seedlings were cultured under normal conditions after vernalization and sampled at the indicated time points.

Gene cloning, sequence alignment and phylogenetic analysis

Total RNA was extracted from coconut pulp using hexadecyltrimethyl ammonium bromide-lithium chloride (CTAB-LiCl) methods (Li 2008). RNA was used for cDNA synthesis by the TIANScript FastQuant RT kit (Tiangen, Beijing, China) following the manufacturer's instructions. PCR was performed with primers *CnMADS1* (Supplemental Table 1). Nucleotide and amino acid sequence analyses were performed on the BLAST program from the NCBI website and ORF Finder2. Clustal Omega3 and MEGA 5.0 were used for amino acid alignment and construction of the phylogenetic tree (Tamura et al. 2007).

Subcellular localization of *CnMADS1*

The coding sequence of *CnMADS1* without the stop codon was cloned into the p35MK1-GFP vector to generate the CnMADS1-GFP fusion expression vector (Fig. S3). The recombinant vector was introduced into Agrobacterium strain GV3101 for further transient transformation. *N. benthamiana* (1-month-old) leaves were infected with the Agrobacterium strain containing the recombinant gene. After 36 h, GFP fluorescence was observed using a confocal laser scanning microscope (TCS SP8, Leica, Heidelberg, Germany).

Plasmid construction and genetic transformation

All primers used for plasmid construction are presented in Supplemental Table 1. The PCR products were cloned into the vector *pMD19-T* (TaKaRa). The cloned *CnMADS1* used for plant transformation was double-digested with *Bam*HI

and *Eco*RI and then ligated into the seed-specific expression vector *pCAMBIA1300s* with the promoter *napin* from *Brassica napus* L (Fig. S1). The constructed *pCAMBIA1300s-CnMADS1-napin* was introduced into *Agrobacterium tumefaciens* strain EHA105 by electroporation. *Arabidopsis thaliana* plants were transformed using a modified floral dip procedure (Jin et al. 2017).

Generation of transgenic plants

Transformed plants were selected by germinating surface-sterilized T_0 seeds on MS medium containing 50 mg/l hygromycin. Resistant seedlings were transplanted to soil after 15 days of growth. Seeds from resistant plants were collected and screened by hygromycin resistance and PCR analysis in this generation. The transgenic plants were classified as T_1 plants. Subsequently, the identified T_1 lines were further self-pollinated to obtain the final T_2 seeds, and the resultant seeds were used as the plant material source for further experiments.

Analysis of gene expression by qRT-PCR

Total RNA was extracted from the fresh seeds of T_2 *A. thaliana* using the RNAprep pure plant kit (Tiangen, Beijing, China) and reverse transcribed to generate cDNA using the TIANScript FastQuant RT kit (Tiangen, Beijing, China) according to the manufacturer's instructions. RT-PCR was used to detect seed-specific expression levels, and real-time quantitative PCR was used to determine the relative gene expression levels in *A. thaliana. β-Actin* of *A. thaliana* was used as a reference control. All the gene-specific primer pairs are shown in Supplemental Table 1. Real-time qPCR was used to quantify the expression levels of seed-size-associated genes in transgenic *A. thaliana*. A 20 µl qPCR reaction contained cDNA that was reverse transcribed from 50 ng of total RNA, specific primers with a final concentration of 100 nM each and 10 µl of SYBR Premix Ex TaqII (TaKaRa, Dalian, China). Real-time qPCR was conducted in a LightCycler® 480 system (BIO-RAD, USA) following the program of predenaturation for 3 min at 95 °C and 40 cycles of 10 s at 95 °C, 30 s at 60 °C and 1 min at 72 °C. Relative expression levels for all the genes under the indicated conditions were calculated using the $2^{-\Delta\Delta Ct}$ method. Three biological samples and triplicate qPCRs for each combination of primers and sample were analyzed.

Light microscopy

Seed embryo observation was performed as described previously (Ohto et al. 2005; Chen et al. 2016). Isolated embryos from the WT and overexpression lines were incubated in buffer (50 mM sodium phosphate, 10 mM EDTA, 1% Triton X-100, and 1% dimethyl sulfoxide) at 37 °C overnight, fixed in a formaldehyde-acetic acid solution (10% formalin, 5% acetic acid, 45% ethanol, and 0.01% Triton X-100) for 45 min, and then rehydrated using a graded ethanol series. Embryos were then cleared in Hoyer's solution (chloral hyd rate:water:glycerol = 3:0.8:0.4) and observed with a light microscope (ZEISS Lab. A1).

Analysis of lipid and fatty acid contents

Seed oil was extracted from mature seeds derived from each of the transgenic lines and WT seeds of *Arabidopsis thaliana* using trichloromethane/methanol (2:1), and the relative weight percentage of seed oil was obtained after air-drying the extract (Yuan et al. 2017). The extract was transmethylated with 2% KOH–methanol (m/v) at 80 °C for 2 h. The fatty acid methyl esters (FAMEs) were recovered using n-hexane. All GC–MS analyses were performed using an HP5890 GC instrument equipped with a BPX-70 (30 m × 0.25 mm) chromatographic column. The initial column temperature (90 °C) was held for 10 min and then raised at 4 °C/min until it reached 240 °C, after which it was held at this temperature for another 10 min.

Transcriptome analysis of WT and T2 transgenic plants

Total RNA of mature seeds from WT and T2 transgenic plants was extracted and purified using an RNAprep pure plant kit (Tiangen, Beijing, China). RNA samples were submitted to the Beijing Genomics Institute (BGI, Shenzheng, China), where library preparation and high-throughput sequencing services were performed. The expression level of each unigene was estimated by FPKM. Differentially expressed gene analysis of WT and transgenic plants was calculated using the cuffdiff program, filtering DEGs of significance (FDR < 0.05).

Electrophoretic mobility shift assay (EMSA)

The full-length coding sequence of *CnMADS1* was cloned into the prokaryotic expression vector pCold-TF (Takara) (Fig. S4). This construct was transformed into competent *Escherichia coli* BL21 (DE3) cells. The recombinant protein was generated by induction with IPTG (0.1 mM) for 24 h at 16 °C. The purified protein was obtained using Ni–NTA agarose columns. In addition to recombinant proteins, biotin-labeled probes (Sangon Biotech, Shanghai, China) were prepared for EMSA. The fusion between protein and probes was generated using binding reactions prepared according to the LightShift chemiluminescent EMSA kit (Thermo Fisher Scientific, USA). Mutated probes and unlabeled probes were added 200-fold relative to the labeled probe as

cold competitors. The free and bound probes were separated by polyacrylamide gel electrophoresis at a low temperature using 0.5 × TBE buffer. Then, the reaction was completed by film transfer, crosslinking and detection.

Protoplast transformation and qRT-PCR

The young leaves of coconut seedlings were cut into 0.5–1.5 mm strips and dispersed in 10 ml of enzyme solution (1.5% (w/v) CellulaseR-10, 0.4% (w/v) MacerozymeR-10, 0.4 M mannitol, 20 mM KCl, 20 mM MES, 10 mM CaCl2 and 0.1% (w/v) BSA). Then, enzymatic digestion was performed in the dark for 4 h at room temperature with gentle shaking. The digestion mixture was filtered through a stainless 45 µm-mesh sieve and centrifuged at 800 rpm for 5 min. The filtrate was re-suspended and washed twice with W5 solution (154 mM NaCl, 125 mM CaCl2, 5 mM KCl and 2 mM MES, pH 5.8) and then resuspended in 1 ml of MMG solution (0.4 M Mannitol, 15 mM $MgCl_2$ and 4 mM MES, pH 5.8).

