# ARTICLE

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

Kristian Mark P. Caldo<sup>1,2,3</sup>, Roberta N. Garcia<sup>\*1</sup>, and Evelyn Mae Tecson-Mendoza<sup>1</sup>

<sup>1</sup>Biochemistry Laboratory, Institute of Plant Breeding, Crop Science Cluster, College of Agriculture University of the Philippines Los Baños, College, Laguna 4031 Philippines

<sup>2</sup>Institute of Chemistry, College of Arts and Sciences, University of the Philippines Los Baños, College, Laguna 4031 Philippines

<sup>3</sup>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

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

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 Nterminal 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	CGGTGTGCGGGGGGTTTCT	RCAGVS	FORWARD
Csnfor-3	CGGCGGGTTATTGAGCCG	RRVIEP	FORWARD
Csnrev-1	NGTYTCYTCYAANCCRTT	NGLEET	REVERSE
Csnrev-2	RTGNGCRTTDATRTTCCA	WNINAH	REVERSE
CSN1-5RA CE GSPA	CAGGATTCTCCCCATCGTTGT	Actual cDNA	REVERSE
CSN2-5RA CE GSPA	CCAAAAGCCGCTGCCAGCAACTC	Actual cDNA	REVERSE
CSN1-3RA CE GSPA	CGCCGTGAGGAGAGCATCAAGGA	Actual cDNA	FORWARD
CSN1-3RA CE 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<sub>2</sub>, 0.20 mM dNTP, 0.75  $\mu M$  Csnfor-1, 0.75  $\mu M$  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<sup>®</sup>-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<sup>TM</sup> Quick Plasmid Miniprep Kit according to the manufacturer's instructions (Invitrogen, USA). The recombinant plasmids were sequenced under the BigDye<sup>TM</sup> 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 ( $T_{m\pm}5^{\circ}C$ ) 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<sup>™</sup> 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.

			$\sim$						
		10	20	20	<u> </u>	50	60	20	Fia
		. [	1	1	401 1			11	g.
CnCos-1	AGCAGCAG	CCTCCA	GCGTTTATCT	CTCTCTTTCTA	AAGCCATGGC	TCCTCCTCI	TTGCTCTCC	TTTTCC 70	0-0
CnCos-2	AGCAGCAG	CCTTCA	GCGTTTCTC	CTTTCTA	AAG <mark>CCATGTC</mark>	FTCCTCCTCI	TTGCTCTCC	TTTTCC 70	Cnu
									Cn
		80	90	100	110	120	130	140	star
			1	1	.			11	LITE
CnCos-1	CTTTGCTI	GTTGCT	CCTGTGTCAT	TTGTCCCAGGC	CCAGTTTGGG	CCAGTCAGG	AGAGCCCAT	TTCAGA 140	/181
CnCos-2	CTTTGCTI	GTTGCT	CCTGTGTCAT	GTGTCCCAGGC	TCAGTTTGGAT	CCAGTCAG	AGAGCCCAT	TTCAGA 140	
									DO
		150	160	170	180	190	200	210	are
			1	1	.			1	u. 0
CnCos-1	GCCCACGA	CGGTCG	GTATCATCTO	GGAACGAGTGC	CGGATCGAGAG	GCTGAATGC	TCTCGAGCC	GACAAG 210	
CnCos-2	GCTCACGA	CGGTCG	GTATCAACTO	GGGACGAGTGC	CAAATCGAGAG	GCTGAATGC	CCTCGAGCC	GACAAG ZIU	
		220	230	240	250	260	270	280	
CnCog-1	CACCCTCA		A CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		. mmc.mc.cc.	···· ····	mmercemer	 CCCCCC 200	
CnCos-2	GACCGTGA	GGTCCC	AAGCTGGTAT	CACGGATTACT	TTGATGAGGA	AATGAACAG	TTCCCGTCC	GCCGGC 280	
011000 1	011000101				110/110/10/07	0.010		000000 200	
							0.40		
		290	300	310	320 	330	340	350	
CnCos-1	GTGTCCAC	GATCCG	CCGAGTGAT?	GAACCCAGAGG	COTTOTCOTO	CCTCCATGT	CCAATGCCC	crece 350	
CnCos-2	GTGTCCGC	AATCCG	CCGAGTGATA	GAACCCAGAGG	CCTTCTCCTCC	CCTCCATGI	CCAATGCCC	CTCGCC 350	
		360	370	380	390	400	410	420	
			1	1	]			1	
CnCos-1	TCGTCTAC	ATCGTC	CAAGGAAGGG	GTATCGTTGGA	CTTGTGATGCC	TGGCTGCCC	AGAAACTTT	CCAATC 420	
CnCos-2	TCGTCTAC	ATCGTC	CAAGGAAGGG	GTATGGTTGGA	CTTGTGATGCC	TeecTeccc	CGAAACTTT	CCAATC 420	
			INDEL 1						
		430	440	450	460	470	480	490	
		•   • • • •	• • • •   • • • •	1	.			1	
CnCos-1	CTTCCAGO	GATCCG	A	ACGCGAGGAAG	GCGAGCGACA	CGGTGGTCC	AGAGATGAA	CACCAA 481	
CnCos-2	CTTCCAGC	GATCCG	AGCAGTATGA	ACGCGAGGAAG	GCGAGCGACAC	CGGAGGICI	AAAGACGAA	CACCAA 490	
		500	510	520	530	540	550	560	
and and 1		-		C A C C M C C M C C M C C M C C M C C M C C M C C M C C M C C M C C M C C M C C M C C M C C M				EE1	
ChCos-1	AAAGTCTA	CCAGTT	CCAAGAGGGG	GACGTCCTGGC	TGTGCCTAATC	CAMPACTI	ACTGGTGCT.	ACAACA JJI	
Cheus-2	AAAGTCTA	CCAGTT	CCAAGAGGGGA	GACGTCCTGGC	TGTGCCTAATG	GATTIGCTI	ACTGGTGCT	ACAACG J60	
		570	580	590	600	610	620	630	
CnCos-1	ATGGGGAG	:A ATCCT	GTTGTCGCG2	TCACCETC	GACACCAGCA	CGACGCTAA	CCAGCTCGA	TCGCAG 621	
CnCos-2	ATGGGGAG	AATCOT	GTGGTCGCGA	TCACCGTCCTC	GACACCAGCA	CGACGCTAA	CCAGCTCGA	TCGCAG 630	
011000 0									
			650	660	670 	680		700	
CnCos-1	CCACAGAC	AATTCT	TATTGGCTGG	AAGGCAGGAGC	AAGGCCGGCAA	CGATATGG	CGTGAGGGG	AGCATC 691	
CnCos-2	CCACAGAC	AATTCT	TATTGGCTGG	AAGGCAGGAGG	AAGG <mark>CC</mark> GGCA <i>I</i>	CGATATCGC	CGTGAGGAG	AGCATC 700	
		710	720	730	740	750	760	770	
			1	1	.			1	
CnCos-1	AAGGAAAA	CATCCT	GAGAGGGTTC	AGCACCGAGTT	GCTGGCAGCGG	CTTTTGGC	TTAACATGG	AG <mark>CT</mark> GG 761	
CnCos-2	AAGGAAAA	CATCCT	GAGAGGGTTC	GGCACCGAGTT	GCTGGCAGCAG	CTTTTGGCA	TTGACATGG.	AG <mark>CT</mark> GG 770	
		790	790	900	810	820	830	840	
							• • • •   • • • •	1	
CnCos-1	CAAGGAAG	CTCCAG	TGCAGGGAT	GACACGAGGGGC	GAGATCGTCC	ggg <mark>c</mark> ggaga	ACGGGCTTCA	GGTTCT 831	L
CnCos-2	CAAGGAAG	CTCCAG	TGCAGGGAT	GACACAAGGGGC	GAGATCGTCC	GGG <mark>C</mark> GGAGA	ACGGACTTCA	GGTGCT 840	)
				IND	FI 2				
		820	830	840	850	860	870	880	
		820	830	840	850 • • • •   • • • •	860 • • • •   • • • •	870 	880 	
CnCos-1	GAGGCCC	820 -   PCAGGG#	830    \ <b>T</b> GGAGGAAG.	840    AAGAGAGGGAAG	850    AGGGTAGAAG	860   TATAAATGG.	870    ATTCGAGGAG	880    GACCTAT 90:	1
CnCos-1 CnCos-2	GAGGCCC	820 •   • • • PCAGGGA PCAAGGA	830    \ <b>T</b> GGAGGAAG. \GGGAGGAAG.	840     AAGAGAGGGAAG AAGAGAGA	850    AGGGTAGAAG AAAAG	860   TATAAATGG. TATAAATGG.	870    ATTCGAGGAG ATTGGAGGAG	880    GACCTAT 90: GACCTAT 90:	L L
CnCos-1 CnCos-2	GAGGCCC!	820 .   PCAGGGA PCAAGGA	830    A <b>T</b> GGAGGAAG. AGGGAGGAAG.	840 -       AAGAGAGGGAAG AAGAGAGA	850    AGGGTAGAAG AAAAG	860   TATAAATGG. TATAAATGG.	870    ATTCGAGGAG ATTGGAGGAG	880    SACCTAT 90: SACCTAT 90:	L L
CnCos-1 CnCos-2	GAGGCCC	820 	830    ATGGAGGAAG. AGGGAGGAAG. 930	840     AAGAGAGGGAAG AAGAGAGA 940	850    AGGGTAGAAG AAAAG 950	860   TATAAATGG. TATAAATGG. 960	870    ATTCGAGGAG ATTGGAGGAG 970	880    SACCTAT 90: SACCTAT 90: 980	1
CnCos-1 CnCos-2	GAGGCCC:	820 .   PCAGGG/ PCAAGG/ 920 .	830    ATGGAGGAAG. AGGGAGGAAG. 930 	840    AAGAGAGGGGAAG AAGAGAGA 940 	850    AGGGTAGAAG AAAAG 950 	860   TATAAATGG. TATAAATGG. 960 	870    ATTCGAGGAG ATTGGAGGAG 970 	880    SACCTAT 90: SACCTAT 90: 980 	1
