Expression of Aintegumenta-like Gene Related to Embryogenic Competence in Coconut Confirmed by 454-pyrosequencing Transcriptome Analysis

H D D Bandupriya*

Abstract

A member of the Aintegumenta sub-family of Apetala gene family encoding two APETALA2 (AP2) domains was isolated and termed as *Cocos nucifera* Aintegumenta like gene (CnANT). The deduced amino acid sequence of the conserved domains shared a high similarity with Aintegumenta-Like (ANT like) genes in *Arabidopsis thaliana*, *Elaeis guineensis*, *Oryza sativa*. Comparison of transcriptomes in different tissues revealed that CnANT transcripts were high in mature zygotic embryo (12 months after pollination; 12ME). Quantitative RT-PCR results confirmed the higher CnANT transcript accumulation in mature zygotic embryos while transcripts were rarely detected in vegetative tissues such as leaf. The expression data and global transcriptome data were therefore consistent across the embryo maturity stage and showed that CnANT could play a role in embryogenesis.

Keywords: coconut, zygotic embryo, Aintegumenta, transcriptome.

Introduction

Coconut (*Cocos nucifera* L.) is a perennial cross-pollinated plant that is cultivated mainly for edible oil in the tropical and sub-tropical regions. Despite its economic importance, coconut production has not been increased as we see in other oil crops. Production of high quality superior palms in large scale is a prerequisite for the development of coconut industry. Achieving this target using seeds as the planting material is impossible. Its propagation by tissue culture was first reported in 1983 (Branton and Blake 1983). Since then cloning of coconut via somatic embryogenesis has been addressed by several researchers in different research groups (Fernando and Gamage 2000; Hornung 1995; Karunaratne and Periyapperuma 1989; Perera *et al*. 2008; Perera *et al*. 2007; Verdeil *et al*. 1994). However, its highly recalcitrant behavior for *in vitro* conditions limits the success of coconut micropropagation (George and Sherrington 1984).Therefore, understanding of the molecular basis of coconut tissue culture would provide necessary information needed for the improvement of the *in vitro* propagation protocol.

Members of the sub family Aintegumenta-Like (AIL) of the Apetala2/Ethylene-responsive element binding protein (AP2/EREBP) family play an important role during the transition from vegetative to embryogenic growth (Banno *et al*. 2001; Boutilier *et al*. 2002). BABY BOOM (BBM) is one of such genes known for its role in cell proliferation and morphogenesis during embryogenesis (Boutilier *et al.* 2002). These genes are expressed in dividing tissues where they have central roles in

_____

*Senior Research Officer, Tissue Culture Division, Coconut Research Institute, Lunuwila, Sri Lanka.

Correspondence author: Email: dbandupriya@yahoo.com

Tel: +94312262003;   Fax: +94312257391

developmental processes such as embryogenesis. Overexpression of the AIL genes induces somatic embryogenesis and ectopic organ formation (Boutilier *et al.* 2002; Tsuwamoto *et al*. 2010). The characterization and functional analysis of markers such as AIL for somatic embryogenesis offer the possibility of determining the embryogenic potential of coconut in culture long before any morphological changes have taken place. Since zygotic embryo development always mimic the somatic embryogenesis, most studies on gene isolation have been carried out initially using zygotic embryo tissues (Cairney and Pullman 2007; Palovaara *et al.* 2010) .

Next generation sequencing (NGS) technologies enable scientists to analyze the complete transcriptome at a minimal cost. Roche 454 Genome Sequencer (GS) FLX platform is widely used for de novo sequencing and EST analyses in non-model plants (Barakat *et al.* 2009; Graham *et al.* 2010; Li *et al.* 2010; Luo *et al.* 2010; Wang *et al.* 2009; Novaes *et al.* 2008; Alagna *et al.* 2009) and is an ideal way to discover genes and markers, quantify transcripts and discover small RNA (Morozova *et al.* 2009; Brautigam and Gowik 2010). In the present study Aintegumeta like gene was identified in coconut zygotic embryos and expression was checked in four different plant tissues. The occurrence of this Aintegumenta like gene was further identified in a separate transcriptome analysis during AP2/EREBP family gene mining.