For protoplast transformation, the *CnMADS1* overexpression plant vector was successfully obtained by cloning the ORF of *CnMADS1* into *pCAMBIA1300s* (Fig. S2). Two hundred microliters of protoplasts was mixed with 20 µg (10 µl) of *pCAMBIA1300s* and *pCAMBIA1300s-CnMADS1* plasmids. Then, 210 µl of PEG solution (40% (w/v) PEG 4000, 0.4 M mannitol and 100 mM $CaCl_2$) was added, and the mixture was incubated in the dark for 30 min. The protoplasts were washed twice with W5 solution, resuspended in 1 ml of W5 solution and then incubated under weak light at 25 °C for 16 h. The transfected protoplasts were collected for RNA extraction and RT-PCR analysis to determine the regulatory relationship between *CnMADS1* and *CnLECRK3* (the homologous serine/threonine-protein kinase of coconut). All the gene-specific primer pairs are shown in Supplemental Table 1.

GenBank accession numbers for related genes

AtAGL39/AT5G27130, *AtAGL23*/AT1G65360, *AtAGL28*/AT1G01530, *IAA31*/AT3G17600, *AtOSR1*/AT3G59900, *GRP6*/AT1G74670, *ERF039*/AT4G16750, *ERF110*/AT5G50080, *SNF1*/AT4G40010, *KCS18*/AT4G34520, *MADS*/AT1G17310, *GDSL*/AT5G22810, *GAM2*/AT5G56300, *ACP-De*/AT5G16240, *ACP1*/AT3G05020, *MYB75*/AT1G56650, *CnLECRK3*/CCG026298.1

Statistical analyses

One-way ANOVA was used to assess the differences between genotypes for measurements of seed percentage oil content and seed weight. Following significant ($P < 0.05$) F test results, means of interest were compared using the appropriate LSD value at the $P = 0.05$ level of significance on the corresponding degrees of freedom.

Results

Gene clone and sequence analysis of *CnMADS1*

An *AGL* MADS-box gene with up-regulated expression during ripening was derived from the coconut endosperm SSH cDNA library (Liang et al. 2014). The full-length cDNA was named *CnMADS1* and isolated from coconut endosperm by rapid amplification of cDNA ends (RACE). Sequence analyses revealed that *CnMADS1* was 1007 bp in length and contained a 714 bp ORF encoding a single 238-amino acid (AA) peptide, with a predicted molecular mass of 27.45 kDa and a pI of 9.52. *CnMADS1* has a 57-aa MADS domain region in the N-terminus and an 88-aa K domain (Fig. 2a). Phylogenetic analysis using a neighbor joining (NJ) tree with 1000 bootstrap replications indicated that *CnMADS1* was grouped into the Mα subclass of the type I lineage MADS-domain protein (Fig. 2b). The NJ tree was constructed with 8 other *AGL* genes from *Arabidopsis thaliana* (*AtAGL39*/AT5G27130, *AtAGL23*/AT1G65360, *AtAGL28*/AT1G01530 of Mα, *AtAGL92*/AT1G31640, *AtAGL86*/AT1G31630 of Mβ and *AtAGL82*/AT5G58890, *AtAGL49*/AT1G60040, *AtAGL50*/AT1G59810 of Mγ). However, the amino acid sequence of *CnMADS* shared low identities with the other three *MαAtAGLs* (Fig. 2c).

CnMADS1 localizes to the nucleus

To examine the localization pattern of *CnMADS1* in plants, we generated recombinant CnMADS1-GFP and transferred it into the leaves of *N. benthamiana* for transient expression. Fluorescence microscopy analysis indicated that GFP fluorescence overlapped with the nuclei stained by 4′,6-diamidino-2-phenylindole (DAPI). These data indicated that CnMADS1 localized to the nucleus (Fig. 2d).

Ectopic expression of *CoMADS1* in *Arabidopsis*

The plant expression vector containing *CnMADS1* driven by the seed-specific promoter was sequenced and transferred into *Agrobacterium tumefaciens EHA105* for *Arabidopsis* transformation. After gene transformation and positive plant selection on MS medium, 16 independent transgenic plants were obtained. Subsequently, eight positive transformants were detected by PCR analysis using the DNA obtained from leaves. These positive transformants were self-pollinated to produce T_1 lines. Eight T_1 transgenic plants for each positive family were grown for segregation of character. The mature seeds from positive transgenic plants were collected

