

Cloning and Characterization of the Oleosin cDNA Isoforms from Coconut (*Cocos nucifera* L.)

Ellen S. Regalado, Jorge Gil C. Angeles, Abigail Joy D. Rodelas, Roberta N. Garcia*,
Antonio C. Laurena and Evelyn Mae Tecson-Mendoza

Institute of Plant Breeding, Crop Science Cluster, College of Agriculture, University of the Philippines Los Baños, College, Laguna 4031, Philippines

*Author for correspondence; e-mail: rngarciaipb@yahoo.com

Oleosins are proteins in the oil bodies of oil-bearing seeds. Two cDNAs of 500 and 300 bp, OLE500 and OLE300, were cloned by RT-PCR from coconut endosperm using a gene specific primer (FSPVLV) based on the conserved and characteristic proline knot of oleosins. Three distinct oleosins were shown by their cDNA sequences: OLE300a, OLE500a and OLE500c, with OLE500a being the major isoform. These isoforms were observed in solid endosperm of 6–7 and 9–10 month old samples but not in the mature 11–12 month sample. Southern blot analysis showed two copies of OLE 500a gene. The partial cDNA sequences of OLE300a, OLE 500a and OLE500c had coding regions of 176, 242, and 233 bp, which code for 58, 80 and 77 amino acids, corresponding to molecular weights of 6040, 8600 and 8566, respectively. Analysis of the amino acid sequences revealed the highly conserved hydrophobic domain from the right half of the proline knot, the C-terminal amphipathic region characteristic of oleosins and that the coconut oleosins are of the L-form due to the absence of the 18-amino acid insert characteristic of the H-form. OLE 500a had 77.0% similarity with OLE500c and 85.96% with OLE 300a and the latter had 91.23% similarity with OLE 500c. OLE 500a and OLE 500c exhibited highest similarity with oil palm oleosin (86%) and OLE300a with soybean oleosin (72%), respectively.

Key Words: oleosin, coconut, *Cocos nucifera*, isoforms, cDNA sequence, amino acid sequence

INTRODUCTION

Oleosins are small and largely hydrophobic proteins of 15–26 kDa size found in the oil bodies of different plant species (Huang 1992). They abound in the lipid-storing bodies of seeds comprising up to 2–8% of the total seed proteins and have also been found in other parts of plants such as the tapetum cells of anthers (de Oliveira et al. 1993; Roberts et al. 1994; Hsieh and Huang 2004) in the Brassicaceae family, such as *Brassica* and *Arabidopsis*.

Oleosins are composed of three regions: an N-terminal hydrophilic region of variable length (from 6 to 60 residues), a central highly conserved hydrophobic domain of 72 residues and a C-terminal amphipathic region of variable length, from 28 to as long as 1000 residues in *Arabidopsis* oleosins (Hsieh and Huang 2004; Huang 1996). The central hydrophobic domain consists of 72 hydrophobic amino acid residues which is a distinctive trait of oleosins. This 72-hydrophobic stretch forms a hairpin with two arms of 30 amino acids joined by a head called the

proline knot because of the three Pro at identical locations (Hsieh and Huang 2004).

In spite of their abundance in seeds, oleosins are encoded by only two to a few genes per haploid genome (Huang 1996) in contrast to large gene families of seed storage proteins. This suggests the high expression of oleosins especially during the embryogenic and immature stages of seeds. Tai et al. (2002) reported single copy or low copy number of the three oleosin genes in sesame genome.

Tzen et al. (1992) have identified two immunologically distinct oleosins classified as high and low- M_r isoforms, the H- and L-oleosins, which always differ in molecular weight by 2 kDa. This difference represents the insertion of 18 residues in the C-terminal part of the H-oleosins (Tai et al. 2002). Thus, maize, barley and rice have 18- and 16-kDa oleosins (Qu and Huang 1990; Aalen 1995; Chuang et al. 1996) and sesame has 17- and 15- kDa oleosins (Tai et al. 2002). However, the oleosin of gymnosperms has only the L form (Wu et al. 1999).