ARTICLE

Biochemical and molecular characterization of two 11S globulin isoforms from coconut and their expression analysis during seed development

Kristian Mark P. Caldo[1,2,3], Roberta N. Garcia*[1], and Evelyn Mae Tecson-Mendoza[1]

[1]Biochemistry Laboratory, Institute of Plant Breeding, Crop Science Cluster, College of Agriculture University of the Philippines Los Baños, College, Laguna 4031 Philippines

[2]Institute of Chemistry, College of Arts and Sciences, University of the Philippines Los Baños, College, Laguna 4031 Philippines

[3]Alberta Innovates Phytola Centre, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, T6G 2P5, Canada

Abstract—Cocosin, the 11S globulin of coconut, is the major storage protein that accumulates in the endosperm during seed development. This work describes the gene structure, derived amino acid sequence, structural homologies, and developmental expression profile of cocosin. Two full-length cocosin cDNAs spanning 1,641 and 1,623 nucleotides and encoding for 466 amino acids were isolated through PCR cloning strategy coupled with 5′ and 3′-RACE technologies. Both sequences displayed 92.3% nucleotide identity and 91.5% amino acid identity. They both exhibit the following conserved regions among 11S globulins/glutelins of various seed plants—signal peptide targeting storage vacuole deposition, highly conserved asparaginyl splice site dividing the polypeptides into acidic (32 kDa) and basic (21 kDa) subunits, bicupin domain and four cysteine residues involved in intra- and inter-chain disulfide bonding. Phylogenetic analysis showed that cocosin is more closely related to *Elaeis guineensis* glutelin, both of which form a distinct clade in the tree between the divergent clades of dicot and cereal 11S globulins. The predicted three-dimensional structure of the cocosin exhibits the bicupin domain separated by a less conserved loop. Relative PCR and western blot analysis showed that synthesis of cocosin started at 6–7 months after pollination (MAP) and increased continuously up to 8–9 MAP. Western blot analysis further showed that majority of cocosin was deposited at 11–12 MAP.

Keywords—cocosin, 11S globulins, *Cocos nucifera* L., storage protein, seed maturation, developmental expression

INTRODUCTION

Over the past two decades, extensive research efforts have been directed towards the isolation and identification of the members of the 11S globulin gene family. The 11S globulins are saline-soluble storage proteins, which constitute the major seed storage proteins in many plants (Shewry 1995). These plants include several legumes, oats, rice, and some oilseeds, which serve as source of dietary proteins for man and livestock. Hence, the study of these proteins is important to characterize their nutritional value, functional properties and potential allergenicity.

The 11S globulins are generally hexamers of around 320–450 kDa assembled through non-covalent random association of monomers which are coded by multiple genes (Casey et al. 1986) Each 11S globulin monomer is initially synthesized as a polypeptide of around 500 residues, which is post-translationally cleaved by an endoplasmic reticulum signal peptidase and subsequently by an asparaginyl endopeptidase between a highly conserved asparaginylglycinyl site. The truncated N-terminal portion forms the larger acidic subunit (30–40 kDa) which is linked prior to cleavage by a disulfide bond with the basic subunit (20–25 kDa) (Staswick et al. 1984; Jung et al. 1998; Shewry and Halford 2001). The aforementioned structure and mode of biosynthesis are also exhibited by the dilute acid/alkali-soluble glutelins, which also belong to the 11S globulin family despite having different solubility behavior (Shewry and Halford 2001).

Coconut is an important oil seed crop widely cultivated in tropical regions. The key products of the coconut industry comprise oil, copra meal, and desiccated coconut, all of which are derived from the coconut endosperm. The coconut endosperm at different stages of development consists of 27–44% oil and about 8% (dry weight) protein (Mendoza et al. 1982). The protein is mostly retained in the

*Corresponding Author
Email Address: rngarciaipb@yahoo.com
Submitted: October 15, 2014
Revised: March 24, 2015
Accepted: April 7, 2015
Published: May 22, 2015

residue after oil extraction. Thus, coconut proteins constitute a bulk by-product of the coconut oil industry. The coconut 11S globulin also known as cocosin is the major protein comprising more than 50% of total seed proteins. We have undertaken biochemical and molecular studies on cocosin to elucidate the nutritional and functional properties of this protein that may lead to its increased utilization. Carr et al. (1990) earlier reported the purification of cocosin with molecular weight of 300–360 kDa. We further purified and conducted physicochemical and functional characterization of cocosin (Garcia et al. 2005; Angelia et al. 2010). Cocosin was found to account for 86% of total globulins and be successfully extracted with 0.35 M NaCl. Anion-exchange chromatography of cocosin separated at least three and up to eight isoforms (Carr et al. 1990; Garcia et al. 2005). The native hexameric form has a molecular weight of 326 kDa and is composed of 6 subunits of acidic and basic chains linked by disulfide bond (Garcia et al. 2005). Cocosin was resolved on SDS-PAGE into two sets of bands: 32–35 kDa acidic subunits and 21–24 kDa basic subunits (Garcia et al. 2005). The basic subunit contains a carbohydrate moiety and is more resilient to chymotrypsin digestion. The hexameric form of cocosin is relatively heat stable with a thermal denaturation midpoint temperature of 100.5°C (Angelia et al. 2010). The N-terminal sequences of the selected acidic and basic bands are SVRSVNEFRXE and GLEETQ, respectively (Garcia et al. 2005).

The physicochemical and functional properties of cocosin have been adequately established at the protein level but its genes remain unreported and uncharacterized. In fact, limited efforts have been undertaken to identify important genes in coconut as compared to other agronomically important crops. Using PCR cloning strategy coupled with 5′- and 3′-RACE technologies, two cDNAs that code for distinct cocosin were identified. Analyses of the gene structure, derived amino acid sequence and structural homologies are reported as well as the developmental expression profile of cocosin at the transcript and protein levels. Emphasis was given on the comparison of cocosin with other monocot 11S globulins which have not been well-studied as a group unlike the other 11S globulins in the dicotyledonous family.