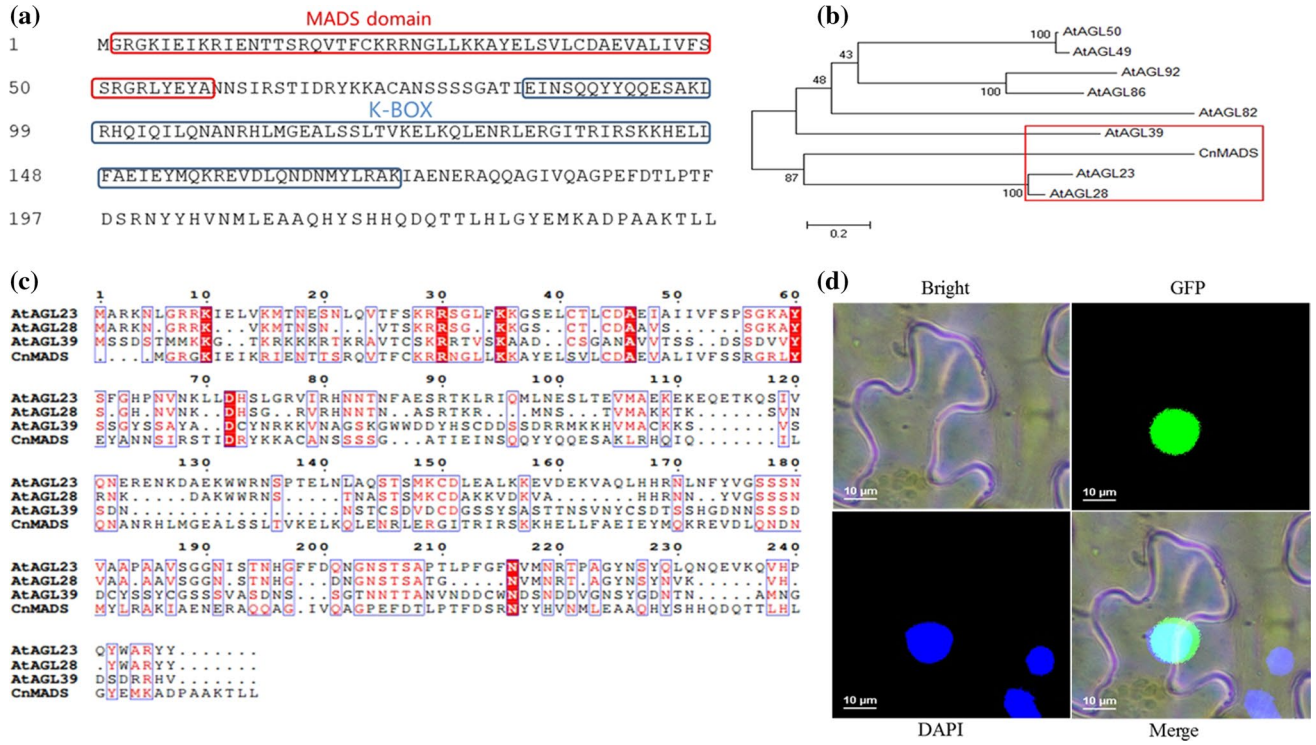

Fig. 2 Sequence analysis and subcellular localization of *CnMADS1*. **a** The amino acid sequence of *CnMADS1*, MADS and K-BOX domains are indicated. **b** Neighbor-joining phylogenetic tree for *CnMADS* and Arabidopsis MADS-domain proteins by Mega 5.0 using default parameters. **c** Comparison of deduced amino acid sequences of *CnMADS*1with type I MADS-domain from *Arabidopsis* by Clustal X 2.0. **d** Bright-field, GFP, DAPI and merged images of the *CnMADS1* protein fused with GFP in *N. benthamiana* leaves

separately after PCR analysis. After the seed germination test, homozygous transgenic lines were selected for subsequent analysis. Ectopic expression of *CnMADS1* in 20 independent T_2 transgenic lines was analyzed by quantitative real-time PCR (qRT-PCR). Different levels of expression of *CnMADS1* were detected in 12 overexpression transgenic (*OE*) lines but not in the negative control. Among the lines, line 11 possessed the highest expression level, which was approximately 528.5-fold that of line 6 (Fig. 3b).

Phenotypic analysis of transgenic *Arabidopsis*

During the experimental process, an unexpected phenomenon was observed, in which the size of seeds from mature transgenic plants was obviously increased compared with that of the wild-type seeds (Fig. 3a). To further elucidate the change in seed size, we first measured the long axis and short axis of approximately 50 grains of transgenic *Arabidopsis* seeds obtained from all the top three high expression level homozygous *OE* lines. The results showed that the sizes of seeds from transgenic plants were significantly increased, with a 19.64% rise in the long axis and an 8.86% rise in the short axis compared to those of seeds from the same batch of wild-type *Arabidopsis* plants (Fig. 3c). Then, we observed the cells of seeds randomly selected from the wild-type and *OE* lines under light microscopy. This result indicated that the cells from different parts of the embryo of the *OE* line 11 were larger than those of the wild-type plant (Fig. 4).

Lipid and fatty acid content analysis in transgenic *Arabidopsis*

To observe the effects of *CnMADS1* overexpression in *OE* lines, the top three *OE* lines and independent T_2 transgenic lines were analyzed for seed lipid content. Analysis of seed lipid content in representative *OE* lines indicated that all three lines had a significant increase ($P < 0.01$) of approximately 50% compared to wild-type plants (Fig. 5a). Further analysis of the fatty acid composition via GC–MS revealed that the contents of palmitic acid (C16:0), oleic acid (C18:1) and eicosanoic acid (C20:1) were significantly increased ($P < 0.01$) in seeds of the transgenic lines compared to their contents in the seeds of wild-type plants, while the arachidic acid (C20:0) content was reduced significantly ($P < 0.01$). In all three lines, the contents of all the other kinds of fatty acids that we detected showed no significant change (Fig. 5b).

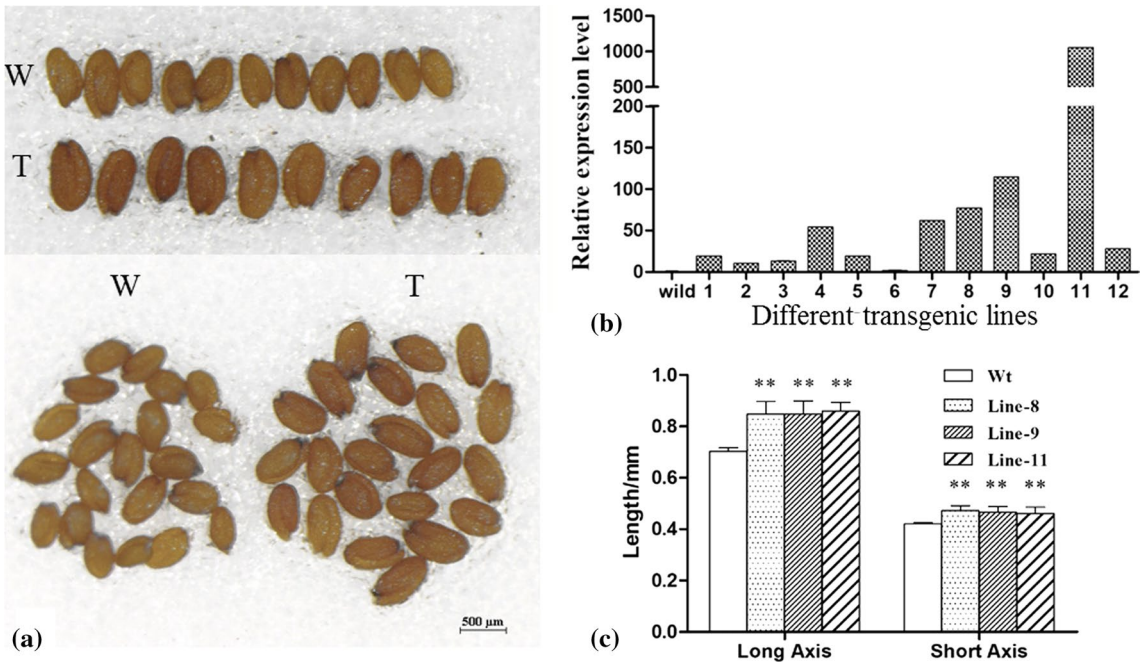

Fig. 3 Observation and measurement of the mature seeds. **a** Observation of the mature seeds under the stereoscopic microscope. W: Wild-type. T: Transgenic. **b** Relative expression levels of *CnMADS1* in different T_3 generation transgenic lines **c** Measurement of the mature seeds. Randomly picked 50 grain from wild-type and transgenic lines (line 8, 9, 11) separately, and deleted 10 most different data. Values are the means and sd of 50 distinct individual with $^{**}p < 0.01$

RNA-seq analysis of transgenic plants

To explain the mechanism underlying the effects of *CoMADS1* in regulating seed size and lipid accumulation, we screened for downstream genes regulated by RNA-seq. These results indicated that ectopic expression of CoMADS1 can cause differential and specific gene regulation in seeds of *OE* lines. Compared to WT plants, 545 genes were up-regulated, and 330 genes were down-regulated in the OE line ($Log_2FC > 1$ and $FDR < 0.05$) according to the RNA-Seq results (Fig. 6a). Moreover, Gene Ontology (GO) enrichment and KEGG pathway enrichment analyses showed that differentially expressed genes (DEGs) tended to take part in metabolism (especially in lipid and carbohydrate metabolism), genetic information processing and cellular processes (Fig. 6b). Based on this information, we focused on DEGs with functions involved in the cell proliferation/expansion process and fatty acid accumulation (Supplemental Table 2). The results indicated that several genes associated with organ size was up-regulated (Fig. 6c), including *ORGAN SIZE RELATED1 (OSR1)*, which regulates cell expansion during organ growth (Feng et al. 2011); gibberellin (GA)-regulated proteins, which lead to a significantly larger size of mature embryos (Hu et al. 2018); and auxin-responsive proteins (IAA31), which have specific domains that interact with auxin response factor (ARF) and inhibit the transcription of genes activated by ARF (Luo et al. 2018).