Materials and methods

Plant Material

Seeds were from variety 'Sri Lanka Tall' and were harvested from bunches at 12-month maturity stage (given that 0 is the most mature unopened inflorescence embryos were used to extract RNA for gene isolation. Four different tissue types from the variety 'Sri Lanka Tall' namely, immature embryo at the age of nine months after pollination (9ME), mature embryo at the age of 12 months after pollination (12ME), a microspore derived embryo (MDE) (Perera *et al.* 2008) and developing leaves from an 8-month old *in vitro* germinated coconut plantlet

were used for the cDNA library construction for the 454 sequence analysis.

RNA Extraction

Total RNA from each tissue sample was extracted using the RNeasy® Plant Kit (Qiagen) according to the manufacturer's instructions. DNA contaminations were eliminated by treating with DNase I (Qiagen, UK) according to the manufacturer's instructions. The amount of RNA was quantified using a NanoDrop® ND-1000 Spectrophotometer and integrity was analyzed by 1% agarose gel electrophoresis.

Cloning of coconut ANT-like cDNA

Primers [2 F (5-TCT ATC TAC CGC GGC GTC- 3′) and 4R (5-ACA AAC TCC TGT CGT GTC A-3′), 4 F (5-TGA CAC GAC AGG AGT TTG T-3′) and 5R2 (5-ATT CCA TTC CAA AGA TGG G-3′)] were designed to amplify the coconut ANT-like cDNA sequence on the basis of the most conserved nucleotide sequences of rice (Accession NM 001060643) and oil palm (Accession AY691196) Aintegumenta-like genes. Approximately 1 μg of total RNA was used to synthesize first strand cDNA using SensiMix 2 step kit (Quantace) according to the manufacturer's instructions. PCR was performed using 1 μL of cDNA in the presence of primers at 0.3 μM concentration. Each PCR reaction was carried out in a 25 μL reaction volume containing 12.5 μL 2X Biomix (Bioline). After an initial 2 min denaturation step at 94°C, 30 cycles were run, each with 30 s of denaturation at 94°C , followed by 30 s of annealing at 57°C and 45s of extension at 72°C. The final elongation cycle was at 72°C for 7 min. Purified PCR products were subjected to BigDye terminator cycle sequencing reaction and sequenced at the Bio-Centre, University of Reading. From these amplifications 1031 bp length fragment was obtained. Flanking sequence determination of 5' and 3' ends was carried out by the RACE method using the GeneRacer kit (Invitrogen) using the primers provided by the kit (GeneRacer 5 primer 5-CGA CTG GAG CAC GAG GAC ACT GA-3, GeneRacer 3′primer 5'-GCT GTC AAC GAT ACG CTA CGT AAC G-3, ′5′ nested primer 5′

GGA CAC TGA CAT GGA CTG AAG TAG AAA-3', 3′nested primer 5′-CGC TAC GTA ACG GCA TGA CAG TG-3′and gene specific primers 4R, 4F, 3R (5'-CGC CTT CTC CTC CTT ATC-3'), ANTF (5′-AAC TGG ATT ATG CAT GAT GA-3)′according to the manufacturer's protocol.

Real-time RT-PCR expression analysis

Two gene-specific primers; ANTF1: (5′-CGG TCT CTT CTC CTC TGG TG-3′) and ANTR: (5′-TCG TAA TTC CCT CCA AAT GC-3′) were designed based on the coding region of the CnANT gene to amplify a 180 bp region. The coconut elongation factor gene was used as the internal control gene (Morcillo *et al*. 2007; Olsvik *et al*. 2005). cDNA from the four tissue types 9ME, 12ME, MDE and LEAF were used for real-time RT-PCR analysis using SYBR premix Ex taqTM (Takara). Experiments were conducted with two biological samples, and the real-time qPCR reactions were performed in triplicate using the CAS-1200 liquid handling system, version 4.7.979 (Corbett Robotics). The real-time RT-PCR was performed on a Rotor-Gene 6000 real-time cycler (software 1.7, Corbett Research). The amplification parameters were one cycle at 95 °C for 1 min, 39 cycles of 95°C for 10s, 60°C for 20s, and 72°C for 8s. The relative expression level in different tissues was calculated by the standard curve method.