MATERIALS AND METHODS

Plant Material

The coconut endosperms were obtained from nuts of the Laguna Tall variety in the experimental field of the Institute of Plant Breeding, Crop Science Cluster, College of Agriculture, University of the Philippines Los Baños. Coconuts at different stages of development were used for the corresponding parts of the study: 8–9 months after pollination (MAP) for RNA extraction and cloning experiment; 11–12 MAP for protein extraction of cocosin for use in antibody production; 5–6, 6–7, 7–8, 8–9, 10–11, and 11–12 MAP for the developmental expression studies.

Total RNA Extraction and Reverse Transcription

Fresh endosperms from nuts at 5–9 MAP were ground in liquid nitrogen to a powder using mortar and pestle. Total RNA was extracted using TRIzol® reagent according to the manufacturer's instruction (Invitrogen, USA). The quality and quantity of the RNA extracts were evaluated through 1% agarose gel electrophoresis and spectrophometrically (A_{260} and A_{280}). Total RNA extract (3 µg) was subjected to first strand synthesis using Superscript® III Reverse Transcriptase with oligo(dT) primers according to the manufacturer's protocol (Invitrogen, USA).

TABLE 1. Primers used in the isolation of the cocosin cDNA isoforms through the PCR Cloning Strategy coupled with 5 and 3-RACE.

	Nucleotide sequence	Conserved oligopeptide	Orientation
Csnfor-1	CCTGGGTGTCCGGAGACCTTT	PGCPETF	FORWARD
Csnfor-2	CGGTGTGCGGGGGTTTCT	RCAGVS	FORWARD
Csnfor-3	CGGCGGGTTATTGAGCCG	RRVIEP	FORWARD
Csnrev-1	NGTYTCYTCYAANCCRTT	NGLEET	REVERSE
Csnrev-2	RTGNGCRTTDATRTTCCA	WNINAH	REVERSE
CSN1-5RACE GSPA	CAGGATTCTCCCCATCGTTGT	Actual cDNA	REVERSE
CSN2-5RACE GSPA	CCAAAAGCCGCTGCCAGCAACTC	Actual cDNA	REVERSE
CSN1-3RACE GSPA	CGCCGTGAGGAGAGCATCAAGGA	Actual cDNA	FORWARD
CSN1-3RACE GSPA	GGTATCGCCCCACTGGAATATCAAT	Actual cDNA	FORWARD

PCR Amplification and 5′- and 3′-RACE

The primers used in this part of the study are summarized in Table 1. Three gene-specific forward primers namely Csnfor-1, -2, and -3 and the reverse primers Csnrev-1, -2, and -3 were designed and synthesized (Life Technologies, Hong Kong) based on conserved regions among 11S globulins. PCR amplification was carried out using the different combinations of the forward and reverse primers. A standard 20-µL reaction mixture was used for each primer pair combination before the components were varied one at a time. The PCR profile consisted of an initial denaturation at 94 °C for 3 min, 35 amplification cycles of 1 min at 94 °C, 1 min at 45–60 °C and 1 min per kb expected at 72 °C, and a final extension step at 72 °C for 5 min. The PCR reaction that yielded the cocosin amplicon consisted of the following components: 1x PCR buffer, 2.0 mM $MgCl_2$, 0.20 mM dNTP, 0.75 µM Csnfor-1, 0.75 µM Csnrev-2 and 1 U Taq polymerase (Vivantis) The PCR products were resolved on a 1.0% agarose gel stained with 0.10% ethidium bromide. 5′-RACE was carried out following the manufacturer's protocol (Invitrogen, USA). Two reverse primers, namely, CSN1-5RACE GSPA and CSN2-5RACE GSPA were designed based on the partial sequence obtained and used for the reverse transcription and PCR amplification of the 5′-portion respectively. The 3′-RACE reaction was carried out according to the manufacturer's protocol (Seegene Capfishing Kit, South Korea). For this, two forward primers namely CSN1-3RACE GSPA and CSN1-3RACE GSPA were also designed based on the partial sequence obtained.

Cloning and DNA Sequencing

The PCR products were cloned using the pGEM®-Teasy cloning system according to the manufacturer's instructions (Promega, USA). The transformed cells were plated on Amp-LB-IPTG-x-Gal plates (1.0% tryptone, 0.5% yeast extract, 1.0% NaCl, and 1.5% agar at pH 7.0) followed by incubation at 37°C for 20 hours. The transformants were screened through blue-white screening and colony PCR. The positive transformants were grown overnight in 10 mL Luria-Bertani Broth (1.0% tryptone, 0.5% yeast extract, and 1.0% NaCl, pH 7.0) containing 100 µg/mL ampicillin. The plasmids were isolated using PureLink™ Quick Plasmid Miniprep Kit according to the manufacturer's instructions (Invitrogen, USA). The recombinant plasmids were sequenced under the BigDye™ terminator cycling conditions. The products were purified through ethanol precipitation and were sequenced using Automatic Sequencer 3730 XL (Macrogen, South Korea).