Validation of regulatory relationship by EMSA

To test the interaction between the *CnMADS1* protein and the *AtOSR1* promoter, we analyzed the relative expression levels of *CoMADS1* and *AtOSR1* in transgenic *Arabidopsis* plants. The results indicated that the expression level of *AtOSR1* in the seeds of transgenic *Arabidopsis* plants was up-regulated significantly compared to that in wild-type plants (Fig. 7a). Moreover, the promoter region of *AtOSR1* and the binding sites for the MADS box were also validated. DNA probes containing the SRF element and mutant *SRF* element were designed according to the elements in the red (Fig. 7b) and incubated with purified pColdTF-*CnMADS1*-His protein for the electrophoretic mobility shift assay (EMSA). As expected, the purified pColdTF-*CnMADS1*-His fusion protein bound to biotin-labeled probes and shifted a band. As a control, no shift was observed when pColdTF-His protein was incubated with biotin-labeled probes (Fig. 7b). There was no effect on the binding capacity of pColdTF-*CnMADS1*-His to the labeled DNA probes when an excess of mutated unlabeled DNA probe was added. In contrast, when unlabeled promoter fragments were used as competitors, the mobility shift was efficiently inhibited. This result demonstrated that *CnMADS1* targets the promoter of the *AtOSR1* gene and positively regulates seed size.

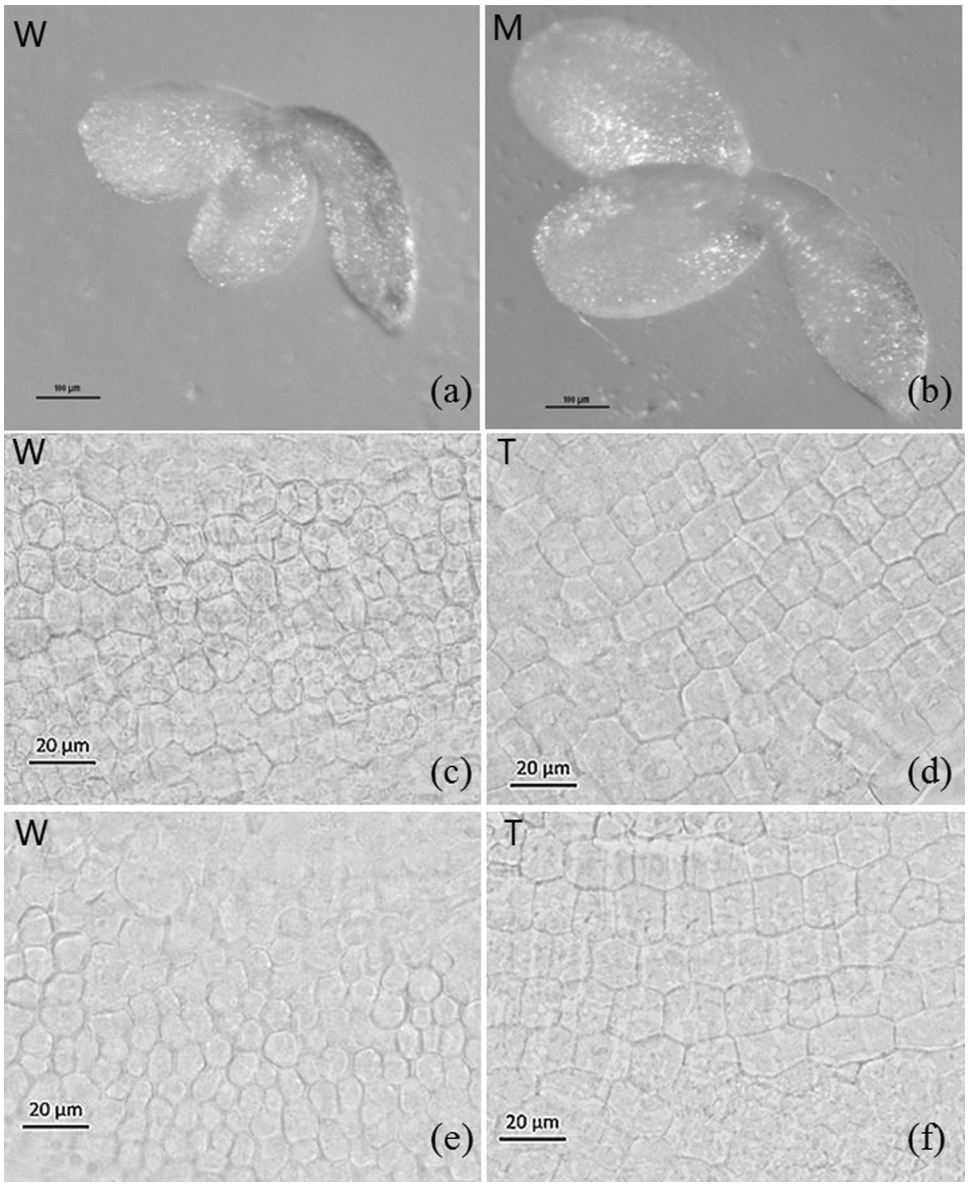

Fig. 4 Observation of the mature seeds under the light microscope. **a**, **c** and **e** are the wild-type seeds. **b**, **d** and **f** are the Line 11 seeds. **a**, **b** are the whole seed without seed coat with the same scale (100 μm), **c**–**f** are the different part of the embryo with the same scale (20 μm)

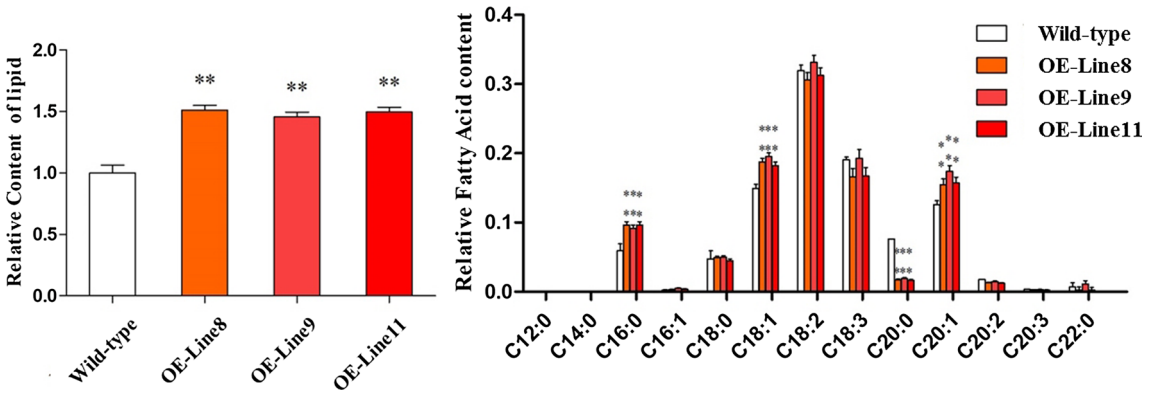

Fig. 5 Lipid content and relative fatty acid (FA) content of the mature T_3 generation *Arabidopsis thaliana* seeds. **a** Lipid content of wild-type and lines 8, 9, 11. Values are the means and sd of three replicates of dry mature seeds with $^{**}p < 0.01$. **b** Relative fatty acid content of wild-type and lines 8, 9, 11 were extracted from around 20 mg dry seeds and determined by GC–MS analysis. Values are the means and sd of 3 replicates from 3 biological individuals with, $^{**}p < 0.01$