CnCos-1 CnCos-2 CnCos-1	GAGGCCCS GAGGCCCS 	820 ICAGGG/ ICAAGG/ 920 ICAAGA	830    ATGGAGGAAG. AGGGAGGAAG. 930    TTAAGCAGAA	840    AAGAGAGGGAAG AAGAGAGA 940     CATTGGGGGATCC	850    AGGGTAGAAG AAAAG 950    AAGGCGTGCC	860 TATAAATGG. TATAAATGG. 960 GACGTCTTC.	870    ATTCGAGGAG ATTGGAGGAG 970    AACCCAAGGG	880    SACCTAT 90: SACCTAT 90: 980    SGCGGAA 97:	L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCS GAGGCCCS TGCTCGAS TGCTCGAS	820 CAGGG/ FCAAGG/ 920 .   FGAAGAT	830    ATGGAGGAAG. AGGGAGGAGGAAG. 930    TTAAGCAGAA	840     AAGAGAGGGAAG AAGAGAGAG 940 .    CATTGGGGATCC CATTGGGGATCC	850    AGGGTAGAAG AAAAG 950    AAGGCGTGCC	860 TATAAATGG. TATAAATGG. 960 GACGTCTTC. GACGTCTTC.	870 II ATTCGAGGAG ATTGGAGGAG 970 II AACCCAAGGG AACCCAAGGG	880 	L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGA TGCTCGA	820 FCAGGG/ FCAAGG/ 920 . I FGAAGAT	830    ATGGAGGAAG. AGGGAGGAGGAAG. 930    TTAAGCAGAA	840 AAGAGAGGGAAGA AAGAGAGGAGA AAGAGAGAGA 940 JATTGGGGATCC CATTGGGGATCC	850    AGGGTAGAAG AAAAG 950    AAGGCGTGCC	860 TATAAATGG. TATAAATGG. 960   GACGTCTTC. GACGTCTTC.	870 II ATTCGAGGAG ATTGGAGGAG 970 II AACCCAAGGG AACCCAAGGG	880    SACCTAT 90: SACCTAT 90: 980    SGCGGAA 97: SGAGGAA 97:	1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC GAGGCCC TGCTCGA TGCTCGA	820 CAGGGA PCAAGGA 920 . I PGAAGAT PGAAGAT 990	830 II AGGGAGGAAG. 930 II TAAGCAGAA TAAGCAAAA 1000	840 AAGAGAGGGAAGA AAGAGAGGGAAGA 940 	850    AGGGTAGAG 950    AAGGCGTGCC AAGGCGTGCC 1020	860 TATAAATGG TATAAATGG 960 	870 II ATTCGAGGAG ATTGGAGGAG 970 II AACCCAAGGG 1040	880 SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGAGGAA 97: 1050	1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT	820 PCAGGGA PCAAGGA 920 PGAAGAT PGAAGAT 990 - 1	830 	840 .       AAGAGAGGAAG AAGAGAGAGA 	850    AGGGTAGAAG AAAAG 950    CAAGGCGTGCC 1020 	860 TATAAATGG. TATAAATGG. 960 	870 	880  SACCTAT 90: SACCTAT 90: 980  3GCGGAA 97: 3GAGGAA 97: 1050 	L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCS GAGGCCCS TGCTCGAS TGCTCGAS GGATTACS	820 PCAGGGA PCAAGGA PCAAGGA P20 PGAAGAD PGAAGAD PGAAGAD P90 PACCCTC	830 MTGGAGGAAG. AGGGAGGAAG. 930 MTAAGCAGAA MTAAGCAGAA 1000 1 CAACAGCGAG.	840 AAGAGAGGAAG AAGAGAGGAAG AAGAGAGAGA 940 CATTGGGGATCC CATTGGGGATCC 1010 111	850 	860 TATAAATGG TATAAATGG 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA	870 ATTCGAGGAG ATTGGAGGAG 970 II AACCCAAGGG AACCCAAGGG 1040 II GTGCTGAGAG	880 SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGAGGAA 97: 1050 11 SGGTGGT 100	L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGA TGCTCGA GGATTACC GAATTAC	820 	830 TGGAGGAAG. GGGAGGAAG. 930 1 TAAGCAGAA. TAAGCAGAA. 1000 1 CAACAGCGAG. CAACAGCGAG.	840 840 AAGAGAGGAAG AAGAGAGGAAG AAGAGAGAGA 940 	850 850 AAAGGTAAAG 950 950 1	860 TATAAATGG TATAAATGG 960 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA TCCAAATGA	870 ATTCGAGGAG ATTGGAGGAG 970  AACCCAAGGG AACCCAAGGG 1040  GTGCTGAGAG GTGCTGAGAG	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGCGGAA 97: 1050 11 SGCGGAT 104 SGGTGGT 104	L L L L £1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGA TGCTCGA GGATTACC	820 PCAGGGA PCAGGGA 920 PGAAGAD PGAAGAD PGAAGAD 990 PGACCTCTC TACCCTCC	830 	840 	850 	860 TATAAATGG. 960  GACGTCTTC. GACGTCTTC. 1030  TCCAAATGA TCCAAATGA	870 IJATTCGAGGAG ATTGGAGGAGAG 970 IJ AACCCAAGGG AACCCAAGGG 1040 IJ GTGCTGAGAG GTGCTGAGAG	880 SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGAGGAA 97: 1050 11 SGGTGGT 10: SGGTGGT 10:	1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT GGATTACC GAATTACC	820 	830 TGGAGGAAGAAGA AGGGAGGAAGA 930 TTAAGCAGAA TTAAGCAGAA 1000 1 CAACAGCGAGA CAACAGCGAGA 1070	840 	850 BAGGGTAGAAG 950 AAAGGCGTGCC AAAGGCGTGCC 1020 1020 CTCAGGTTCA CTCAGGTTCA 1090	860 TATAAATGG TATAAATGG 960 	870 ATTCGAGGAG ATTGGAGGAG 970 	880 SACCTAT 90: SACCTAT 90: 980 1 960 1 975 960 1 960 1 975 960 1 975 960 1 975 960 1 975 960 1 1 960 1 1 960 1 1 960 1 1 975 1 1 960 1	1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT TGCTCGAT GGATTACC GAATTAC	820 PCAGGA PCAAGGA 920 PGAAGAT 990 PACCTOC 1060	830 TGGAGGAAG, TGGAGGAAG, 930 TTAAGCAGAA TTAAGCAGAA 1000 	840 AAGAGAGGGAAG AAGAGAGGA 940 940 	850 SAGGGTAGAAG 950 950 1020 1020 1020 CTCAGGTTCA CTCAGGTTCA 1090	860 TATAAATGG TATAAATGG 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA TCCAAATGA 1100	870 II.ATTCGAGGAG ATTCGAGGAGAG 970 AACCCAAGGG AACCCAAGGG 1040 II.GTGCTGAGAG GTGCTGAGAG GTGCTGAGAG	880 SACCTAT 90: SACCTAT 90: 980 980 980 980 980 980 980 980	1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT GGATTACC GAATTACC	820 FCAGGA FCAAGGA 920 FGAAGAT FGAAGAT 990 FACCCTC FACCCTC 1060 	830 	840 	850 SSO SAGGGTAGAAG SSO SSO SSO SSO SSO SSO SSO SS	860 	870 I	880 SACCTAT 90: SACCTAT 90: 980 11 960 960 11 960 960 960 960 97: 1050 11 960 960 97: 1050 11 960 11 97: 1.050 11 97: 1.050 11 11 9	L L L L L L L L
CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGA1 TGCTCGA1 GGATTACC GAATTACC TCTCTAC TCTCTAC	820 PCAGGGA PCAGGGA 920 1060 PGAGAT PGAGAT PGAGAT PGAGAAC 1060 AGGAACC	830 	840 AAGAGAGGGAAG AAGAGAGGAGA 940 940 940 940 940 940 940 940	850 SAGGGTAGAAG 950 1020 1020 1020 1020 1020 CTCAGGTTCA CTCAGGTTCA 1090 1.4CATCAATGC	860 TATAAATGG TATAAATGG 960 GACGTCTTC 1030 TCCAAATGA TCCAAATGA 1100 100 CCACAGCAT CCACAGCAT	870 ATTCGAGGAG ATTCGAGGAG 970 970 1 AACCCAAGGG AACCCAAGGG 1 1040 1 GTGCTGAGAG GTGCTGAGAG GTGCTGAGAG CATGTATTGG CATGTATTGG	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 980 980 980 980 980 980 980	L L L L L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGA: TGCTCGA: GGATTACC GAATTACC TCTCTAC, TCTCTAC,	820 	830 AGGGAGGAAG, 930 TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, 1000 1 CAACAGCGAG, CAACAGCGAG, CAACAGCGAG, 1070 1 SCCATGGTAT, SCCATGGTAT,	840 AAGAGAGGGAAG AAGAGAGA 940 940 SATTGGGGATCC CATTGGGGATCC 1010 1010 1010 1010 1010 1010 1010 1	850 850 850 950 950 1020	860 TATAAATGG 760 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA 1100 1100 CCCACGCAT	870 	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 1050	L L L L L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAC TGCTCGAC GGATTACC GGATTACC GGATTACC TCTCTAC	820 	830 Aregargaaaa Aregargaaaa Data area area area area area area area a	840 AAGAGAGGGAAGA AAGAGAGAGA AAGAGAGAGA CATTGGGGATCC 1010 1010 1010 1010 1080 10	850 AGGGTAGAAG 950 AAAGGCGTGCC AAAGGCGTGCC AAGGCGTGCC 1020 CTCAGGTTCA CTCAGGTTCA 1090 ACATCAATGC ACATCAATGC 1160	860 <b>TATAAATGG</b> <b>960</b> 	870 ATTGGAGGAC 970 970 1040 STGCTGAGGAC 1040 STGCTGAGGAC 1110 CATGTATTGC CATGTATTGC 1180	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGAGGAA 97:	L L L 1 1 1 1 L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT GAATTACC TCTCTAC, TCTCTAC,	820 	830 ATGGAGGAAG, AGGGAGGAAG, 930 TTAAGCAGAA TTAAGCAGAA 1000 1000 CAACAGCGAG, CAACAGCGAG, 1070 CCCATGGTAT CCCATGGTAT 1140 1	840 AAGAGAGGGAAG AAGAGAGGA 