Characterization of the Full-length Cocosin cDNA

The cDNA sequences obtained were analyzed in silico. The deduced amino acid sequences were derived by translating the nucleotide sequences at different reading frames using the Translate tool (http://expasy.org). The composition of the nucleotide and amino acid sequences were analyzed using ApE and ProtParam (http://expasy.org). The resulting sequences were aligned with known 11S–12S globulin genes from the database using BLASTp and Clustal W to establish their identities (Altschul et al.1997; Thompson et al. 1994). The presence of the highly conserved regions in the nucleotide and amino acid levels as well as the percent similarity and identity to known 11S globulins were noted. Evolutionary analysis of cocosin together with highly homologous sequences was performed using Neighbor-Joining method (Saitou and Nei 1987). A molecular model of CnCos-1 was constructed through homology modelling using Phyre server (http://www.sbg.bio.ic.ac.uk/phyre/).

Developmental Expression Analysis: Relative RT-PCR

Primers targeting the cocosin cDNA were designed using Primer-BLAST (http://ncbi.nlm.nih.gov/tools/) and Primer 3 (http://frodo.wi.mit.edu/primer3/). The actin gene, a housekeeping gene, was used as the internal control. Relative PCR was optimized by varying the annealing temperature through gradient PCR (T_m±5ºC) using 8–9 MAP cDNA.

The first strand cDNAs obtained from coconut endosperm at different stages of developments were used as templates for relative PCR. The components of relative PCR are as follows: The reaction mixture was initially incubated at 94 °C for 3 min, followed by 35 amplification cycles of 40 s at 94 °C, 40 s at 60 °C and 40 s at 72 °C, and a final extension at 72 °C for 5 min. The intensities of the amplicons corresponding to the cocosin and actin genes were semi-quantified using QuantityOne software (Biorad, USA). The relative expression was obtained by getting the cocosin : actin amplicon ratio at different stages of development.

Antibody Production Against Cocosin and Characterization

Cocosin was extracted and purified according to the protocol described by Garcia et al. (2005). Polyclonal antibodies to cocosin were raised in white rabbits. The immunoreaction of the polyclonal antibodies against purified cocosin was determined using double immunodiffusion assay and enzyme-linked immunosorbent assay.

SDS-PAGE and Protein Content Determination

Electrophoresis was done according to the method described by Laemmli (1970). SDS-PAGE was performed on 11% discontinuous denaturing gels using a minigel electrophoresis apparatus (BIORAD). Protein samples were run at 110 V for 1 h and 30 min then stained with 0.25% Coomassie blue R250 for 20–30 min. The gels were destained with 40% methanol-7% acetic acid in water solution. The molecular weights of the subunits were estimated using BenchMark™ Protein Ladder (Invitrogen, USA). The protein content of the samples was determined according to the dye-binding method of Bradford (1986) using bovine serum albumin (BSA) as protein standard.

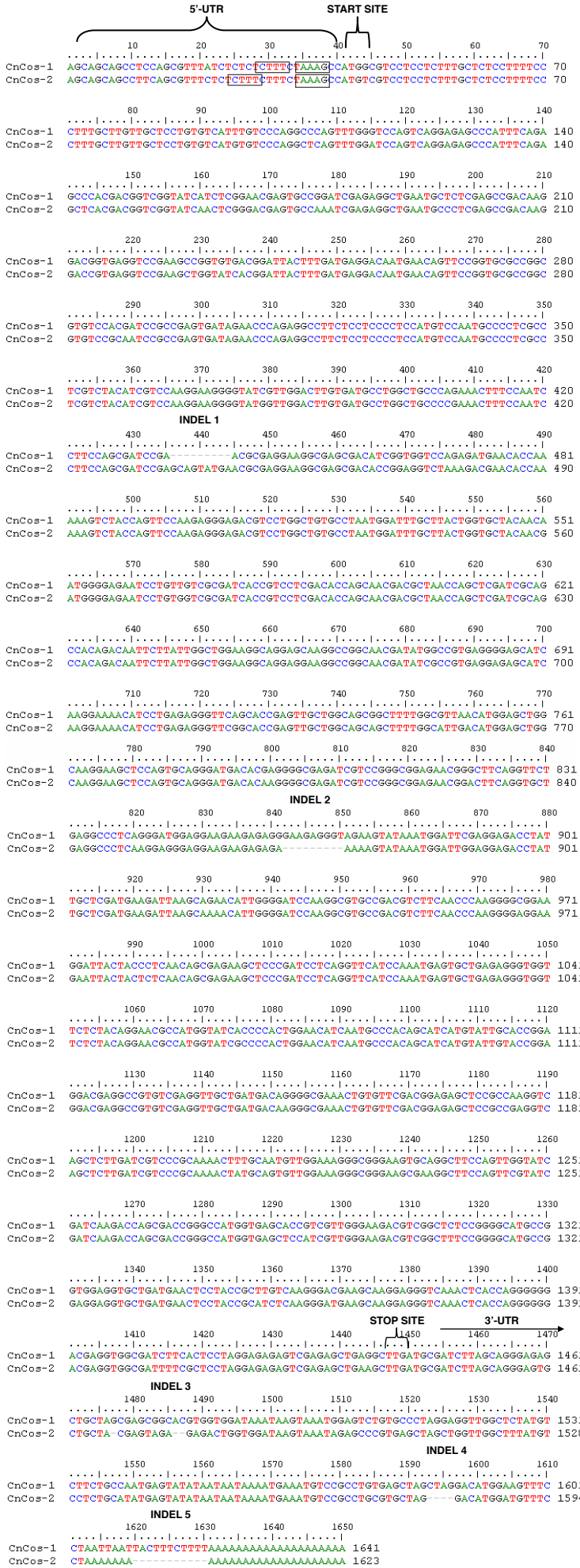

Figure 1. Pairwise nucleotide sequence alignment of CnCos-1 (KP902412) and CnCos-2 (KP902413). The start and stop sites, 5′ and 3′-UTRs, and the major gaps (INDELS) are labeled. The DOFCOREZM cis-elements are boxed.