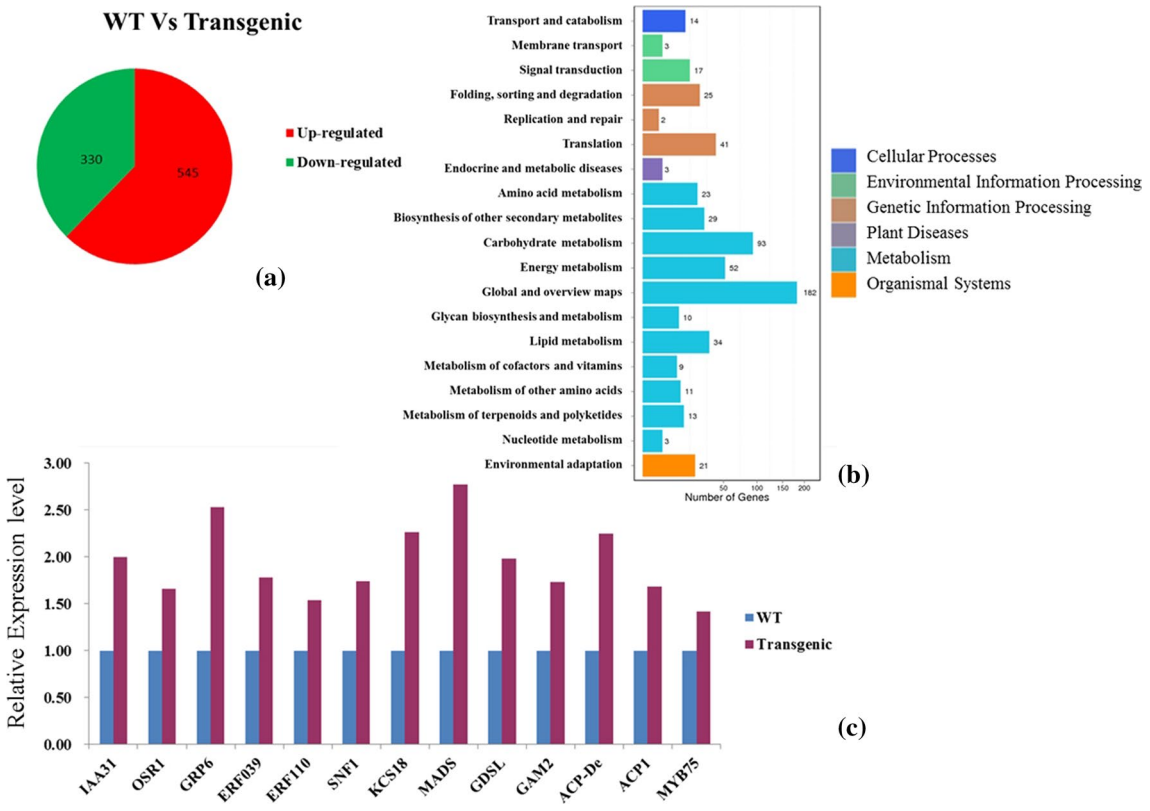

Fig. 6 RNA-Seq analysis of gene expression changes in WT and Transgenic plants. **a** Up- and down-regulated genes in *CnMADS1* overexpression transgenic *Arabidopsis*. **b** KEGG study of up- and down-regulated genes including significantly enriched pathway. **c** mRNA expression levels of selected genes with differential expression levels. The values are Log_2 (fold change) of expression level of transgenic plant to WT

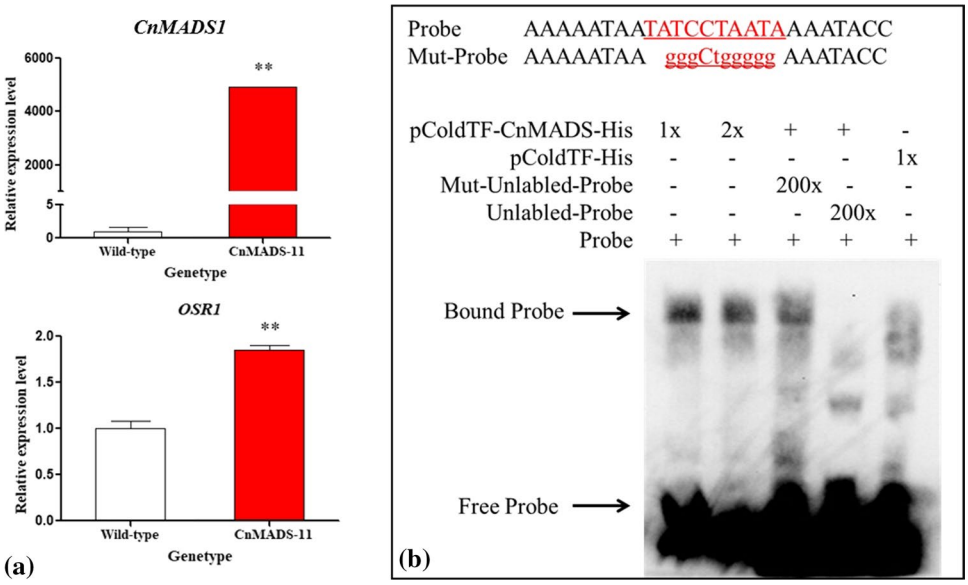

Fig. 7 Screening and identification of downstream regulatory gene of *CnMADS1* in transgenic plant. **a** Relative expression level of *CnMADS1* and *OSR1* in transgenic *Arabidopsis* plants. **b** Verification of binding of *CnMADS1* to the motif in the promoter of the *OSR1* gene by EMSA

CnMADS1 positively regulates the expression of CnLECRK3

To better understand the function of *CnMADS1* in coconut, *CnLECRK3,* the homologous gene of *AtOSR1*, was identified from the cDNA library of the coconut endosperm. Both of these genes belong to the serine/threonine-protein kinase family. Protein sequence alignment analysis showed that *CnLECRK3* and *AtOSR1* share a conserved domain of the serine/threonine kinases (Fig. 8a). Moreover, using the PROMO website, the SRF binding sites for the MADS-box gene were identified on both promoter sequences of *CnLECRK3* and *AtOSR1* (Fig. 8b). To confirm the regulatory relationship between *CnMADS1* and *CnLECRK3*, the overexpression vector *pCAMBIA1300s-CnMADS1* was introduced into coconut protoplasts via PEG-mediated transformation. After incubation, the protoplasts were collected for qRT-PCR analysis. As expected, the overexpression of *CnMADS1* significantly up-regulated the expression of *CnLECRK3* (Fig. 8c). Therefore, all the results suggested that there should be a conservative regulatory pathway that regulates the serine/threonine kinase gene *(CnLECRK3)* by the MADS-box gene (*CnMADS1)*, finally affecting the cell proliferation and/or cell expansion process in the endosperm of coconuts (Fig. 9).