cDNA synthesis and 454 pyrosequencing

The first-strand cDNA was produced from 0.3µlg of total RNA. A modified SMART-Sfi1A oligonucleotide (5'- AAG CAG TGG TAT CAA CGC AGA GTG GCC ATT ACG GCCrGrGrG-3') was used in combination with the CDS-Sfi1B primer (5'- AAG CAG TGG TAT CAA CGC AGA GTG GCC GAG GCG GCC d (T) 20–3') to synthesize the first strand cDNA in the presence of PowerScript Reverse Trascriptase (BD Biosciences Clontech,UK). For double-stranded cDNA (ds cDNA) synthesis, the cDNA was diluted and amplified using PCR Advantage II polymerase (BD Biosciences Clontech,UK) in the presence of SMART PCR primer (5'-AAG CAG TGG TAT CAA CGC AGA GT- 3'). The following the thermal profile: 1 min at 95ºC followed by 25 cycles of 95ºC for 7 s, 65ºC for

20 s, and 72ºC for 3 min was used for the amplification. Five micro liters of PCR product was electrophoresed in a 1% agarose gel to determine the amplification efficiency. The amplified cDNA PCR product was purified using QIAquick PCR Purification Kit (QIAGEN, CA), concentrated by ethanol precipitation and adjusted to a final concentration of 50 ngµ^{-1}. A total yield of 3 µg of cDNA was prepared for each tissue type by conducting several long distance PCR reactions. DNA sequencing of four libraries was performed at Centre for Genomic Research, University of Liverpool using a 454-GS FLX Genome Sequencer and the sequence data processing was performed with the GS FLX software v2.0.01. Sequences in each library were subjected to a BLAST search against the non-redundant protein database using BLASTX with an e value cut-off of 1E-6. AP2 family proteins were identified in each library based on the BLASTX search.

<div align="center">Results and discussion</div>

Cloning of full length cDNA homologue of Aintegumenta-like gene

A partial sequence (1029 bp) of Aintegumenta-like homologue gene was obtained by PCR using primer pairs 2F, 4R and 4F, 5R2. Amplification products of 5'-RACE and 3'-RACE generated the full length cDNA sequence that was 1782 bp in length. We named it as *Cocos nucifera* L. Aintegumenta (CnANT) gene. This contained a 1425 bp open reading frame (ORF), 62 nucleotides at the 5'-untranslated region (UTR) and 305 nucleotides at the 3' UTR including 26 adenines from the polyA tail. The ORF encoded a putative peptide of a 474 amino acids (Figure 1).