Western Blotting

Coconut endosperms at different developmental stages (6–7, 7–8, 8–9, 10–11 and 11–12 MAP) were defatted using the protocol described by Garcia et al (2005). Total proteins were extracted from 1.0 g ground defatted coconut endosperm by the addition of 18 mL extraction buffer (18 mL 0.4 M NaCl in 35 mM potassium phosphate buffer, pH 7.6 with 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM β-mercaptoethanol and 0.02% sodium azide). The mixture was stirred for 1 hour on an ice bath and the resulting homogenate was passed through four layers of cheese cloth, and was centrifuged at 23,500 x g for 15 min using a Beckman-Coulter T-28 Rotor Refrigerated Centrifuge. The supernatant containing the total proteins was collected and subjected to further analysis. Total proteins (8 µg) were subjected to 11% denaturing SDS-PAGE at 90V. The resolved total proteins were electroblotted onto polyvinyldifluoride (PVDF) membrane. Electroblotting was carried out at 45 mA for 3 hours. The PVDF blot was recovered and screened with the polyclonal antibodies coupled with an anti-rabbit IgG-alkaline phosphatase-BCIP/NBT detection system. The air-dried blot was scanned and the resulting images were analyzed using ScionImage Software (Scion Corporation, Maryland, U.S.A) to determine the pattern of cocosin protein synthesis in coconut at different developmental stages.

RESULTS AND DISCUSSION

Isolation and Characterization of Two Cocosin cDNAs

Homology-based PCR cloning strategy was employed. Conserved consensuses among 11S globulins of *Elaeis guineensis* (Q9SNZ2), *Avena sativa* (P14812), *Oryza sativa* (Q09151) and *Zizania latifolia* (Q0Z945) were used to design and synthesize primers. No PCR product was obtained out of reactions having Csnfor-2, Csnfor-3 and Csnrev-1 as one of the primers. A distinct PCR product of expected length (700 bp) was obtained using the primer pair Csnfor-1 and Csnrev-2 which was subsequently subjected to cloning and DNA sequencing. DNA sequence analysis identified two types of cocosin cDNAs, designated as CnCos-1 (KP902412) and CnCos-2 (KP902413), from eleven and ten clones, respectively, of the 21 clones sequenced. The 5′- and 3′-ends were cloned through 5′- and 3′-RACE PCR, respectively. CnCos-1 and CnCos-2 cDNA sequences span 1,641 and 1,623 nucleotides, respectively, and share 92.3% identical nucleotides (Figure 1). CnCos-1 and CnCos-2 both have 5′-untranslated regions of 40-bp, open reading frames of 1401 bp, and 182–200 bp 3′-untranslated regions. PLACE analysis identified the presence of DOFCOREZM consensus along the 5′-untranslated region of both cDNAs (Higo et al. 1999). This consensus was reported as the core binding site of the maize transcription factors, one of which is found to mediate endosperm-specific expression by binding to a prolamin box (Yanagisa and Schmidt 1999). Both cDNA sequences are GC-rich and have high preference for adenine at the wobble positions. Two major insertion/deletions that are multiples of three are present within the open reading frames of the aligned cocosin sequences, which mostly account for most of the difference between the two.

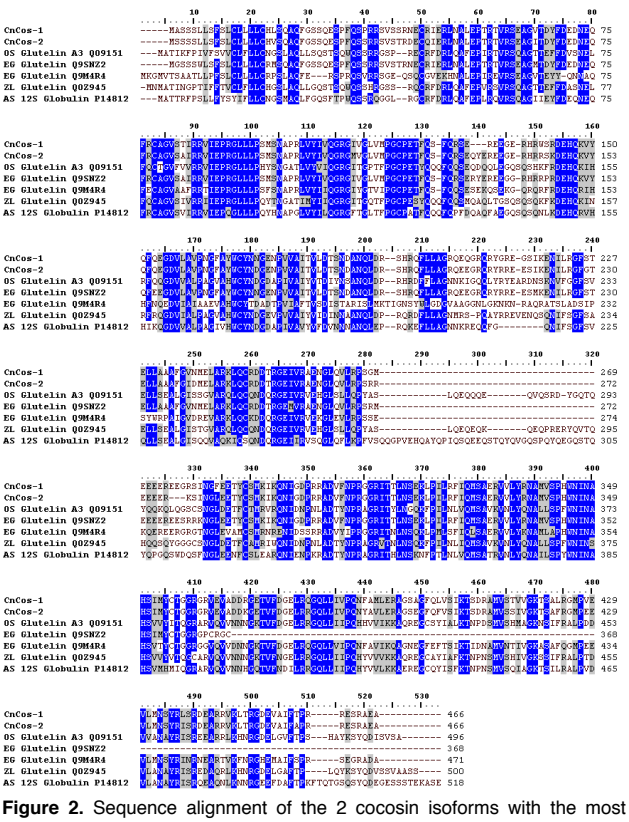

Figure 2. Sequence alignment of the 2 cocosin isoforms with the most homologous monocot 11S globulins (CnCos-coconut cocosin; OS- Oryza sativa, EG- *Elaeis guineensis*, ZL- *Zizania latifolia* and AS-*Avena sativa*).

Analysis of the Derived Amino Acid Sequences

The derived amino acid sequences of CnCos-1 and CnCos-2 both span 466 amino acid residues and have molecular weights of 52,597 and 52,956 Da, respectively (Figure 2). The functional properties of the primary structure were characterized. A hydrophobic signal peptide comprising the first 30 residues most likely directs the post-translational targeting of the cocosin into storage vacuoles as predicted by iPSORT (Bannai et al. 2002). This prediction is consistent with previous studies which showed that 11S globulins are targeted into plant storage vacuoles via the Golgi complex (Sanderfoot and Raikhel 1999). The N-terminal sequence, SVRSVNEFRXE, previously reported closely aligned with the 38th–48th residues of the two isoforms, with 7 identical residues out of 11 in CnCos-1 and 4 out of 11 in CnCos-2 (Garcia et al. 2005). The presence of multiple isoforms may have interfered with the precision of N-terminal sequencing result. Considering this N-terminal sequence, the presence of the propeptide, spanning 7 residues, can be noted as the signal peptides in both isoforms do not directly precede the N-terminal residue.