940 940 940 940 940 940 940 940	850 SAGGGTAGAAG 950 10.10 SAAGGCGTGCC AAAGGCGTGCC 1020	860 TATAAATGG 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA 1100 CCACAGCAT CCACAGCAT	870 ATTCGAGGAGAA 970 4ACCCAAGGGAGAC 970 104	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 980 980 980 980 980 980 980	L L L 41 L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-7	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT GGATTACC GGATTACC TCTCTACC CGGACGAGC GGACGAGC	820 CAGGAGA PCAAGGA PCAAGGA PCAAGA PGAAGA PGAAGA P90 LOCO LOCO AGGAACC AGGAACC AGGAACC LI30 	830 AGGGAGGAAG, AGGGAGGAAG, 930 TTAAGCAAAA TTAAGCAAAA TTAAGCAAAA 1000	840 AAGAGAGGAAG AAGAGAGA 940 940 940 1010 1020 100 10	850 850 AGGGTAGAAG 950 950 1020	860 TATAAATGG 960 960 1030 TCCAAATGA 1030 1030 CCACAGCAT 100 100 100 100 100 100 100 10	870 870 877 970 970 1040 1040 1040 1140 110 110 110	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGGGGAA 97: SGGGGAA 97: SGGGGAA 97: SGGGGGA 97: SGGGGGA 10: SGGTGGT 10: SGGTGGT 10: SGGGGGA 11: SACCGGA 11: SACCGGA 11: SAGGGCC 11: SGGGGCC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGC 11: SGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGGC 11: SGGGC 11: SGGC 11: SGGGC 11: SGGGC 11: SGGC 11: SGGGC 11: SGGC 11: SGGGC 11: SGGGC 11: SGGC 11: SGGC 11: SGGC 11: SGGC 11: SGGGC 11: SGGC 11: SGGGC 11: SGGC 11: SGGC 11: SGGC 11: SGGGC 11: SGGGC 11: SGGC 11: SGGGC 11: SGGGGC 11: SGGGGC 11: SGGGC 11: SGGGGC 11: SGGGC 11: SGGGGC 11: SGGGC 11: SGGGGC 11: SGGGGG 11: SGGGGGG 11: SGGGGG 11: SGGGGG 11: SGGGGG 11: SGGGGG 11: SGGGGG 11: SGGGGG 11: SGGGGGG 11: SGGGGG 11: S	
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGA TGCTCGA GGATTAC GAATTAC TCTCTAC GGACGAGGGGACGAGG	820 	830 ATGGAGGAAG AGGGAGGAAG 930 TTAAGCAGAA TTAAGCAGAA 1000 CAACAGCGAG AACAGCGAG AACAGCGAG CAACAGCGAG 1070 CCATGGTAT CCATGGTAT 1140 CCAGGTTGC CCAGGTTGC	840 AAGAGAGGGAAG AAGAGAGAGA AAGAGAGAGA 940 940 940 940 940 940 940 940	850 AGGGTAGAAG 950 1020 1020 1020 1020 1020 1020 1020 1020 1020 1050 1090 1000 10	860 TATAAATGG 960 GACGTCTTC 1030 TCCAAATGA 1100 CCACAGCAT CCACAGCAT 1170 TTCCACCGA	870 ATTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 10: SGAGGAA 10: SGAGGAA 11: SGAGGAA 11:	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAI TGCTCGAI GGATTACC GGATTACC TCTCTAC, TCTCTAC, GGACGAGG GGACGAGG	820 	830 AGGGAGGAAG, AGGGAGGAAG, 930 TTAAGCAGAA TTAAGCAGAA 1000 1 CAACAGCGAG, CAACAGCGAG, CAACAGCGAG, 1070 1 SCCATGGTAT SCCATGGTAT CGAGGTTGC' CGAGGTTGC'	840 AAGAGAGGGAAGA AAGAGAGGA 940 940 940 940 940 940 940 940	850 850 AGGGTAGAAG 950 950 1020 1020 1020 1020 1020 1020 1020 1090 1090 1090 1090 1090 1090 100	860 TATAAATGG 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA 1100 CCACAGCAT CCCACAGCAT 1170 1170	870 870 ATTGAGGAGAA ATTGGAGGAC 970 970 104	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: 1050 11 SGGTGGT 10- SGGTGGT 10- SGGTGGT 10- SGGTGGT 11- SGGTGGT 11- SGAGGTC 11-	1 1 1 11 11 11 11 11 11
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCC GAGGCCC TGCTCGAI TGCTCGAI GGATTAC GGATTAC TCTCTAC, GGACGAGG GGACGAGG	820 	830 	840 	850 850 AGGGTAGAAG 950 1020 10	860 	870 1 ATTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 960 1 960 1 960 1 960 1 960 1 960 1 960 1 97: 960 1 9 100 100 1 9 100 100 100 100 100 100 100 100 100 10	L L 1 11 11 11 31 31 31
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAC TGCTCGAC GGATTACC GAATTACC TCTCTACC TCTCTACC GGACGAGG GGACGAGG	820 	830 ATGGAGGAAG, SGGGAGGAAG, 930 TAAGCAGAA, TAAGCAGAA, 1000 AACAGCGAG, AACAGCGAG, CAACAGCGAG, 1070 CCATGGTAT, CCATGGTAT, 1140 CCGAGGTTGC CGGAGTTGC	840 840 AAGAGAGGGAAGA AAGAGAGAGA 940 940 940 940 940 940 940 940	850 SAGGGTAGAAG SAGGGTAGAAG SAAGGCGTGCC 1020	860 TATAAATGG 960 960 960 960 960 960 960 960	870 ATTGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGGGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 10: 11 SGGTGGT 10: SGAGGTC 11: SGAGGTC 11: 1.260 11 SGGTATC 12: SGGTATC 12: SGG	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAC TGCTCGAC GGATTACC GGATTACC GGACTACC GGACGAGG GGACGAGG GGACGAGG	820 ICAGGGA PCAAGGA 920 IGAAGAT 930 IGAAGAT 990 IACCTCT 1060 AGGAACC 1130 IL30	830 	840 AAGAGAGGAAG AAGAGAGA 940 940 1010 1000 1000 1000 1000 1000 100	850 850 AGGGTAGAAG 950 950 1020	860 860 TATAAATGG 960 960 960 960 1030 TCCAAATGA 1030 10 10 10 10 10 10 10 10 10 1	870 870 ATTGAGGAGAA ATTGGAGGAG 970 970 1040 105	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGGGGAA 97: SGGGGAA 97: SGGGGGA 97: SGGGGGA 97: SGGGGGA 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGGGGT 11: SGGGGC 11: SGAGGC 1	L L L H H H H H H H H H H H H H H H H H
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGAI TGCTCGAI GGATTAC GGATTAC TCTCTAC TCTCTAC GGACGAG GGACGAG GGACGAG	820 920 920 920 920 930 930 930 930 930 930 930 930 930 93	830 	840 840 AAGAGAGGAAGA AAGAGAGAAGA AAGAGAGAA AAGAGAGAA 1010 1000 1000 1000 1	850 850 AGGGTAGAAG 950 AAAGGCGTGCC 1020 10	860 TATAAATGG 960 960 960 960 960 960 970 960 970 970 970 970 970 970 970 97	870 870 870 877 877 877 877 877	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 11 SGCGGAA 97: SGAGGAA 97: SGAGGAA 97: SGAGGAA 97: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 10: SGGTGGT 11: SACCGGA 11: SACCGGA 11: SGGTGGT 11: SGGTGGT 11: SGGTGTC 12: SGGTATC 12: SGGTGTC 12:	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAT TGCTCGAT GGATTACC GAATTACC TCTCTAC GGACGAGG GGACGAGG GGACGAGG	820 920 920 920 930 940 940 940 940 940 940 940 94	830 	840 840 AAGAGAGGGAAGA AAGAGAGGAGA 940 940 940 940 940 940 940 940	850 SAGGGTAGAAG 950 10.0 1	860 TATAAATGG TATAAATGG 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA 1100 	870 870 870 870 877 877 970 970 970 970 970 970 970 9	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 10: SACCTAT 11: SACCTAT 11:	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAI TGCTCGAI GGATTACC GGATTACC CTCTCAC CTCTCAC GGACGAGG GGACGAGG CGACGAGG AGCTCTTC	820 10-0	830 AGGGAGGAAG, AGGGAGGAAG, 930 TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, 1000 100	840 840 AAGAGAGGAAG AAGAGAGAA AAGAGAGAA 940 940 940 1010 1020	850 850 AGGGTAGAAG 950 AAAGCGTGCC AAAGCGTGCC 1020 100	860 860 TATAAATGG 960 960 1030	870 870 870 877 877 877 877 877	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 980 1 980 1 1050 1 SGGGGAA 97: SGGGGAA 97: SGGGGGA 97: SGGGGGA 97: SGGGGGA 97: SGGGGGA 10: SGGGGGGT 10: SGGGGGGG 10: 1 1.20 1 SGGGGGGG 10: SGGGGGGG 11: SGGGGGG 11: SGGGGGC 11: SGGGGC 11: SGGGCC 11: SGGGGC 11: SGGGGGC 11: SGGGGGC 11: SGGGGGC 11: SGGGGGC 11: SGGGGGC 11: SGGGGGGC 11: SGGGGGC 11: SGGGGGGC 11: SGGGGGGGC 11: SGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGAC TGCTCGAC GGATTACC GAATTAC TCTCTAC TCTCTAC GGACGAG GGACGAG GGACGAG GGACCAC AGCTCTT	820 PCAGG4 PCAGG4 920 920 920 930 940 940 940 940 940 940 940 94	830 ATGGAGGAAG, AGGGAGGAAG, 930 TTAAGCAGAA, TTAAGCAGAA, CAACAGCGAG, AACAGCGAG, 1070 CCATGGTAT, CCATGGTAT, 1140 1140 CCAGGTTGC 1210 CCGGAAAAC CCGGCAAAAC	840 840 AAGAGAGGGAAGA AAGAGAGAGA AAGAGAGA AAGAGAGA AAGAGAGA AAGAGAGA AAGACCCCGATC AAGCTCCCACTGA AGCTCCCCACTGA AGCTCCCACTGA AGCTCCCCACTGA AGCTCCCCACTGA AGCTCCCCACTGA AGCTCCCCACTGA AGCTCCCCCACTGA AGCTCCCCACTGA AGCTCCCCACTGA AGCTCCCCCACTGA AGCTCCCCCCCCACTGA AGCTCCCCACTGA AGCTCCCCCACTGA AGCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	850 SAGGGTAGAAG 950 SAGGGTGCC CAAGGCGTGCC 1020 CTCAGGTTCA CTCAGGTTCA 1090 1160 1160 1160 1230 1230 1300 1300 1300 1000 1	860 TATAAATGG TATAAATGG 960 GACGTCTTC 1030 TCCAAATGA TCCAAATGA 1100 CCCACAGCAT 1170 TTCCACGGAT 1170 1240 GAAGTGCAG GAAGCGAAG 1310	870 870 870 877 877 877 877 877	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 1000 100	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAY TGCTCGAY GGATTACC GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GGACCAAGG GATCAAAG	820 1.1	830 	840 840 AAGAGAGGAAGA AAGAGAGA 940 940 1010 1000 1000 1000 1000 100	850 850 AGGGTAGAAG 950 950 1020	860 860 TATAAATGG 960 960 960 1030 TCCAAATGA 1030 1030 TCCAAATGA 1100 CCACAGCAT 1100 1170 1170 1170 1240 1240 1310 34ACTCGAG	870 870 ATTGAGGAGAA ATTGGAGGAG 970 970 1040 105	880           SACCTAT           SAGGGAA           SACCGA           SACCGGA           SACCGGA           SACCGGA           SAGGGTC           SATGCCG           SATGCCG	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCC GAGGCCC TGCTCGAI TGCTCGAI GGATTAC GGATTAC GGACGAG GGACGAG GGACGAG AGCTCTT GATCAAG GATCAAG GATCAAG	820 1	830 330 345 355 350 350 350 350 350 350 35	840 840 AAGAGAGGGAAG AAGAGAGAA AAGAGAGAA AAGAGAGAA 940 940 940 1010 CATTGGGGATCC 1010 1000 100	850 850 AGGGTAGAAG 950 1020 10	860 860 TATAAATGG 960 960 960 1030 TCCAAATGA 1030 TCCAAATGA 1000 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1170 1170 1170 1170 1170 1240 GAAGGCAGGAG 1310 GACGTCGGC GACGTCGGC	870 870 870 877 877 877 877 877	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: 960 1 950 1 950 1 950 1 950 1 950 1 950 1 950 1 950 1 950 1 950 1 950 1 1050 1	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCCC TGCTCGAY TGCTCGAY GGATTACC GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACGAGG GATCAAGG GATCAAGG	920 920 920 920 920 930 940 940 940 940 940 940 940 94	830 	840 840 AAGAGAGGGAAGA AAGAGAGAGA 940 940 940 940 940 940 940 940	850 850 AGGGTAGAAG 950 1020 100	860 TATAAATGG TATAAATGG 960 GACGTCTTC GACGTCTTC 1030 TCCAAATGA 1100 CCACAGCAT 1100 1170 1170 1170 1240 1240 1310 GACGTCGCC GACGTCGCC 1380	870 870 870 870 877 877 877 877	880 SACCTAT 90: SACCTAT 10: SACCTAT 90: SACCTAT 10: SACCTAT 11: SACCTAT 12: SACCTAT 12:	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAI TGCTCGAI GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACCAGG GGACCAGG GATCAAGG GATCAAGG	820 10	830 AGGGAGGAAG, AGGGAGGAGA, 930 TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, 1000 1000 CAACAGCGAG, CAACAGCGAG, 1070	840 840 AAGAGAGGAAG AAGAGAGAA AAGAGAGAA 940 940 1010 100 1010 1000 1000 1000 1000	850 850 AGGGTAGAAG 950 AAGGCGTGCC AAGGCGTGCC 1020 10	860 860 TATAAATGG 7ATAAATGG 960 960 1030 10 1030 10	870 870 ATTGAGGAGAA ATTGGAGGAA 970 970 1040 105	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 1050 105	L L L L L L L L L L L L L L L L L L L
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAI TGCTCGAI GGATTACC GAATTACC TCTCTACL TCTCTACL GGACGAGG GGACGAGG GGACGAGG GATCAAG GATCAAGG	820         1	830 	840 840 AAGAGAGGAAGA AAGAGAGAGA AAGAGAGAA AAGAGAGA AAGAGAGA AAGAGAGA AAGCTCCCGATC AAGCTCCCGCTCCA AAGCTCCCGCTCCA AAGCTCCCGCTCCAAGCTCCA	850 850 AGGGTAGAAG 950 AAAGGCGTGCC 1020 1020 CTCAGGTTCA CTCAGGTTCA 1090 10	860 860 TATAAATGG 960 960 960 960 970 960 970 970 970 970 970 970 970 97	870 870 ATTGAAGAAAAAACCCAAGGAA 970 104	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 1000 100	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGA TGCTCGA GGATTAC GGATTAC GGATTAC GGATTAC GGACGAG GGACGAG GGACGAG GGACGAG GATCAAG GATCAAG GATCAAG GATCAAG	820         820           1         1           PCAS65/         1           PCAS65/         920           920         920           930         1           PGASGAT         1           RACCCTC         1           1060         1           1130         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1200         1           1340         1           1340         1	830 	840 840 AAGAGAGGAAGA AAGAGAGAA 940 940 1010 1000 1000 1000 1000 100	850 850 AGGGTAGAAG 950 1020 100	860 860 TATAAATGG 7ATAAATGG 960 960 1030 1040	870 870 ATTGAAGAA ATTGGAGGAC 970 970 1040 1050	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGGGA 11:           SACCGGA 11:           SAAGGT:           SAAGGT:           SAAGGT:           SGAGGGG:           SAAGGT:	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAI TGCTCGAI GGATTACC GGATTACC CTCTTAC, CTCTCTAC, CTCTCTAC, GGACGAGG GGACGAGG GGACGAGG GATCAAG, GATCAAG, GATCAAG, GAGGAGGG	820           1	830 AGGGAGGAAG, AGGGAGGAGA, 930 TTAAGCAAAA, TTAAGCAAAA, TTAAGCAAAA, TTAAGCAAAA, 1000 10	840 840 AAGAGAGGAAGA AAGAGAGAGA 940 940 940 940 1010 1000 1000 100	850 850 AGGGTAGAAG 950 AAGGCGTGCC AAGGCGTGCC 1020 1020 CCTCAGGTTCA 1090 1090 1090 1090 1090 1090 1090 1160 1160 1160 1160 1160 1230 1230 1300 1300 1370 17	860 860 TATAAATGG 960 960 960 960 1030 1	870 870 870 877 970 970 1040 1040 1040 1040 1040 1040 1040 10	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 1050 105	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCC GAGGCCC TGCTCGAC TGCTCGAC GGATTAC GGATTAC TCTCTAC TCTCTAC TCTCTAC GGACGAG GGACGAG GGACGAG GGACGAG GATCAAG GATCAAG	820           PCAAGG4           PCAAGG4           920           PCAAGG4           920           PCAAGG4           930           PCAAGG4           940           PCAAGG4           950           1060           1130           PCCCCTCG1           1200           PCCCCCCCG12           1200           PCCCCCCCCG12           1340           PCCCCGCG1340           1410	830 	840 840 AAGAGAGGGAAGA AAGAGAGAGA AAGAGAGAA 940 940 1411 CATTGGGGATCC AAGCTCCCGATC