Sequence database searches revealed that the deduced polypeptide sequence of CnANT shows similarity to AP2/EREBP family proteins. This putative CnANT protein sequence contains two AP2 domains from protein residues 128 to 204 in repeat one and residues 230 to 298 in repeat two and a conserved linker region from residues 205 to 229 (Figure 1). Within the CnANT polypeptide sequence the highly conserved YRG elements encoded by amino acid residues from 128 to 149 and 230 to 251 were

```
                                                                          AA
AGCGTTCAAATCAAAACCTTCTTTCTCAACTTGTTATTCCGGGGATACCCTGGCAGCGCC
atggacatggacgcttcacacagctggctcgccttctcgctatccaaccaccagccctac
  M  D  M  D  A  S  H  S  W  L  A  F  S  L  S  N  H  Q  P  Y     20
cttctcgaggctctgtcctccgcccatccacatggtggggtggagatgacggcggaggag
  L  L  E  A  L  S  S  A  H  P  H  G  G  V  E  M  T  A  E  E     40
cggggaggatcggcggaagtggcggcggctggcggtgatgggccgaagctggagacttc
  R  G  G  S  A  E  V  A  A  L  A  V  M  G  P  K  L  E  D  F     60
ctcggcgggtgtgggagccgagggggcgatactctggcggcgaggccggggatgcagcg
  L  G  G  C  G  E  P  R  G  R  Y  S  G  G  E  A  G  D  A  A     80
gggatctacgactctgaactgaagcacatcgccgccgggtacctgcaggggctaccgccg
  G  I  Y  D  S  E  L  K  H  I  A  A  G  Y  L  Q  G  L  P  P    100
gcggagcatcaggactccgagatggcgaaggtggtggcgccggcggagtcacggaaggcc
  A  E  H  Q  D  S  E  M  A  K  V  V  A  P  A  E  S  R  K  A    120
gtcgaaaccttcggccagcgcacatctatctaccgcggcgttcaccaggcatcgatggacg
  V  E  T  F  G  Q  R  T  S  I  Y  R  G  V  T  R  H  R  W  T    140
ggtcgatacgaggcgcacctctgggacaacagctgccgccgggagggtcagagccggaag
  G  R  Y  E  A  H  L  W  D  N  S  C  R  R  E  G  Q  S  R  K    160
ggaaggcaagtctatcttggtggatacgataaggaggagaaggcggcgagggcttacgat
  G  R  Q  V  Y  L  G  G  Y  D  K  E  E  K  A  A  R  A  Y  D    180
ctcgcagcgctcaagtattggggtccgaccaccaccactaactttccgatcttcaactac
  L  A  A  L  K  Y  W  G  P  T  T  T  T  N  F  P  I  F  N  Y    200
gagaaggagctggaggagatgaagaacatgacacgacaggagtttgtcgcatcgctcaga
  E  K  E  L  E  E  M  K  N  M  T  R  Q  E  F  V  A  S  L  R    220
aggaagagctctgggttctccagagggcgcatcgatctataggggagtcaccagacaccat
  R  K  S  S  G  F  S  R  G  A  S  I  Y  R  G  V  T  R  H  H    240
caacatggccggtggcaagctagaattggaaggggtagcaggcaacaaagatctttatctt
  Q  H  G  R  W  Q  A  R  I  G  R  V  A  G  N  K  D  L  Y  I    260
ggaacctttagcacgcaggaagaggcagcggaggcctatgacattgcagcaattaagttt
  G  T  F  S  T  Q  E  E  A  A  E  A  Y  D  I  A  A  I  K  F    280
cggggggctcaatgcagtgacaaattttgacattagccgctacgatgtcaagagcatcgcc
  R  G  L  N  A  V  T  N  F  D  I  S  R  Y  D  V  K  S  I  A    300
aacagcaaccttcccattgggggatgacaggccggccctccaaggccacagaatcttcg
  N  S  N  L  P  I  G  G  M  T  G  R  P  S  K  A  T  E  S  S    320
ccatcatcctcatctgacgccatgactgtggaagccaagcagctgttggacggccgggat