Sequence alignment of the two cocosin isoforms with reported 11S globulins revealed the presence of 4 conserved cysteine residues (Figure 2). The two cysteines (C_{45} and C_{77}) near the N-terminal portion form an intrachain disulfide linkage within the acidic subunit. On the other hand, the other two (C_{121} and C_{228}) are responsible for the formation of the interchain disulfide bond between the mature acidic and basic subunits, as exhibited by the well-studied legumins (Staswick et al. 1984; Horstmann 1983). The highly conserved asparaginyl splice site is identified in the alignment as N_{281}-G_{282} of both isoforms (Dickinson et al. 1989; Jung et al. 1998). It divides the precursor polypeptides into the expected acidic and basic subunits with 281 and 185 residues, respectively, corresponding to pI values of 5.6–5.8 and 9.6–9.8, respectively. The predicted sizes for the acidic and basic subunits are 32 kDa and 21 kDa, respectively, which are in good agreement with the values earlier obtained (Garcia et al. 2005; Angelia et al. 2010).

The CnCos-1 and CnCos-2 amino acid sequences share 91.5% identical amino acid residues. BLAST analysis revealed that both sequence exhibits close similarity with glutelins and 11S globulins of monocots, dicots and a magnoliid. The two isoforms exhibit highest homology (58–93% and 56%, respectively) with *Elaeis guineensis* glutelins (Q9SNZ2 and Q9M4R4) and 11S globulins of *Magnolia salifocia* (Q40346 and Q40347). In addition, the two isoforms manifest 43–51% identity with glutelins, 11S globulins and 12S globulins of other monocots including *Oryza sativa, Avena sativa, Musa acuminata* and *Zizania latifolia* and considerable identity (<50%) with dicot 11S globulins.

The 11S globulin sequences of reported monocot orthologues of cocosin exhibit distinct regions of variability and similarity (Figure 2). The alignment reveals that the N-terminal portion of the basic subunits is the most highly conserved region in this subunit while the acidic subunits consist of alternating conserved and variable regions. On the other hand, the presence of five major regions of variability is also noted in the alignment. Of these regions, a hypervariable segment is particularly noted along the C-terminal portion of the acidic subunit. This hypervariable segment has also been identified among 11S globulins of dicots (Jain 2004; Argos et al. 1985). Cocosin together with glutelin of African oil palm contains runs of glutamic acid residues within this portion (amino acid residues number 321–327) which is not observed in other non-Araceae monocots (Figure 2). On the other hand, the 11S globulins/glutelins of the other monocots belonging to the cereal family contain extended oligopeptide not present in the two palm species. It can also be noted that these hypervariable regions of cereal 11S globulins are glutamine-rich having an average of 33.5% glutamine, which is observed in other dicot 11S globulin. The second and third regions of variability span 139 to 152 and 215 to 231, respectively. These regions were also noted among 11S globulins of legumes (Jain 2004; Argos et al. 1985). Both regions in palm 11S globulins/glutelins are rich in glutamic acid and arginine, which is not observed in other monocots.

Evolutionary analysis of CnCos-1, CnCos-2 and their orthologues (Figure 3) revealed that these proteins form two subfamilies, one composed of monocots and the other composed of dicots. The monocot subfamily forms two subclades. The two cocosin isoforms form one subclade together with similar proteins from *Elaeis guineensis* and *Musa acuminata*. Within this subclade, it can also be noted that the two cocosin isoforms together with one of the *E. guineensis* glutelin (Q9SNZ2) diverged from other reported Eg glutelin. It is possible that there are other cocosin isoforms belonging to this divergent clade that have yet to be isolated. The other monocot subclade on the other hand is composed of cereal 11S/12S globulin/glutelins.

The deduced amino acid sequence was mapped for the presence of conserved domains. A bicupin domain, a conserved barrel found in 11S and 7S plant seed storage proteins, is detected along the primary structure of both isoforms. The cupin domain is comprised of β-strands separated by less conserved loop. Figure 4 shows the predicted three-dimensional structure of CnCos-1 with the bicupin domain illustrated in red and blue. CnCos-1 and CnCos-2 manifest high levels of similarity with cupin domains at the N- and C-terminal portions contributed by the acidic and basic subunits. The N-terminal motifs of CnCos-1 and CnCos-2 consist of 127 and 129 amino acid residues, respectively. On the other hand, the C-terminal motifs consist of 135 and 134 residues, respectively. CnCos-1 and CnCos-2 are both classified under the 00190 Pfam family of 11S plant seed storage proteins when compared with the conserved domain database (CDD). The presence of the bicupin domain may be able to account for the high thermal stability of cocosin. Our previous study showed that the hexameric form of cocosin has a thermal denaturation midpoint temperature of 100.5°C which was higher than those of other 11S globulins (Angelia et al. 2010).