AAGCTCCCGATC 1010 1000 1010 1000 10	850 SAGGGTAGAAG SS0 SS0 SS0 SS0 SS0 SS0 SS0 SS	860 TATAAATGG 960 960 GACGTCTTC 1030 TCCAAATGA TCCAAATGA 1100 CCACAGCAT CCACAGCAT 1170 TTCCACGGA TTCCACGGA 1240 1240 1310 GACGTCGGC GAAGTGCAAG GAAGTGCAAG 1310 GACGTCGGC 1380 AGGAGGGTC AGGAGGGTC 1380 AGGAGGGTC 1450 STCCACGCACGCACGCACGACGACGACGACGACGACGACGAC	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGGTGGT 10:           SGGTGGT 10:           SGGTGGT 10:           SGGTGGT 10:           SGGTGGT 10:           SGGTGGT 10:           SGGTGGT 11:           L120           L1           SACCGGA 11:           SGAGGTC 11:           SGAGGTC 11:           SGAGGTC 11:           SGAGGTC 12:           SGAGGTC 13:           SATGCCG 13:           SATGCCG 13:           SGGGGGG 13:           SGAGGGG           SGAGGGG	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGA TGCTCGA GGATTAC GGATTAC GGACGAG GGACGAG GGACGAG GGACCAG GGACCAG GGACCAG GGACCAG GGACCAG GGACCAG GGACCAG GATCAAG GATCAAG	820         820           PCAAGG4         PCAAGG4           PCAAGG4         920           PEAAGA1         PCAAGA1           PEAAGA1         PCAAGA1           PA0         920           P10         PCAAGA1           PA0         920           P11         PCAAGA1           PA0         920           P11         PCACCACC           P11         PCACCACC           P11         PCACCACC           P11         PCACCACC           P1200         PCACCACCC           P1200         PCACACACCC           P1200 <td>830 </td> <td>840 840 AAGAGAGGAAGA AAGAGAGA 940 940 1010 1000</td> <td>850 850 AGGGTAGAAG 950 AAGGCGTGCC AAGGCGTGCC 1020 100</td> <td>860 TATAAATGG TATAAATGG 960 GACGTCTTC 1030 TCCAAATGA 1100 CCACAGCAT CCACAGCAT 1170 1170 1170 1170 1240 GAAGGCGAC 1310 1240 1310 1310 GACGTCGGC 1380 1380 AGGAGGGTC 1380 1390 1380</td> <td>870 870 870 877 877 877 877 877</td> <td>880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGAA 10:           SGAGGAA 10:           SGAGGAT 10:           SACCGGA 11:           CACCGGA 11:           SAAGGTC 116           SGAGGGT 112:           ICGTATC 12:           ICGTATC 12:           ICGTATC 12:           SATGCCG 13:           SATGCCG 13:           IATGCCG 13:           IATGCCG 13:           IATGCCG 13:           IATGCCG 13:           IATGCCG 13:           ITR           IATGC</td> <td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>	830 	840 840 AAGAGAGGAAGA AAGAGAGA 940 940 1010 1000	850 850 AGGGTAGAAG 950 AAGGCGTGCC AAGGCGTGCC 1020 100	860 TATAAATGG TATAAATGG 960 GACGTCTTC 1030 TCCAAATGA 1100 CCACAGCAT CCACAGCAT 1170 1170 1170 1170 1240 GAAGGCGAC 1310 1240 1310 1310 GACGTCGGC 1380 1380 AGGAGGGTC 1380 1390 1380	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGAA 10:           SGAGGAA 10:           SGAGGAT 10:           SACCGGA 11:           CACCGGA 11:           SAAGGTC 116           SGAGGGT 112:           ICGTATC 12:           ICGTATC 12:           ICGTATC 12:           SATGCCG 13:           SATGCCG 13:           IATGCCG 13:           IATGCCG 13:           IATGCCG 13:           IATGCCG 13:           IATGCCG 13:           ITR           IATGC	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCC GAGGCCC TGCTCGAI TGCTCGAI GGATTAC GGATTAC TCTCTAC TCTCTAC GGACGAG GGACGAG GGACGAG GATCAAG GATCAAG GATCAAG GATCAAG	820           1.1	830 AGGGAGGAAG, AGGGAGGAAG, 930 TAAGCAGAAA TAAGCAGAAA 1000 CAACAGCGAG, 1070 CAACAGCGAG, 1070 CCATGGTAT 1140 1210 CCGGCAAAAC 1210 CCGGCAAAAC 1280 SACCGGGCCA SACCGGGCCA 1350 TGAACTCCTA 1420 CCTCATCCTA	840 840 AAGAGAGGGAAGA AAGAGAGAGA AAGAGAGAGA AAGAGAGAG	850 850 AGGGTAGAAG 950 1020 100	860 7ATAAATGG 7ATAAATGG 960 960 960 1030 TCCAAATGA 1000 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1170 TTCGACGGA 1100 GACGTCGGC GACGTCGGC 1380 AGGAGGGTC 1380 AGGAGGGTC 100 STTGATGGC	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 10:           SACCTAT 10:           SACCTAT 10:           SACCTAT 10:           SACCTAT 11:           SACCTAT 11:           SACCTAT 12:           SATGCC 13:           SATGCC 14:           SATGCC 14:	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCCC TGCTCGAC TGCTCGAC GGATTACC GGATTACC GGATTACC GGATCACG GGACGAGG GGACGAGG GGACGAGG GATCAAGG GATCAAGG GATCAAGG GATGAAGG GAGGAGG CAGGAGG CAGGAGG	820 1000 1	830 	840 840 AAGAGAGGAAGA AAGAGAGAGA 940 940 1010 1000 10	850 850 850 850 950 1020	860 TATAAATGG TATAAATGG 960 960 960 1030 TCCAAATGA 1030 TCCAAATGA 1000 CCACAGCAT 1100 CCACAGCAT 1100 1170 1240 1240 1240 1240 1240 1380 AGGAGGGCC 1380 AGGAGGGCC 1380 AGGAGGGCC 1380 CTTGATGCG CTTGATGCG CTTGATGCG	870 870 870 870 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGAA 10:           SGAGGAC 11:           SGAGAGC 11:           SGAGAGC 12:           SGAGAGC 13:           SGGAGGA 14:           SGGAGGA 14:           SGGAGGG 14:	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCCC TGCTCGAC TGCTCGAC GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACGAGG GATCAAG GATCAAG GATCAAG GACGAGGC CAGGAGGC	820         820           PCASG5/         PCASG5/           PCASG5/         PCASG5/           P20         P20           PGASG5/         PCASG5/	830 	840 840 840 840 840 840 840 840 840 840	850 850 AGGGTAGAAG 950 1020 100	860 860 TATAAATGG 960 960 960 1030 TCCAAATGA 1000 TCCAAATGA 1100 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1100 1170	870 870 ATTGAGGAGAA ATTGGAGGAG 970 1040 1050 10	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 1050 105	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCC GAGGCCC TGCTCGAI TGCTCGAI GGATTAC GGATTAC TCTCTAC TCTCTAC GGACGAG GGACGAG GGACGAG GATCAAG GATCAAG GATCAAG GATCAAG GATGAAGG GAGGAGG	820         1	830 ATGGAGGAAG, AGGGAGGAAG, 930 TAAGCAGAAA TAAGCAGAAA 1000 CAACAGCGAG, 1000 CAACAGCGAG, 1070 CCATGGTAT CCATGGTAT 1140 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAAAC 1210 CCGCGCAAACC 1210 CCGCGCCAC 1210 CCGCGCCACCCCA 1350 CGAACTCCTA CGAACTCCTA CGAACTCCTA 1420 TTCACTCCT. DEL 3 1490	840 840 AAGAGAGGAAGA AAGAGAGAGAAGA AAGAGAGAG	850 850 850 850 850 850 850 850	860 860 TATAAATGG 960 960 960 960 960 960 960 960	870 870 870 870 877 877 877 877	880 SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: SACCTAT 90: 1000 10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2	GAGGCCC GAGGCCC TGCTCGA TGCTCGA GGATTAC GGATTAC GGATTAC GGACGAG GGACGAG GGACGAG GGACGAG GGACGAG GATCAAG GATCAAG GATCAAG GATCAAG GAGGAGG GACGAGG CAGGGGT	820         820           1         1           PCASGG/         1           PCASGG/         