  P  S  S  S  D  A  M  T  V  E  A  K  Q  L  L  D  G  R  D    340
ctctcggcctcacttgggtttgctgcactccccattaagcacgatcaggacttctggtcc
  L  S  A  S  L  G  F  A  A  L  P  I  K  H  D  Q  D  F  W  S    360
ttgtttgccctccaacagcagcagcagcaaagtaatcaagcttcaggcttcggtctcttc
  L  F  A  L  Q  Q  Q  Q  Q  Q  S  N  Q  A  S  G  F  G  L  F    380
tcctctggtgtgactgtggacttcttctccacagcttctaatggtgttataagccaaggtgtt
  S  S  G  V  T  V  D  F  S  T  A  S  N  G  V  I  S  Q  G  C    400
ggggggagccttgtgtggaacggtggtgttggagaacagcagcaggagcaatcactgaac
  G  G  S  L  V  W  N  G  G  V  G  E  Q  Q  Q  E  Q  S  L  N    420
aatagttgctcttctatcccatatgcaacacctattgcatttggaggggaattacgagggc
  N  S  C  S  S  I  P  Y  A  T  P  I  A  F  G  G  N  Y  E  G    440
tccagctatgtggggagctgggtcacaccaccaccctcttactaccacgaagctgccaag
  S  S  Y  V  G  S  W  V  T  P  P  P  S  Y  Y  H  E  A  A  K    460
cccaatgtggcattcttcagacacccatcttggaatggaatgaTATGGATGGGTGGGA
  P  N  V  A  F  F  Q  T  P  I  F  G  M  E  -
AGGTGAGAGGAGATCACACTTGCTTGAAGAGAAAGAGAGTCAAAAGTGAGAGAGACCATG
GGAGGTGATGCCAACTGACCACACTCACAAACATGAAGAAGGGAAGGAGAACAGAGGGAG
GACTGACAAGAAGGCTAACAAGATATCAAAAGTTTTGCTCATGGTCATGCGCCCAACTTC
TTTAACTTTTTGGTTTTTCTTTTCATGACATGATTCTATATTGTCATCATCATCATCAGTG
TAATGGAAACAAGAACCCTTTTGTAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 1. cDNA sequence of coconut AINTEGUMENTA like homologue gene *(CnANT)*. The nucleotide sequence and the putative protein sequences are shown. Two conserved AP2 domains are underlined. 5' and 3' UTRs are shown along with cDNA sequence in upper case letters. Numbers given in the right hand side are only for the putative amino acid sequence. Amino acid residues that encode the highly conserved YRG elements (128 to 149 and 230 to 251 respectively) in AP2 repeat 1 domain and repeat 2 domain are highlighted in light grey and the two RAYD element encoding amino acid residues (161 to 204 in repeat 1 and residues from 255 to 298 in repeat 2) are highlighted in dark grey. The specific central core of 18 amino acids within the RAYD element in each AP2 domain is illustrated in italic letters.
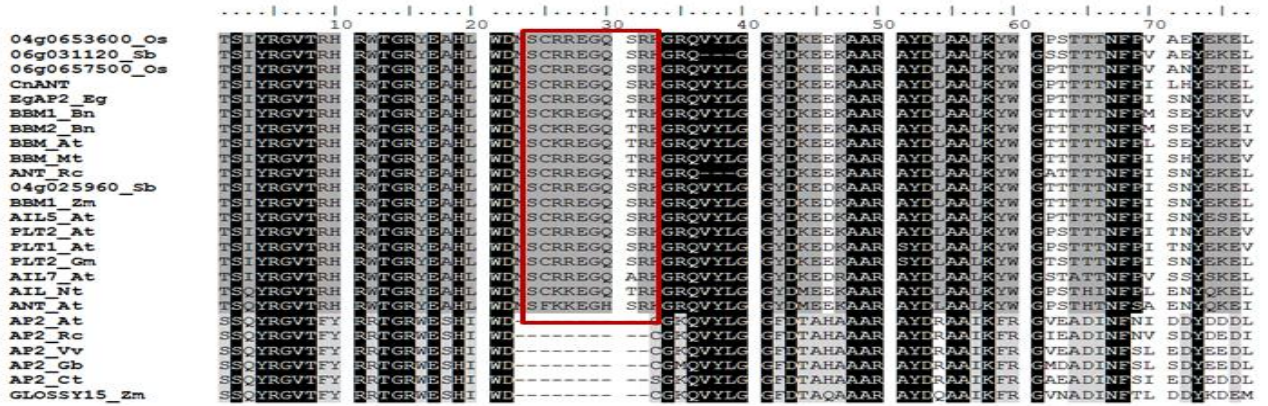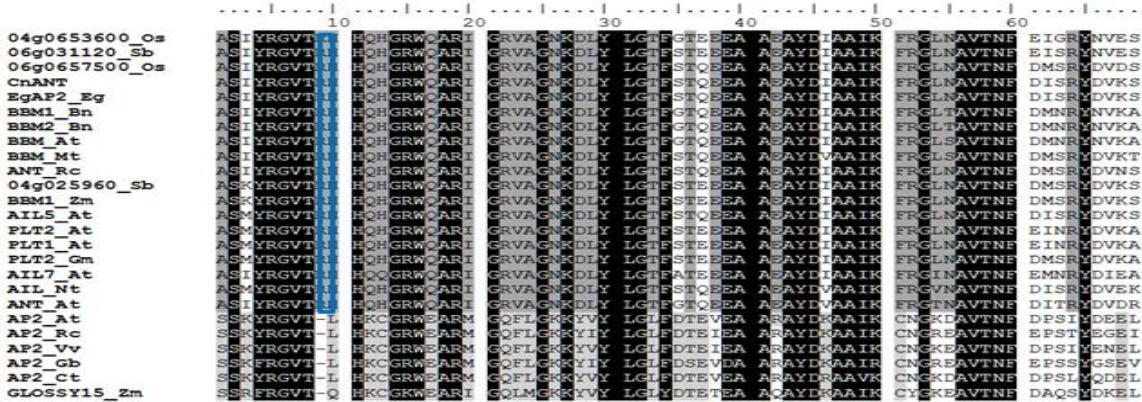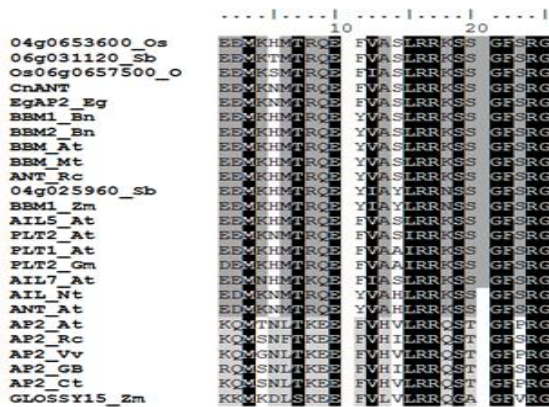
Figure 2. Comparison of CnANT protein sequence and related AP2 domain containing proteins. The amino acid sequences of two AP2 domains and the linker region connecting the two repeats of CnANT were aligned with the related AP2/EREB proteins. Black boxes indicate the amino acids that are identical in all members of the AP2 subfamily. Dark grey boxes shows the amino acids that are identical in all AIL sub group proteins. Amino acids that are identical in AP2 sub group are coloured in light grey. Note the 10 amino acid insertion in the AP2 repeat 1 (indicated by a red box) and one amino acid insertion in the AP2 repeat 2 (indicated by a blue box) which distinguish ANT genes from the rest of AP2 genes.