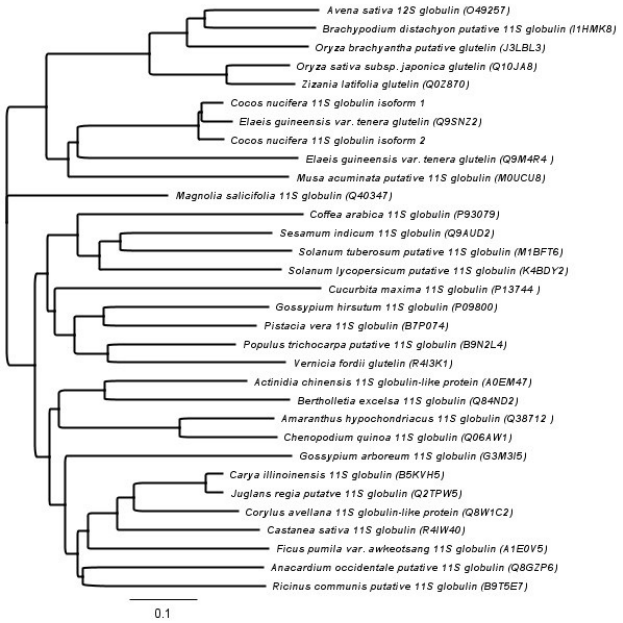

Figure 3. Phylogenetic tree analysis of cocosin and selected monocot and dicot 11S globulins

Developmental Expression Analysis of Cocosin

Cocosin transcripts were analyzed through relative PCR in the coconut endosperm at four developmental stages: 5–6, 6–7, 7–8 and 8–9 MAP (Figure 5). The constitutively expressed actin gene was also amplified to serve as internal control. Actin had earlier been shown to be an appropriate internal control for coconut studies (Dela Cruz et al. 2011).

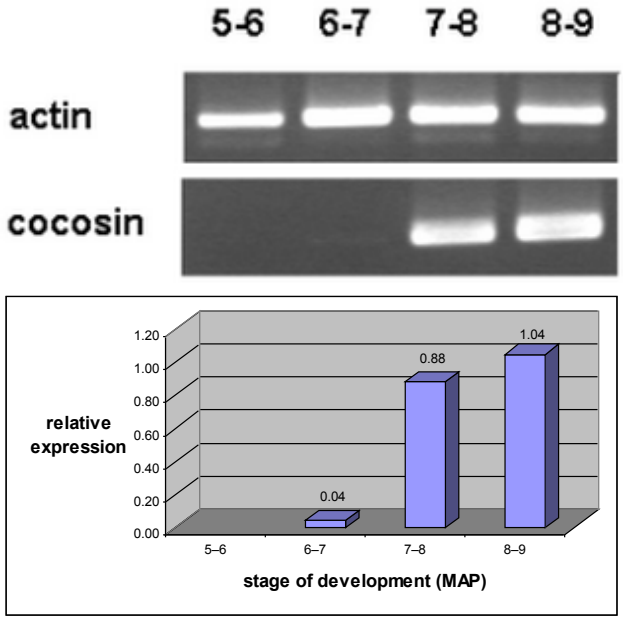

Figure 5. Relative PCR of cocosin with actin as internal control using cDNA templates derived from endosperms at different stages of development (MAP).

Results of relative PCR confirmed the constitutive expression of actin across the developmental stages considered (Figure 5). On the other hand, it was shown that cocosin transcripts were not yet detectable by PCR at 5–6 MAP. The initiation of cocosin transcription started at past mid-maturation stage, as indicated by the very faint cocosin amplicon at 6–7 MAP (Figure 5). The level of cocosin transcripts increased dramatically at 7–8 MAP. The high level of cocosin transcripts was maintained in the next stage of development considered (8–9 MAP). The first and second phases of seed development are marked with rapid cell division and differentiation, respectively (Weber et al. 1998). These phases are highly exhibited by the coconut endosperm at 5–6 MAP, during which the coconut endosperm is starting to undergo a transition from a liquid to a gelatinous state. This transition continues and is almost completed in the next stage of development. At 6–7 MAP, the nascent solid endosperm has fully covered the inner endocarp. This phenomenon is demonstrated by oil palms once the embryos have completed their cell multiplication period (Ferdinando et al. 1985). The initiation of cocosin transcription starts at this stage. The initiation of seed storage synthesis has been previously shown to coincide with the end of cell division, or the start of the cell

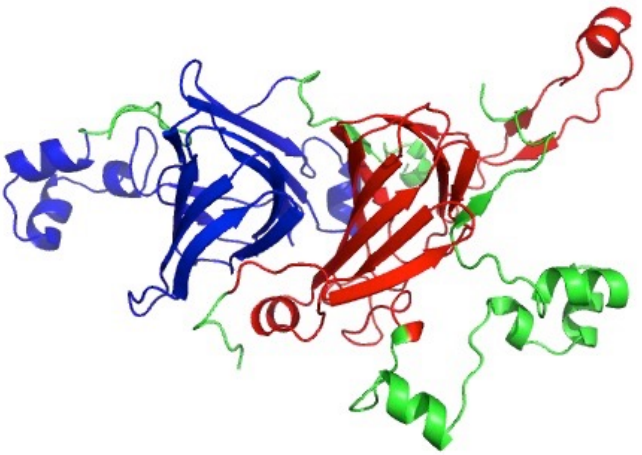

Figure 4. Homology model of CnCos-1 showing the predicted bicupin domain in red and blue. The red and blue regions correspond to the domains contributed by the acidic and basic subunits, respectively. The model was constructed by the Phyre server using recombinant *Cucurbita maxima* pro-11S globulin (PDB ID 2E9Q) as template with an E-value of 1e-40.

expansion or maturation in developing embryos (Dure 1985; Millerd and Spencer 1974). In addition, seed storage protein synthesis has been regarded as a marker for maturation (Weber 1998). At 6–7 MAP, it can be inferred that the coconut fruit is entering the third phase of seed development which is marked by the assimilation of storage protein reserves and cell expansion. This is further validated by results showing that the coconut endosperm continues to accumulate 11S globulin transcripts at 7–8 MAP and 8–9 MAP (Figure 6), with the concomitant expansion of the coconut endosperm. Transcription level of cocosin in more mature endosperm (>10 MAP) could not be ascertained as attempts to obtain high quality RT-PCR grade RNA from samples of that age were unsuccessful.