Discussion

As one of the key plantation crops of the tropics, the current distribution of coconut spans the coastal areas between 20° north and 20° south of the equator and ranges in altitude from sea level to 1000 m (Gunn et al. 2011). Coconut is also the most widely grown plantation crop and is one of the three major plantation crops on Hainan Island of China. However, coconut cultivation is currently in relative decline in many countries due to strong competition from other oil crops, such as oil palm, low productivity, increased demand

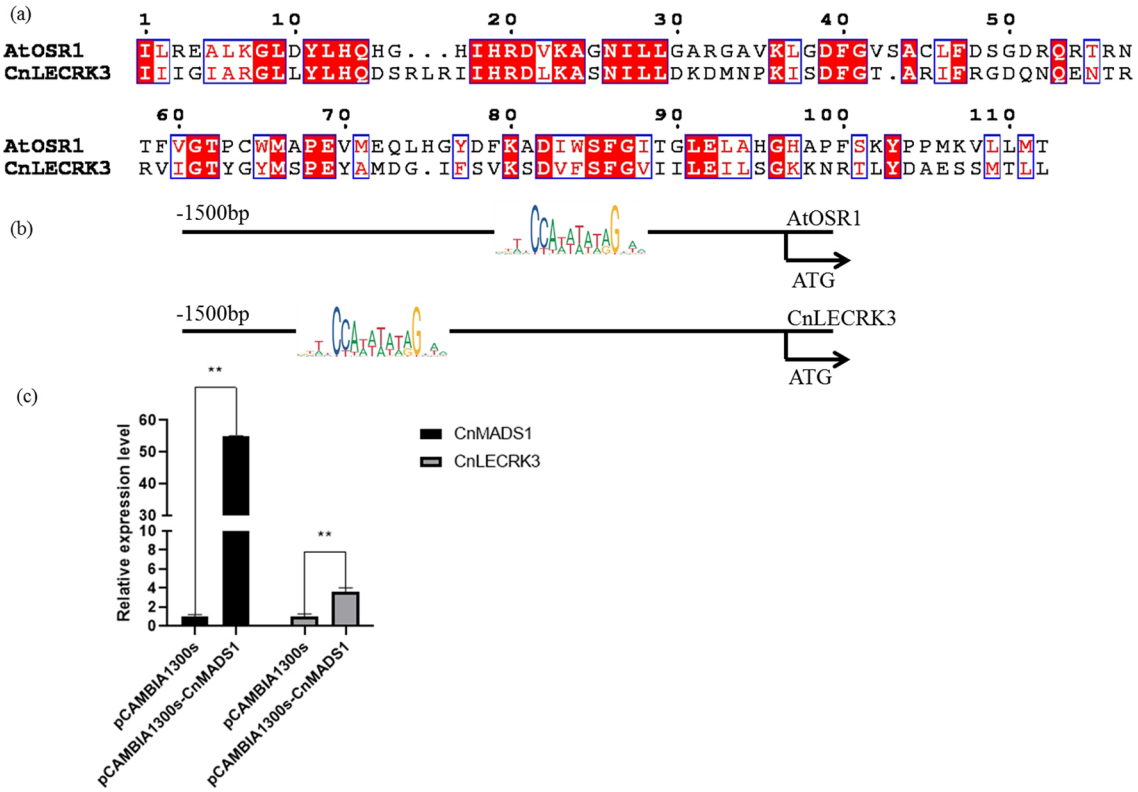

Fig. 8 Verification of *CnMADS*s and *AtOSR1* homologous gene (CnLECRK3) in protoplasts coconut. **a** Protein sequence alignment analysis. **b** Binding sites SRF on promoter of both genes. **c** Overexpression of *CnMADS1* in the protoplasts of coconut

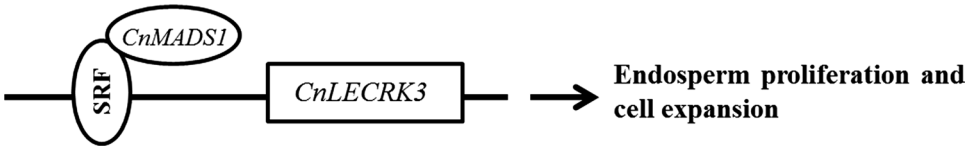

Fig. 9 A working model proposed for *CnMADS1* function in endosperm development of coconut

for timber and devastating diseases (Sandoval-Cancino et al. 2016; Nejat et al. 2015). Compared to that on other important crops, research on coconut has evolved slowly, especially with respect to the transcriptional regulation of fruit and endosperm development. In past research, most of the studies were related to the synthesis and accumulation of medium-chain fatty acids in coconut endosperm (Yuan et al. 2017, 2015; Zheng et al. 2019). Only one transcription factor related to oil metabolism, WRI1, was cloned and identified from coconut endosperm development (Sun et al. 2017b). Therefore, the MADS-box gene obtained in our present study represents a novel gene that has never been reported in the endosperm of coconut.

Increasing evidence shows that during the development of plants, the MADS-box family is involved not only in the process of flower transformation and floral organ development but also in the development of the endosperm and material accumulation (Smaczniak et al. 2012). In *Arabidopsis thaliana*, the MADS-box gene ARABIDOPSIS BSISTER (ABS) affects the coordination of cell divisions in ovules and endosperm formation by repressing SHP1, SHP2, and FRUITFUL expression (Ehlers et al. 2016). In rice (*Oryza sativa*), OsMADS6 plays an essential role in endosperm starch filling and epigenetically regulates trimethylated H3K27. ADP-glucose pyrophosphorylase genes, encoding the rate-limiting step enzyme in the starch synthesis pathway, are subject to the regulation of OsMADS6 (Zhang et al. 2010). Currently, although a series of MADS-box family transcription factors have been identified during the development of fruit in palmitic plants by transcriptomic analysis (Tranbarger et al. 2011), there are few reports on the related metabolic regulation pathways and functions in endosperm (Ehlers et al. 2016). However, in contrast to previous references, the MADS-box gene cloned in this paper significantly increased the seed size and oil content of transgenic plants by regulating the expression of the *ORGAN SIZE RELATED1 (OSR1)* gene in *Arabidopsis* (Feng et al. 2011). Although it is the result of ectopic expression in model plants, the obtained results suggest that similar regulatory mechanisms may exist in the development of the coconut endosperm.