observed for AP2 repeat 1 domain and repeat 2 domain, respectively. Two RAYD elements were identified from residues 161 to 204 in repeat 1 and residues from 255 to 298 in repeat 2. Furthermore, the specific central core of 18 amino acids within the RAYD element which has been predicted as forming an alpha helix were also detected (residues 169-186 in repeat 1 and 263-280 in repeat 2; Figure 1) (Okamuro *et al.* 1997). The 10 amino acid insertion in the AP2 repeat 1 and one amino acid insertion in the AP2 repeat 2 which distinguish ANT genes from the rest of AP2 genes were identified in the CnANT sequence (Figure 2). This clearly showed that *CnANT* belongs to the ANT sub group (Nole-Wilson *et al.* 2005).

Furthermore, four conserved motifs identified (Kim *et al.* 2006) in the pre-domain region (EuANT 1, EuANT 2, EuANT 3 and EuANT 4) which are specified for the ANT sub group proteins were identified in the CnANT putative protein. Two AP2 domains and the linker region of the CnANT protein shows the highest similarity with EgAP2-1 (oil palm), ANT-like (rice), BBM, AIL5, AIL7, PLT1 and PLT2 (*Arabidopsis*), BBM (*Brassica*) and a number of recently identified hypothetical proteins available publicly. Amongst these genes, BBM (Boutilier *et al.* 2002), AIL5 (Tsuwamoto *et al.* 2010) *Brassica* BBM1 and BBM2 (Boutilier *et al.* 2002; Srinivasan *et al.* 2007), EgAP2-1 (Morcillo *et al.* 2007) have been characterized as embryogenesis related genes. Within the two AP2 repeat domains, these sequences share ⸴ 28% identity when consider both AP2 and ANT sub groups. However, the sequence identity is greatly increased (>75%) when only ANT group proteins are considered. When examined carefully it was noted that pairs of genes share similarity within the entire AP2 domain sequences. Coconut and oil palm ANT protein sequences show 98% identity with each other in the conserved domain regions while CnANT shares more than 80% identity in the same region with other ANT subfamily proteins when a pair wise comparison was conducted (Figure 2). This observation emphasized that CnANT is strictly related with oil palm EgAP2-1 gene (Morcillo *et al.* 2007) and indicates its

close relationship with the palm species sequence. In a recent study, Ouakfaoui *et al.* (2010) studied the conserved motifs of AP2 sub family outside of AP2 DNA binding domains and found ten sequence motifs of euANT lineage. Three of them identified at N terminal have been described by Kim et al (2006) previously. The euANT sub group was further categorized into BBM-like, PLT-like and AIL5-like and oil palm EgAP2-1 was grouped as an AIL5-like gene (Ouakfaoui *et al.* 2010). Thus, similarity between oil palm EgAP2-1 and CnANT suggests CnANT to be grouped as AIL genes according to Ouakfaoui and colleagues classification and hence suggests that CnANT could play a role in embryogenesis as proposed for those orthologs.