We also determined the developmental expression at the protein level of cocosin using antibodies prepared against pure cocosin. At an early stage of fruit maturity (6–7 MAP), minute amount of 11S globulins (~4.5% of total cocosin at the end of seed development) was already detected in the Western blot (Figure 6). This confirmed the relative PCR results demonstrating that cocosin transcription was initiated at this stage. The level of cocosin manifested a 4-fold increase to 17.6% at 7–8 MAP, which correlated with the sudden accumulation of cocosin transcripts. The amount of cocosin detected in the blot as well as the level of cocosin transcripts gradually increased from 7–8 to 8–9 MAP. Furthermore, it was shown in the immunoblot that the gradual increase of cocosin transcripts continued until 10–11 MAP, reaching 45% at this stage. The blot further showed that majority of storage protein deposition (~55% of total cocosin) took place at the end of fruit development (between 10–11 and 11–12 MAP), right before post-germination where these proteins will be needed to nourish the seedling.

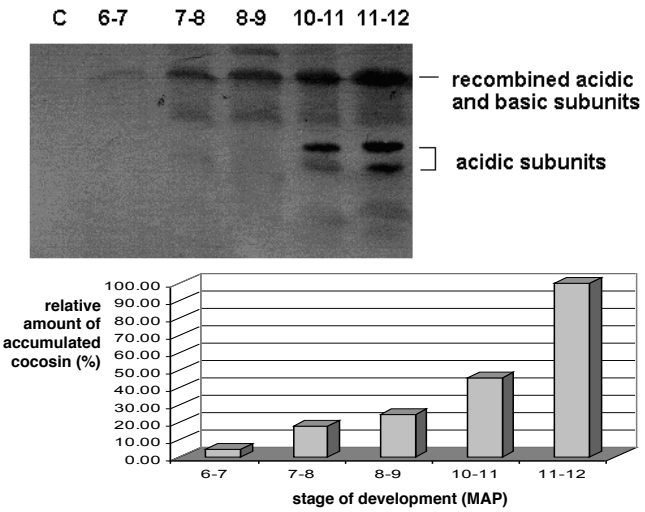

Figure 6. Cocosin levels in coconut endosperm at different developmental stages (MAP) determined by western blot. (a) Western blot of cocosin; (b) relative expression of accumulated cocosin.

In this study, we report two isoforms of the coconut 11S globulin gene family. The presence of other isoforms cannot be disregarded given that 11S globulin gene families commonly consist of multiple members. The presence of at least three cocosin isoforms was noted when cocosin was resolved on an anion-exchange chromatography column (Garcia et al. 2005) while Carr et al. (1990) noted at least 8 cocosin peaks on anion-exchange chromatography, indicative of the presence of multiple isoforms.

The 11S globulins exhibit three characteristic features which are all present in the primary structure of the two cocosin isoforms: the asparaginylglycinyl cleavage site, four conserved cysteine residues involved in intra- and inter-chain disulfide bonding and the bicupin domain. The SDS-PAGE profile of the cocosin protein under reducing and non-reducing conditions showed the presence of the two sets of bands corresponding to the dissociated acidic and basic subunits which are linked via disulfide bond (Garcia et al. 2005). The first set is composed of two closely migrating bands of 35 and 32 kDa (M_r), corresponding to the larger acidic subunit. On the other hand, the second set of closely migrating bands of 24 and 21 kDa (M_r) correspond to the smaller basic subunit. These two sets of acidic and basic polypeptides may have originated from at least two precursor polypeptides cocosin isoforms upon the action of the asparaginyl endopeptidase on the aforementioned cleavage site. The expression profile of cocosin was also reported in this study at the transcript and protein levels. It is found to be comparable with the expression pattern of the major storage protein of *Elaeis guineensis* and coffee 11S globulin. At around 14 weeks after anthesis (WAA), the *Elaeis guineensis* endosperm begins to accumulate 7S globulins while its endosperm changes into a gelatinous state. 7S globulins continually accumulate in the next three weeks prior to drying stage at 17th WAA (Morcillo et al. 1998). In addition, *Elaeis guineensis* accumulates over two-thirds of its total 7S globulins at a later stage of seed development (between 14th and 17th WAA). Coffee 11S globulin transcription begins at past mid-maturation stage of 126 days after flowering (DAF) and is maintained at high level until 189 DAF and become undetectable at 245 DAF (Roger et al. 1999). Cocosin, like most 11S globulins, was found to be produced at 6-7 MAP and increased steadily up to mature stage of the endosperm at 11-12 MAP.

CONCLUSION

We have isolated and characterized two isoforms of the coconut 11S globulin gene family, which is a first report on the genes of the coconut storage proteins. CnCos-1 and CnCos2 together with *Elaies guineensis* glutelins constitute one of the two subfamilies of monocot 11S globulins. The full-length cocosin isoforms, CnCos-1 and CnCos-2, are both composed of 466 aa (91.5% identical). These isoforms are similar (58–93%) to 11S globulins/glutelins of monocots and dicots. The conserved residues/domains were identified as bicupin domain separated by a less conserved loop, thus, bridging the gap in knowledge on the biochemistry of coconut proteins. The developmental expression profile was elucidated at the transcript and protein levels. It was shown that cocosin is initially expressed at past mid-maturation stage of 6–7 MAP (cell expansion), gradually accumulated as the fruit matures up to 10-11 MAP (continuous cell expansion) and massively deposited at the end of fruit development. These data provided insights on the molecular events related to coconut seed development.