920           920         920           940         920           950         1           PCACTCO	830 	840 840 AAGAGAGGAAGA AAGAGAGAGA 940 940 1010 1000 10	850 850 850 850 950 1020	860 860 TATAAATGG 960 960 960 960 1030 TCCAAATGA 1100 CCACAGCATC 1030 100 CCACAGCAT 1100 CCACAGCAT 1170 1170 1170 1170 1240 1240 1310 1240 1310 1380	870 870 870 870 870 870 970 970 1040 10	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGGAGGA 10:           SGAGGAC 11:           SGAGGAC 11:           SGAGGAC 11:           SGAGGAC 11:           SATGCCG 11:           SATGCCG 12:           SATGCCC 13:           SATGCCC 13:           SATGCCC 13:           SATGCCC 13:           SATGCCC 13:           SATGCCC 13:           SGAGAG 14:           SGGAGAG 14:           SGGAGAG 14:           SGAGAG 14: <td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAI TGCTCGAI GGATTACC GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GGACGAGG CATCAAG GATCAAG GATCAAG GATCAAG GACGAGGT ACGAGGT ACGAGGT	20     20	830 837 337 347 357 357 357 357 357 357 357 35	840 840 AAGAGAGGGAAG AAGAGAGGAAG AAGAGAGAGA 940 940 940 1010 CATTGGGGATCC 1010 1010 CATTGGGGATCC 1010 1010 CATCGGGATCCCGATC 1080	850 850 AGGGTAGAAG 950 1020 100	860 860 TATAAATGG 960 960 960 960 960 960 960 960	870 870 870 870 970 1040 1040 1040 1040 1040 1040 1040 10	880           SACCTAT 90:           SACCTAT 10:           SACCTAT 10:           SACCTAT 10:           SACCTAT 10:           SACCTAT 10:           SACCTAT 11:           SACCTAT 11:           SACCTAT 12:           Incontract 12:           SAGGTAT 12:           SAGGTAT 12:           Incontract 12:           SAGGTAT 12:           Incontract 13:           SATGCCG 13:           SATGCCG 13:           Incontract 13:           SAGGAGGT 14:           SAGGAGT 14:           SAGGAGT 14:           SAGGAGT 14:           SAGGAGT 12:	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCCCCCCCCCCCCCCCCCCCCCCCC	820         1           PCAAGGA         1           PCAAGGA         920           PCAAGGA         920           PCAAGGA         920           PCAAGGA         920           PCAAGGA         920           PCAAGAT         930           PCAAGCCTC         1060           1.100         1.1.0           PCCCCTCG         1130           1.200         1.2.00           1.200         1.2.00           1.200         1.2.00           1.200         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00           1.2.00         1.2.00	830 	840 840 840 840 840 840 840 840 840 840	850 850 850 850 950 1020	860           TATAAATGG           960           GACGTCTTC.           1030           TCCAAATGA           1100           CCCACGCAT           1100           CCCACGCAT           1100           CCCCCCGCAT           1100           1170           1170           1170           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1240           1250           GACGTCGCC           1300           AGGAGGGTCA           AGGAGGGTCGC           1320           GTTGATGCG           1520	870 870 870 870 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGGTGGT 10-           LACCGGA 11:           CACCGGA 11:           CACCGGA 11:           SGAGGTC 112           SGAGGGC 13:           LATGCCG 13:           CATGCCG 13:           LATGCCG 13:           SGGGGGG 13:           SGGGAGG 14:           SGGAGGA 14:           SGGAGGG 14:           SGGAGGG 14:           SGGAGGG 14:           SGGAGG 14:           SGGAGG 14:           SGGAGG 14:           SGGAGG 14:           SGGAGG 14:           SGGAGGG 14:           SGGAGGG 14:	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1	GAGGCCC GAGGCCC TGCTCGA TGCTCGA GGATTAC GGATTAC GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACGAGG GATCAAG GATCAAG GATCAAG GATCAAG GACGAGGT CTGCTAG CTGCTAG	820         820           1         CCAAGG4           CCAAGG4         CCAAGG4           920         Saccard           Incoaccard         Saccard           1         Saccard	830 AGGGAGGAGA, AGGGAGGAGA, 930 TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, TTAAGCAGAA, 1000 1	840 840 AAGAGAGGAAG AAGAGAGAA 940 940 940 940 940 940 940 940	850 850 850 850 850 850 850 850	860 860 TATAAATGG 960 960 960 960 1030 TCCAAATGA 1030 TCCAAATGA 1100 CCACAGCATC 1030 CCACAGCATC 1030 100 CCACAGCATC 1170	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGAA 10:           SGAGGGA 10:           SGAGGGA 11:           SACCGGA 11:           SACCGGA 11:           SACCGGA 11:           SGAGGTC 116           SGAGGTC 116           SGAGGTC 116           SGAGGGC 13:           SATGCCG 13:           SATGCCG 13:           SATGCCG 13:           SATGCCG 13:           SATGCCG 13:           SATGCCG 13:           SGAGGGG 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 15:           STTATGT 15:           STTATGT 15:	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGA/ TGCTCGA/ GGATTACC GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GATCAAG, CTGCTAG, CTGCTAGAG, CTGCTAG, C	820 1.1	830 	840 840 AAGAGAGGGAAGA AAGAGAGAGA AAGAGAGAGA AAGAGAGAG	850 850 850 950 950 1020	860 860 TATAAATGG 960 960 960 1030 TCCAAATGA 1000 CCACAGCAT 1000 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1240 GAAGGCGGAG GAAGGCGGAG 1310 GACGGCGGAC 1380 AGGAGGGGCC 1380 AGGAGGGGCC 1520 GTGCCTAGCAC GTGCCCTAGCAC GTGCCCTAGCAC 1520 CTTGATGCG GTGCCCTAGCAC 1520 CTTGATGCG CTTGATGCC CTTGATGCG CTTGATGCA CTTGATGCA CTTGATGC CTTGATGCA CTTGATGCA CTTGATGCA CTTGATGCA CT	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 10:           SACCGGA 11:           SACCGGA 11:           SACCGGA 11:           SACCGGA 11:           SACCGGA 11:           SACCGGA 11:           SAGGGGC 11:           SAGGGGC 12:           SAGGGGGG 13:           SATGCCG 13:           SATGCCG 13:           SATGCCG 13:           SATGCCG 13:           SAGGGGGG 14:           SGGAGAG 14:           SGGAGAG 14:           SGGAGAG 14:           SGGAGAG 14:           SGGAGGG 14: </td <td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-1	GAGGCCCC GAGGCCCCCCCCCCCCCCCCCCCCCCCCCC	820         820           1         1           PCASGE/         1           PCASGE/         920           920         1           PGASGA/	830 	840 840 AAGAGAGGGAAGA AAGAGAGAGAGA 940 940 124TTGGGGATCC CATTGGGGATCC 1010	850 850 850 850 950 1020	860           TATAAATGG           TATAAATGG           960           GACGTCTTC           1030           TCCAAATGA           TCCAAATGA           1100           CCCACGCAT           1100           