The four cDNA libraries were sequenced using a GS FLX sequencer (454/Roche), resulting in a total of 979428 reads. These reads were assembled into 32621 putative unigenes and 155017 singletons. ESTs had an average length of 460 bp, and represented 223.7 Mb. Assembled sequences were functionally annotated with blast2go. Homology-based functional assignment for putative sequences was accomplished through BLASTX queries against non-redundant protein database with an e value cut-off of 1E-6 to identify AP2 family proteins in each library.

The highest number of APETALA family genes was found in the 9ME library which was represented by 16 contigs. The other three libraries 12ME, MDE and LEAF had 11, 12 and 10 contigs respectively which were encoded for AP2 family genes. As per the putative annotations made from BLASTX hits, most of them were of the ERF subfamily which has a single AP2 domain. Also there were contigs encoded for RAV subfamily genes (genes containing one AP2 domain and a second B3 domain) and AP2/ANT subfamily genes. Interestingly, CnANT gene which was described early in this paper could be identified in all embryo tissue libraries (Table 1; marked with an asterisk). These contigs returned the highest significant similarity [Evalue ranged from 1.04E-33 (9ME & MDE) to 1.66E -166 (12ME)] with the oil palm Aintegumenta-like gene;

Table 1. Identification of AP2 family genes and *CnANT* in LEAF, 9ME, 12ME and MDE Library ESTs

Contig name	No of ESTs	Putative identity	*Species*	GI number	E value
9ME					
contig00041	6	AP2/ERF domain containing protein	*Oryza sativa*	25548125	3.22E-16
contig00367	11	AP2 domain containing protein	*Vitis vinifera*	297743010	7.17E-15
contig01025	33	AP2/ERF domain containing protein	*Populus trichocarpa*	224075453	7.66E-28
contig01586	5	AP2 domain containing protein	*Oryza sativa*	25603736	3.00E-09
contig01721	8	AP2/ERF transcription factor	*Lycoris logituba*	295913651	2.36E-09
contig02544	3	RAV	*Nicotiana tabacum*	194475604	2.45E-46
contig02679	16	AP2/ERF factor	*Actinidia deliciosa*	300079016	1.95E-21
contig03137	4	AP2/ERF FACTOR	*Populus trichocarpa*	224128256	1.00E-07
contig03897	8	AP2/ERF FACTOR GENE	*Lycoris logituba*	295913454	1.79E-25
contig04205	2	AINTEGUMENTA Protein	*Ricinus communis*	255547490	2.22E-44
contig04757	2	ERF3 Transcription factor	*Vitis vinifera*	22544716	3.57E-11
contig05190	63	ERF transcription factor 3	*Vitis pseudoreticulata*	289466347	1.44E-60
contig05434	19	AP2 protein*	*Elaeis guineensis*	56567285	1.05E-33