Generally, for mature seeds with an endosperm, seed weight strongly depends on the development of the endosperm, while seed size is closely related to the formation of the seed coat (Zhang et al. 2016). To date, a number of gene and transcriptional factors that control organ size by regulating the cell proliferation and/or cell expansion process have been identified in different kinds of plants. For instance, the AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS) gene from *Arabidopsis* transduces the auxin signal to affect the cell proliferation of organs, thus controlling organ size (Hu et al. 2006). ARGOS-LIKE (ARL), which has some sequence homology to the ARGOS gene, is also involved in cell expansion during organogenesis (Hu et al. 2003). MINISEED3 (MINI3), a transcription factor of the WRKY class, regulates the seed size of *Arabidopsis* via positively regulating IKU2 by binding to the W-box motif in the IKU2 promoter (Luo et al. 2005). The ORGAN SIZE RELATED (OSR) family was characterized in the regulation of aerial organ size in *Arabidopsis*. Among these genes, OSR1 promoted organ growth by affecting both cell proliferation and expansion, while OSR2 participates in the regulation of cell expansion during organogenesis and thus organ size (Qin et al. 2014). However, no references have shown that the MADS-box gene has a significant effect on seed size thus far.

As a perennial palm family plant, coconut still exhibited a very weak in vitro culture and genetic transformation system. At present, there are no successful case reports related to coconut genetic transformation. In addition, the environmental factors and developmental cycle of the coconut endosperm are very different from those of other types of plants. In particular, coconut water has high salinity, with especially high concentrations of potassium, malate and chloride. Therefore, the coconut endosperm grew in a unique environment of high salinity, hyperosmotic stress, hypothermia, low light and oligotrophic qualities (Jayalekshmy et al. 1986). It can be speculated that the endosperm of coconut undergoes very special mechanisms during growth and development regulation. Therefore, although a certain number of endosperm-specific expressed genes have been screened in previous studies (Liang et al. 2014), determining the role and function of these genes during the development of coconut endosperm is still an urgent problem to be solved. Although there are obvious genetic background differences, the functional characterization of genes from coconuts can only be studied with the help of genetic transformation systems of *Arabidopsis* and other model plants.

In conclusion, the results presented here identified and proved that one MADS-box transcriptional factor, *CoMADS*1, which is expressed specifically in the endosperm of coconut, will increase the seed size and lipid content though upregulation of the *OSR1* gene in transgenic *Arabidopsis*. Although the functional validation in this study was performed in model plants with different genetic backgrounds, the obtained results suggest that a similar regulatory mechanism may operate in coconut. Taken together, the results obtained in this study could provide a reference to elucidate the molecular regulation mechanism of *CoMADS1* in the coconut endosperm and provide insight into the coloration of other palmitic plants.

Author contribution statement DL, YZ contributed to the experimental design. RS and LG carried out mainly experiments, drafted the manuscript, and analyzed data from all

experimental results. ZM prepared the coconut figures. RS and LG contributed equally to the article.

Acknowledgements This research was supported by the Hainan Provincial Natural Science Foundation of China (No. 2019CXTD397), the National Natural Science Foundation of China (NSFC) (No. 31660222) and National Key R&D Program of China, 2018YFD1000500.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Becker A, Theissen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 29(3):464–489. https://doi.org/10.1016/s1055-7903(03)00207-0

Chen C, Begcy K, Liu K, Folsom JJ, Wang Z, Zhang C, Walia H (2016) Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. Plant Physiol 171(1):606–622. https://doi.org/10.1104/pp.15.01992

Cutter VM Jr, Wilson KS, Dube GR (1952) The isolation of living nuclei from the endosperm of *Cocos nucifera*. Science 115(2977):58–59. https://doi.org/10.1126/science.115.2977.58

Cutter VM, Wilson KS, Freeman B (1955) Nuclear behavior and cell formation in the developing endosperm of *Cocos nucifera*. Am J Bot 42(2):7

Dasanayaka PN, Everard JMDT, Karunanayaka EH, Nandadasa HG (2009) Analysis of coconut (*Cocos nucifera* L.) diversity using microsatellite markers with emphasis on management and utilisation of genetic resources. J Natl Sci Found Sri 37(2):99–109

Dreni L, Kater MM (2014) MADS reloaded: evolution of the AGAMOUS subfamily genes. N Phytol 201(3):717–732. https://doi.org/10.1111/nph.12555

Dutt M (1953) Dividing nuclei in coconut milk. Nature 171(4357):799–800. https://doi.org/10.1038/172799a0

Ehlers K, Bhide AS, Tekleyohans DG, Wittkop B, Snowdon RJ, Becker A (2016) The MADS box genes ABS, SHP1, and SHP2 Are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in *Arabidopsis thaliana*. PLoS ONE 11(10):e0165075. https://doi.org/10.1371/journal.pone.0165075

Elitzur T, Vrebalov J, Giovannoni JJ, Goldschmidt EE, Friedman H (2010) The regulation of MADS-box gene expression during ripening of banana and their regulatory interaction with ethylene. J Exp Bot 61(5):1523–1535. https://doi.org/10.1093/jxb/erq017

Feng G, Qin Z, Yan J, Zhang X, Hu Y (2011) Arabidopsis ORGAN SIZE RELATED1 regulates organ growth and final organ size in orchestration with ARGOS and ARL. N Phytol 191(3):635–646. https://doi.org/10.1111/j.1469-8137.2011.03710.x

Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. PLoS ONE. https://doi.org/10.1371/journal.pone.0021143

Guo X, Chen G, Naeem M, Yu X, Tang B, Li A, Hu Z (2017) The MADS-box gene SlMBP11 regulates plant architecture and affects reproductive development in tomato plants. Plant Sci 258:90–101. https://doi.org/10.1016/j.plantsci.2017.02.005

Harris M, Fryda L, Hutchins A (2015) The impact of coconut and safflower oils on body composition. Faseb J 29:747

Hu Y, Xie Q, Chua NH (2003) The *Arabidopsis* auxin-inducible gene ARGOS controls lateral organ size. Plant Cell 15(9):1951–1961. https://doi.org/10.1105/tpc.013557

Hu Y, Poh HM, Chua NH (2006) The *Arabidopsis* ARGOS-LIKE gene regulates cell expansion during organ growth. Plant J 47(1):1–9. https://doi.org/10.1111/j.1365-313X.2006.02750.x

Hu Y, Zhou L, Huang M, He X, Yang Y, Liu X, Li Y, Hou X (2018) Gibberellins play an essential role in late embryogenesis of *Arabidopsis*. Nat Plants 4(5):289–298. https://doi.org/10.1038/s41477-018-0143-8

Jin YH, Yuan YJ, Gao LC, Sun RH, Chen LZ, Li DD, Zheng YS (2017) Characterization and functional analysis of a type 2 diacylglycerol acyltransferase (DGAT2) gene from oil palm (*Elaeis guineensis* Jacq.) mesocarp in *Saccharomyces cerevisiae* and transgenic *Arabidopsis thaliana*. Front Plant Sci. https://doi.org/10.3389/fpls.2017.01791

Knutzon DS, Lardizabal KD, Nelsen JS, Bleibaum JL, Davies HM, Metz JG (1995) Cloning of a coconut endosperm cDNA encoding a 1-acyl-sn-glycerol-3-phosphate acyltransferase that accepts medium-chain-length substrates. Plant Physiol 109(3):999–1006. https://doi.org/10.1104/pp.109.3.999