CCCCCAGCAT           1100           CCCCAGCAT           1170           1170           1170           1170           1240           1310           1240           1310           1240           1240           1240           1240           1310           1250           CTGATCGC           1380           AGGAGGCTAG           GTGCCTAGC           1520           GGCCTAGC           1590           GACCTAGCTAGC	870 870 870 870 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGAA 10:           SGAGGAG 11:           CACCGGA 11:           SGAGGGC 11:           SGAGGGC 11:           SGAGGGC 11:           SGAGGGC 12:           SGAGGGG 13:           ATGCCG 13:           SGGAGGG 14:           SGGAGGT 15:           SGGAGT 15:           SGGAGT 15:           SGGAGT 14:	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAC TGCTCGAC GGATTACC GGATTACC GGACGAGG GGACGAGG GGACGAGG GGACGAGG GGACGAGG GATCAAG GATCAAG GATCAAG GATCAAG GACGAGGT CTGCTAGC CTGCTAGC CTGCTAGC	820         820           PCAAGG4         PCAAGG4           PCAAGG4         920           PGAAGA1         PGAAGA1           PGAAGAACCA         PGAACAGACA           PGAAGA1         PGAACAGACA           PGACAGACA         PGAACAGACA           PGACAGACA         PGAACAGACA           PGCGAT         PGAAGACA           PGAAGAACA         PGAAGAACA           PGAAGAACAACA         PGAAGAACAACAACAACAACAACAACAACAACAACAACAAC	830 	840 840 AAGAGAGGAAGA AAGAGAGAGA 940 940 940 940 1010	850 850 AGGGTAGAAG 950 1020 100	860 860 TATAAATGG 960 960 960 1030 TCCAAATGA 1030 TCCAAATGA 100 CCACAGCATC 1030 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1170	870 870 870 877 877 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGAGGAA 10:           SGAGGGA 10:           LACCGGA 11:           CACCGGA 11:           CACCGGA 11:           CACCGGA 11:           CACCGGA 11:           CAGGGGC 11:           CAGGGGC 12:           CATAGCC 13:           CATAGCC 13:           CATAGCC 13:           CATAGCC 13:           CATAGCC 13:           CATAGCC 14:           SGGAGGG 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 14:           SGGGGGT 14:           SGGGGGT 14:           SGGGGGT 14:           SGGGGGT 14: <td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2	GAGGCCCC GAGGCCCC TGCTCGAC TGCTCGAC GGATTACC GGATTACC GGATTACC CTCTCAC GGACGAGG GGACGAGG GGACGAGG GGACGAGG GATCAAG GATCAAG GATCAAG GATCAAG GATCAAG GATCAAG GACGAGGT CTGCTAGC CTGCTAGC	820 1.1	830 1	840 840 840 840 840 840 840 840	850 850 AGGGTAGAAG 950 AAAGGCGTGCC AAAGGCGTGCC 1020 1020 CTCAGGTTCA CTCAGGTTCA 1090 1090 1090 1000 1	860           TATAAATGG           960           SACGRETTC           1030           TCCAAATGA           TCCAAATGA           100           CCACAGCAT           1100           CCACAGCAT           1100           CCACAGCAT           1100           CCACAGCAT           1240           GAAGTGCAG           GAAGTGCAG           GAAGTGCGAG           1310           GACGTCGGC           1380           AGGAAGGGTC           CTGATGCS           1520	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGGTGGT 10:           SGGGGG 11:           LacceGGA 11:           SGAGGTC 11:           SGAGGTC 11:           SGAGGGC 13:           LarGCCG 13:           LarGCCG 13:           SGGAGGG 13:           SGGAGGG 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 14:           SGGAGGT 14: <td>1 1 1 1 1 1 1 1 1 1 1 1 1 1</td>	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2 CnCos-2	GAGGCCCC GAGGCCCCCCCCCCCCCCCCCCCCCCCCCC	820         820           1         Image: Constant Section Sectin Section Sectin Section Sectin Sectin Section Sectin Section Sect	830 	840 840 AAGAGAGGAAGA AAGAGAGAGA 940 940 124TTGGGGATCC CATTGGGGATCC 1010 100	850 850 850 850 950 150 150 150 1020 10	860           TATAAATGG           TATAAATGG           960           GACGTCTTC.           1030           TCCAAATGA           TCCAAATGA           1100           CCACAGCAT           CCACAGCAT           CCCACGCAT           1100           CCACAGCAT           1170           TTCCACGGAT           1240           1240           1240           1310           GACGTCGCGC           AGGAGGTCC           AGGAGGTCC           CTTGATEGG           IS20           GTGCCTAGC           IS20           GTGCCTAGCTAGC           GTGCCTAGCTAGC	870 870 870 877 877 877 877 877	880           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SACCTAT 90:           SGAGGAA 97:           SGGAGGA 11:           SGAGGAGT 116:           SGAGGAGT 116:           SGAGGAGT 120:           SGAGGAG 13:           SGGAGGAG 13:           SGGAGGAG 14:           SGGAGAG 14:           SGGAGAG 14:           SGGAGGAG 14:           SGGAGGG 15:           SGGAGGG 15:           SGGAGGG 15:	1 1 1 1 1 1 1 1 1 1 1 1 1 1
CnCos-1 CnCos-2 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1 CnCos-2 CnCos-1	GAGGCCCC GAGGCCCCC TGCTCGAC TGCTCGAC CTCCTAC	820 10	830 830 AGGGAGGAAG, AGGGAGGAAG, 930 TTAAGCAAAA TTAAGCAAAA TTAAGCAAAA 1000 1 CAACAGCGAG, 1070  CCATGGTAT. CCATGGTAT. CCATGGTAT. CCAGGGTGGT 1140 1 CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGAAAAC CCGGGACAAAC CCGGGACAAC CCGGGACAAC CCGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGACACC 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGGCCA 1 CCGGGCCA 1 CCGGGCAAAC CCGGGCCA 1 CCGGGCAAAC CCGGGCCA 1 CCGGGGCCA 1 CCGGGCAAAC CCGGGCCA 1 CCGGGCCA 1 CCGGGCCA 1 CCGGCAAACC CCGGCAAACC CCGGCAAAC CCGGGCCA 1 CCGGGCCA 1 CCGGGCAAAC CCGGCAAACC CCGGCAAAC CCGGGCCA 1 CCGGCAAACC CCGGCAAAC CCGGCAAACC CCGGCAAACC CCGGCAAAC CCGGCAAC CCGCAAC CCGCAC CCGCAAC CCGC CCGCAAC CCGCCAC CCGCAC CCGCCCCCCCCCC	840 840 AAGAGAGGGAAGA AAGAGAGAGAA AAGAGAGAGA	850 850 850 850 850 850 850 850	860 860 TATAAATGG TATAAATGG 960 960 960 1030 TCCAAAGCATC 1030 TCCAAAGAATGA 1100 CCACAGCAT 1100 CCACAGCAT 1100 CCACAGCAT 1100 1170 1	870 870 877 877 877 877 877 877	880           SACCTAT 90:           SACCGAA 97:           SGAGGAA 97:           SGAGGAA 97:           SGAGGAA 10:           SACCGGA 11:           SACCGGA 11:           CACCGGA 11:           SACCGGA 11:           SCAGGGC 11:           SCAGGGC 11:           SCAGGGC 12:           SCAGGGG 13:           SCAGGGG 13:           SCAGGGG 13:           SGAGGG 14:           SGGAGA 14:           SGGAGGT 15:           SCATTGT 15:           SCATTGT 15:           SCATTGT 16:           SCATTGT 16:	1 1 1 1 1 1 1 1 1 1 1 1 1 1

#### 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  $\beta$ -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.