ACKNOWLEDGEMENT

The grant from the University of the Philippines System Research and Creative Grant to Roberta N. Garcia and the support of the Institute of Plant Breeding, Crop Science Cluster, College of Agriculture, University of the Philippines Los Baños are gratefully acknowledged. The first author (KMPC) sincerely thanks the Philippine Council for Advanced Science and Technology of the Department of Science and Technology for the scholarship grant and to the Institute of Chemistry, College of Arts and Sciences, UPLB for the two-year study leave provided him.

REGISTRATION OF COCOSIN DNA SEQUENCES

CnCos-1 and CnCos-2 are registered with the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). The accession numbers are KP902412 KP902413, respectively.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

CONTRIBUTION OF INDIVIDUAL AUTHORS

RNG, EMTM and KMPC conceptualized the study. All authors participated in the design of the experiments. KMPC performed the experiments. All authors analysed results and wrote the paper.

REFERENCES

Altschul SF, Maden TL, Schaffer AA, Zhang J, Zheng A, Miller, Lipman DJ. Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997; 25: 3389–3402.

Angelia MRN, Garcia RN, Caldo KMP, Tecson-Mendoza EM. Physicochemical and functional characterization of cocosin, the coconut 11S globulin. Food Sci. Tech. Res. 2010; 16: 225–232.

Argos P, Narayana SVL, Nielsen NC. Structural similarity between legumin and vicilin storage proteins from legumes. EMBO Journal 1985; 4:1111-1117.

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 2002; 18(2): 298–305.

Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. Analytical Biochem 1976; 72: 248-254.

Carr HJ, Plumb GW, Parker ML, Lambert N. Characterization and crystallization of an 11S seed storage globulin from coconut (*Cocos nucifera* L. Food Chem 1990; 38: 11–20.

Casey R, Domoney C, Smith AM. Legume storage proteins and their genes. In: Miflin BJ (ed). Oxford Surveys of Plant Molecular and Cell Biology. Oxford University Press: Oxford 3 (1986) 1–95.

Dela Cruz RY, Laude RP, Diaz MCQ, Laurena AC, Mendioro MS, Mendoza EMT. Gene for actin is a suitable internal reference for relative RT-PCR-based expression analysis in normal and mutant 'Makapuno' endosperms of coconut (*Cocos nucifera* L.). Philipp Agric Scientist 2011; 94: 118-123.

Dickinson CD, Hussein EHA, Nielsen NC. Role of post-translational cleavage in glycinin assembly. Plant Cell 1989; 1: 459–469.

Dure L. Embryogenesis and gene expression during seed formation, Oxford Surveys. Plant Mol Cell Biol 1985; 2: 179-197.

Ferdinando D, Hulme JH, Hughes WH. Oil palm embryogenesis: A biochemical, morphological study. In: Chapman GP, Mantell SH, Daniels RW, (eds). Experimental Manipulation of Ovule Tissue. France: Longman, 1985: 135–150.

Garcia RN, Arocena RV, Laurena AC, Tecson-Mendoza EM. 11S and 7S globulins of coconut (*Cocos nucifera* L.): Purification and characterization. J Agric Food Chem 2005; 53:1734–1739.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database. Nucl Acids Res 1999; 27(1): 297–300.

Horstmann C. Specific subunit pairs of legumin from *Vicia faba*. Phytochemistry 1983; 22:1861–1866.

Jain AK. Cloning and structural analysis of a cDNA clone encoding glycinin (Gly-1) seed storage protein of peanut. Electronic Journal of Biotechnology 2004; 7: (3). http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-34582004000300003

Jung R, Scott MP, Nam YW, Beaman TW, Bassuner R, Saalbach I, Muntz K, Nielsen NC. The role of proteolysis in the processing and assembly of 11S seed globulins. Plant Cell 1998; 10: 343–357.

Laemmli UK. Cleavage of structural proteins during the assembly of the bacteriophage T4. Nature 1970; 227:680-685.

Mendoza EMT, Mujer CV, Rodriguez FM, Ramirez DA. Ontogeny of some chemical constituents in makapuno and normal coconut endosperms. Kalikasan, Philipp J Biol 1982; 11: 99107.

Millerd A, Spencer D. Changes in RNA-synthesizing activity and template activity in nuclei from cotyledons of developing pea seeds. Aust J Plant Physiol 1974; 1: 331–341.

Morcillo F, Aberlene-Bertossi F, Hammon S, Duval SY. Accumulation of storage protein and 7S globulins during zygotic and somatic embryo development in *Elaeis guineensis*. Plant Physio Biochem 1998; 36(7): 509–514.

Roger WJ, Bezard G, Deshayes A, Meyer I, Petiard V, Maraccini P. Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea Arabica* endosperm. Plant Physiol Biochem 1999; 37: 261–272.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4:406-425.

Sanderfoot AA, Raikhel NV. The specificity of vesicle trafficking: Coat proteins and SNAREs. The Plant Cell 1999; 11: 629–641.

Shewry PR. Plant storage proteins. Biol Rev Camb Philos Soc 1995; 70: 375–426.

Shewry PR, Halford NJ. Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 2001; 370: 947–958.

Staswick PE, Hermodson MA, Nielsen NC. Identification of the cysteines which link the acidic and basic components of the glycinin subunits. J Biol Chem 1984; 259: 13431–13435.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673–4680

Weber H, Heim U, Golombek S, Borisjuk L, Wobus U. Assimilate uptake and the regulation of seed development. Seed Science Research 1998; 8: 331–345.

Yanagisawa S, Schmidt RJ. Diversity and similarity among recognition sequences of Dof transcription factors. Plant J 1999; 17: 209–214.