Laureles LR, Rodriguez FM, Reano CE, Santos GA, Laurena AC, Mendoza EM (2002) Variability in fatty acid and triacylglycerol composition of the oil of coconut (*Cocos nucifera* L.) hybrids and their parentals. J Agric Food Chem 50(6):1581–1586. https://doi.org/10.1021/jf010832w

Li DD, Fan YM (2008) Construction and characterization of a cDNA library from the pulp of coconut (*Cocos nucifera* L.). Agric Sci China 7(9):1071–1076

Li S, Chen K, Grierson D (2019) A critical evaluation of the role of ethylene and MADS transcription factors in the network controlling fleshy fruit ripening. N Phytol 221(4):1724–1741. https://doi.org/10.1111/nph.15545

Liang YX, Yuan YJ, Liu T, Mao W, Zheng YS, Li DD (2014) Identification and computational annotation of genes differentially expressed in pulp development of *Cocos nucifera* L. by suppression subtractive hybridization. BMC Plant Biol 14:205

Liu J, Zhang J, Zhang J, Miao H, Wang J, Gao P, Hu W, Jia C, Wang Z, Xu B, Jin Z (2017) Genome-wide analysis of banana MADS-box family closely related to fruit development and ripening. Sci Rep 7(1):3467. https://doi.org/10.1038/s41598-017-03897-1

Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. Proc Natl Acad Sci USA 102(48):17531–17536. https://doi.org/10.1073/pnas.0508418102

Luo J, Zhou JJ, Zhang JZ (2018) Aux/IAA gene family in plants: molecular structure, regulation, and function. Int J Mol Sci. https://doi.org/10.3390/ijms19010259

Nayar S, Kapoor M, Kapoor S (2014) Post-translational regulation of rice MADS29 function: homodimerization or binary interactions with other seed-expressed MADS proteins modulate its translocation into the nucleus. J Exp Bot 65(18):5339–5350. https://doi.org/10.1093/jxb/eru296

Nejat N, Cahill DM, Vadamalai G, Ziemann M, Rookes J, Naderali N (2015) Transcriptomics-based analysis using RNA-Seq of the coconut (*Cocos nucifera*) leaf in response to yellow decline phytoplasma infection. Mol Genet Genom 290(5):1899–1910. https://doi.org/10.1007/s00438-015-1046-2

Qin Z, Zhang X, Zhang X, Feng G, Hu Y (2014) The *Arabidopsis* ORGAN SIZE RELATED 2 is involved in regulation of cell expansion during organ growth. BMC Plant Biol 14:349. https://doi.org/10.1186/s12870-014-0349-5

Rivera R, Edwards KJ, Barker JH, Arnold GM, Ayad G, Hodgkin T, Karp A (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. Genome 42(4):668–675

Sandoval-Cancino G, Sáenz L, Chan JL, Oropeza C (2016) Improved formation of embryogenic callus from coconut immature inflorescence explants. Vitro Cell Dev Biol Plant 52(4):367–378. https://doi.org/10.1007/s11627-016-9780-7

Shan H, Zahn L, Guindon S, Wall PK, Kong H, Ma H, DePamphilis CW, Leebens-Mack J (2009) Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. Mol Biol Evol 26(10):2229–2244. https://doi.org/10.1093/molbev/msp129

Smaczniak C, Immink RG, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development 139(17):3081–3098. https://doi.org/10.1242/dev.074674

Sun R, Ye R, Gao L, Zhang L, Wang R, Mao T, Zheng Y, Li D, Lin Y (2017a) Characterization and ectopic expression of CoWRI1, an AP2/EREBP domain-containing transcription factor from coconut (*Cocos nucifera* L) endosperm, changes the seeds oil content in transgenic *Arabidopsis thaliana* and rice (*Oryza sativa* L). Front Plant Sci 8:63. https://doi.org/10.3389/fpls.2017.00063

Sun RH, Ye RJ, Gao LC, Zhang L, Wang R, Mao T, Zheng YS, Li DD, Lin YJ (2017b) Characterization and ectopic expression of CoWRI1, an AP2/EREBP domain-containing transcription factor from coconut (*Cocos nucifera* L.) endosperm, changes the seeds oil content in transgenic *Arabidopsis thaliana* and rice (*Oryza sativa* L.). Front Plant Sci 5:10. https://doi.org/10.3389/fpls.2017.00063

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599. https://doi.org/10.1093/molbev/msm092

Tranbarger TJ, Dussert S, Joet T, Argout X, Summo M, Champion A, Cros D, Omore A, Nouy B, Morcillo F (2011) Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. Plant Physiol 156(2):564–584. https://doi.org/10.1104/pp.111.175141

van der Linden CG, Vosman B, Smulders MJ (2002) Cloning and characterization of four apple MADS box genes isolated from vegetative tissue. J Exp Bot 53(371):1025–1036. https://doi.org/10.1093/jexbot/53.371.1025

Weigel D (1995) The APETALA2 domain is related to a novel type of DNA binding domain. Plant Cell 7(4):388–389. https://doi.org/10.1105/tpc.7.4.388

Yamaguchi T, Hirano HY (2006) Function and diversification of MADS-box genes in rice. ScientificWorldJournal 6:1923–1932. https://doi.org/10.1100/tsw.2006.320

Yamaguchi T, Lee DY, Miyao A, Hirochika H, An G, Hirano HY (2006) Functional diversification of the two C-class MADS box genes OSMADS3 and OSMADS58 in *Oryza sativa*. Plant Cell 18(1):15–28. https://doi.org/10.1105/tpc.105.037200

Yuan YJ, Liang YX, Li BZ, Zheng YS, Luo XQ, Li DD (2015) Cloning and function characterization of a beta-Ketoacyl-acyl-ACP synthase I from coconut (*Cocos nucifera* L.) endosperm. Plant Mol Biol Rep 33(4):1131–1140. https://doi.org/10.1007/s11105-014-0816-z

Yuan YJ, Gao LC, Sun RH, Yu T, Liang YX, Li DD, Zheng YS (2017) Seed-specific expression of an acyl-acyl carrier protein thioesterase CnFatB3 from coconut (*Cocos nucifera* L.) increases the accumulation of medium-chain fatty acids in transgenic *Arabidopsis* seeds. Sci Hortic-Amsterdam 223:5–9. https://doi.org/10.1016/j.scienta.2017.05.029

Zhang J, Nallamilli BR, Mujahid H, Peng Z (2010) OsMADS6 plays an essential role in endosperm nutrient accumulation and is subject to epigenetic regulation in rice (*Oryza sativa*). Plant J 64(4):604–617. https://doi.org/10.1111/j.1365-313X.2010.04354.x

Zhang X, Hirsch CN, Sekhon RS, de Leon N, Kaeppler SM (2016) Evidence for maternal control of seed size in maize from phenotypic and transcriptional analysis. J Exp Bot 67(6):1907–1917. https://doi.org/10.1093/jxb/erw006

Zheng Y, Jin Y, Yuan Y, Feng D, Chen L, Li D, Zhou P (2019) Identification and function analysis of a type 2 diacylglycerol acyltransferase (DGAT2) from the endosperm of coconut (*Cocos nucifera* L.). Gene 702:75–82. https://doi.org/10.1016/j.gene.2019.03.060