CnCos-1 CnCos-2 OS Glutelin A3 Q09151 E6 Glutelin Q95NZ2 E6 Glutelin Q94RA ZL Glutelin Q02945 AS 125 Globulin P14812	10 MASSSLLSSLS MASSSLLSSLS MATIKFPIVSSVV 	20 LLLCHLSQAO LLLCHVSQAO LLLCRGSLAO LLLCRGSLAO LLLCRGSLAO LLLCRGSLAO LFLLCNGSMAO	30 FGSSQESPI FGSSQESPI LLSQSTSQ FGSSQESPI FERSPI LLGQSTSQ LFGQSFTPI	40 225 PRRSVSS: 205 SRRSVST: 225 SRRSVST: 255 SRRSVST:	50 RNECRIERLM RDECQIERLM -RECREDRLQ RNECRIERLM 2500 GVEKHN -ROCREDRLQ -RGCREDRLQ	60 ALEPTRTVRS ALEPTRTVRS ALEPTRTVRS ALEPTRTVRS ALEPTREVRS AFEPURSVRS	70 EAG UTD YF DED ZAG ITD YF DED QAG TEFFDVS ZAG YTD YF DED AG YTD YF DED QAG TEFFDAS QAG ITEYFDED	80 NEQ 75 NEQ 75 NEL 75 NEQ 75 NEQ 75 NEL 77 NEL 77 NEQ 75
CnCos-1 CnCos-2 OS Glutelin A3 Q09151 EG Glutelin Q95NZ2 EG Glutelin Q95HA4 ZL Glutelin Q02945 AS 125 Globulin P14812	90 SRCAOVSTIRRVIEPRO PCAOVSAIRRVIEPRO CCICUPVURVIEPRO PCAOVSAIRRVIEPRO PCAOVSTURRIEPRO PCAOVSTURRIEPRO	100 LLLPSMS(AP) LLLPSMS(AP) LLLPSMS(AP) LLLPSMS(AP) LLLPQYT(GA' LLLPQYT(GA')	110 LUYIVQGR LVYIVQGR LVYIVQGR LVYIVQGR LVYIVQGR LVYIQGR FLVYILQGR	120 TVSLVHPGC ITSCVHPGC IVSLVHPGC IVSLVHPGC IVSTVIPGC ITSQTFPGC FTSLTFPGC	130 2 170 S - 7 08 2 170 S - 7 08 2 170 Q 7 0 0 2 170 S - 7 08 2 170 S - 7 08 2 170 S - 7 0 2 170 Q 7 00 2 170 Q 7 00 4 170 Q 00 7 0 1 4 170 Q 00 7 0 1 7 170 Q 00 0 0 0 1 7 170 Q 00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	140 ERESGE EQUERESGE EQUQULSGQ ERYERESGG ESEKQSEKG NQAQLTGSQ FDQAQFAE GQ	150 -RHRWSRDEHO -RHRRSKDEHO SQSHKFRDEHO -RHRRPRDEHO -QRQRFRDEHO SQSQNLKDEHO	160 KVY 150 KVY 153 KIH 155 KVY 153 RIH 153 RIH 153 KIN 157 RVH 155
CnCos-1 CnCos-2 OS Glutelin A3 Q09151 EG Glutelin Q95NZ2 EG Glutelin Q95HA4 ZL Glutelin Q02945 AS 125 Globulin P14812	170 CG CEOULDURING PAY CG CEOULDURING PAY RG CEOULDURING PAY G CEOULDURING PAY CG ESOULDURING PAY RG CEOUTAL BAR RG CEOUTAL BAR HIK COUVAL BAR	180 ICVINGENEVV ICVINGENEVV ICVINGENEVV ICVINGENEVV ICVINGENEVV ICVINGEVEVV ICVINGEVEVV	190 ITVLDTSN ITVLDTSN ITVLDTSN ITVLDTSN FTVSDIST IYVIDINN	200 MNQL DRS: ANOL DRS: ANOL DPR ANOL DRS: RISL MKTIGI ANOL DPR IANOL DPR	210 HROFLLAGRO HROFLLAGRO HROFLLAGRO NSYULGDGVA QROFLLAGNH OKEFLLAGNH	220 EQGRORYGRE EEGRORYRRE KIGGOLYRYE EEGRORYRRE AGGNLGKNKN RS-POLYRRE KRECOFG	230 -GSIKENILRG -ESIKENILRG ARDNSRNVFGG -ESMKDKILRG -RAQRATSLAD: VENQSQNIFSG QNIFSG	240 ST 227 GT 230 SV 233 ST 230 SIP 232 SA 234 SV 225
CnCos-1 CnCos-2 OS Glutelin A3 Q09151 EG Glutelin Q95NZ2 EG Glutelin Q95H44 ZL Glutelin Q02945 AS 125 Globulin P14812	250 ELLAAP CUNNELARA ELLEAD CIDHELARA ELLEAD CISSONARO SLLAAP CUNNELARA STWRPAL CUDEVARD ELLEAL CISTONARO CLISEAL CISCONACKI	260 CCRDDTRGET CCRDTTRGET CCRDTTRGET CCRDDTRGET CCRDDCRGET CCRDDCRGET CCRDDCRGET	270 TRADINGLOV TRADINGLOV TRADINGLOV TRADIGLOV TRADIGLOV TRADIGLOV TRADIGLOV	280 R0 SGM R0 SRR R0 SRR R0 SSE Q2 YAS R0 FVSQQGP	290 	300 	-QEQPRERYQ	320 1 269 272 QTQ 293 272 274 VTQ 295 STQ 305
CnCos-1 CnCos-2 OS Glutelin A3 Q09151 EG Glutelin Q9SNZ2 EG Glutelin Q9M4R4 ZL Glutelin Q02945 AS 12S Globulin P14812	330 EEEEEEECS 100 PET EEEERKS100 PET YQQKQLQSSSNGLDET KQEREERKOTINGE PT HQQSQYGGCSNGLDET YQPCQSWDQSFNGLED	340 TYC SHKIKONI TYC SHKIKONI TYC SHKIKONI TYC SHKIKONI TYC SHKIKONI TYC SLEARONI TYC SLEARONI	350 SDPRRADVF SDPRRADVF SNRNLADTY SDPRRADVF SSRRADVY SNRNLADTY SNRNLADTY	360 IPPGGRITTA IPPGGRITTA IPPGGRITTA IPPGGRITTA IPPGGRITTA IPPAGRUTTA	370 NSERLPILRF NGORFPILNL NSERLPILRF NSORLPILSF NSORFPILNL NSERNFPILNL	380 IQNSAERVVL IQNSAERVVL IQNSAERVVL IQNSAERVVL IQNSAERVVL IQNSATRVNL VQNSATRVNL	390 YRWANVSPHWN YRWANUSPHWN YRWANSPHWN YRWANLAPHWN YRWANLAPHWN YWALLSPFWN Y ONALLSPFWN Y ONALLSPFWN	400 INA 349 INA 349 INA 373 INA 352 INA 354 INA 354 INA 385
CnCos-1 CnCos-2 OS Glutelin A3 Q09151 EG Glutelin Q95NZ2 EG Glutelin Q95NZ2 EL Glutelin Q02945 AS 125 Globulin P14812	410 ISING COSTORNAL SING COSTORNAL SVIP IN COSTORNAL SVIP COSTORNAL SVIP COSTOR COSTON ISV VEVIOCARVAVIN ISV VEVIOCARVAVIN ISV NEUTOCOSTARVAVIN ISV NEUTOCOSTARVAVIN	420 DDRGETVFDGE DDRGETVFDGE NNGRTVFDGE NNGRTVFDGE NNGRTVFNGE NNGCTVFNDI	430 ROGOLLIV RRGOLLIV RRGOLLII RRGOLLVI RRGOLLII	440 ONFAMLERA ONYAVLERA HIVVIERA ONFAVIKOA OHYVVVERA	450 SSAFQUVSI SSEFQFVSI QRESCSVIAL GNESFEFTSI QRESCAVIAF ERESCQVISF	460 SISDRAHVST KISDRAHVSS KINPDSHVSH KIIDNAHVNT KINPNSHVSH KINPNSHVSQ	470 IV ON TO ALRON IV ON TO AFROM NA GONO IFRAL IV ON AS AFROM IV ON ON IFRAL IA GONTO ILRAL	480 PVE 429 PDE 429 PDD 453 368 PEE 434 PTD 455 PVD 465
CnCos-1 CnCos-2 OS Glutelin A3 Q09151 EG Glutelin Q05N82 EG Glutelin Q08484 ZL Glutelin Q0845 S. 125 Globmin P14812	490 ULMISTUISTOPPARTU ULMISTUISTOPPARTU UVANATSISSPERRIJ ULMISTEINSNEARTU ULMISTEINSNEARTU ULMISTEINSNEARTU	500 LTRODEVATE LTRODEVATE HNRODELGVE HNRODELGAE HNRODELGAE	510 RR RR SHAY RS RLQY RLQY	520 SRAFA SRAFA SRAFA	530 46 46 49 36 36 47 &S3 50	6 6 3 1 2		

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 38<sup>th</sup>-48<sup>th</sup> 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<sub>45</sub> and C<sub>77</sub>) near the N-terminal portion form an intrachain disulfide linkage within the acidic subunit. On the other hand, the other two (C<sub>121</sub> and C<sub>228</sub>) 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<sub>281</sub>-G<sub>282</sub> 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  $\beta$ -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.





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 (Mr), 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  $14^{th}$  and  $17^{th}$  WAA). Coffee 11S globulin transcription begins at past midmaturation 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 (<u>http://www.ncbi.nlm.nih.gov/</u>). 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.