contig05623	19	APETALA2	*Hyacinthus orientalis*	5360996	2.53E-35
contig05624	33	AP2/ERF factor gene	*Populus trichocarpa*	295913424	6.92E-21
contig05812	9	APETALA2-LIKE	*Pisum sativum*	13173164	1.87E-30
12ME					
contig00041	211	AP2 protein*	*Elaeis guineensis*	56567285	1.66E-166
contig00486	25	ERF factor 2	*Prunus salicina*	222427675	2.91E-68
contig00853	3	EREB-like protein	*Oryza sativa*	25559179	1.01E-12
contig01069	31	ERF transcription factor 3	*Vitis pseudoreticulata*	289466347	5.10E-62
contig02415	7	AP2/ERF transcription factor	*Lycoris longituba*	295913651	4.90E-24
contig03252	6	AP2/ERF domain containing protein	*Populus trichocarpa*	295913424	6.29E-10
contig03842		ethylene responsive protein	*Zea mays*	226497004	8.57E-12
contig04354	20	Aintegumenta like 1	*Vitis vinifera*	47832538	1.09E-41
contig04480	5	APETALA2-LIKE	*Pisum sativum*	13173164	2.11E-34
contig04711	14	AP2/ERF factor	*Oryza sativa*	25548125	1.18E-17
contig04898	45	AP2/ERF FACTOR	*Populus trichocarpa*	224075453	7.81E-28
MDE					
contig00907	5	RAV	*Nicotiana tabacum*	194475604	1.30E-71
contig01217	7	APETALA2	*Hyacinthus orientalis*	5360996	1.47E-35
contig01626	28	ERF transcription factor 3	*Vitis pseudoreticulata*	289466347	1.55E-33
contig02007	34	AP2 protein*	*Elaeis guineensis*	56567285	1.04E-33
contig02444	14	AP2/EREBP transcription factor	*Vitis vinifera*	297742325	7.15E-17
contig03285	8	APETALA2 protein	*Oryza sativa*	222629673	2.82E-27

contig03381	7	AP2/ERF factor	*Oryza sativa*	115463423	1.23E-23
contig05046	4	AP2/ERF factor	*Actinidia deliciosa*	300079018	9.61E-11
contig07109	4	AP2/ERF transcription factor	*Lycoris logituba*	295913424	1.14E-19
contig08165	18	AP2 domain DRE binding protein	*Cymbidium insigne*	49980678	4.29E-11
contig08519	2	AP2/ERF transcription factor	*Lycoris logituba*	295913424	4.96E-15
contig09160	27	APETALA protein	*Vitis vinifera*	297746184	4.98E-39
LEAF					
contig00169	4	APETALA2	*Hyacinthus orientalis*	5360996	6.03E-27
contig04147	17	AP2 domain containing transcription factor	*Arabidopsis thaliana*	15221435	1.62E-09
contig04233	4	AP2/ERF factor 10	*Actinidia deliciosa*	300079016	5.34E-31
contig08025	3	APETALA2	*Hyacinthus orientalis*	5360996	3.78E-21
contig08521	3	Aintegumenta protein	*Triticum aestivum*	82780772	7.41E-25
contig08923	4	AP2/EREBP transcription factor	*Vitis vinifera*	225463950	1.36E-10
contig09150	15	Aintegumenta protein	*Ricinus communis*	255578779	3.35E-40
contig09168	33	AP2 domain containing protein	*Vitis vinifera*	225435173	1.03E-45
contig09396	33	APETALA2	*Hyacinthus orientalis*	5360996	7.56E-38
contig11376	29	ERF factor 3	*Nicotiana tabacum*	30961941	2.07E-12

*similar contigs for *CnANT* gene showing BLASTX sequence similarity with oil palm AINTEGUMENTA-like gene (*EgAP2-1*)
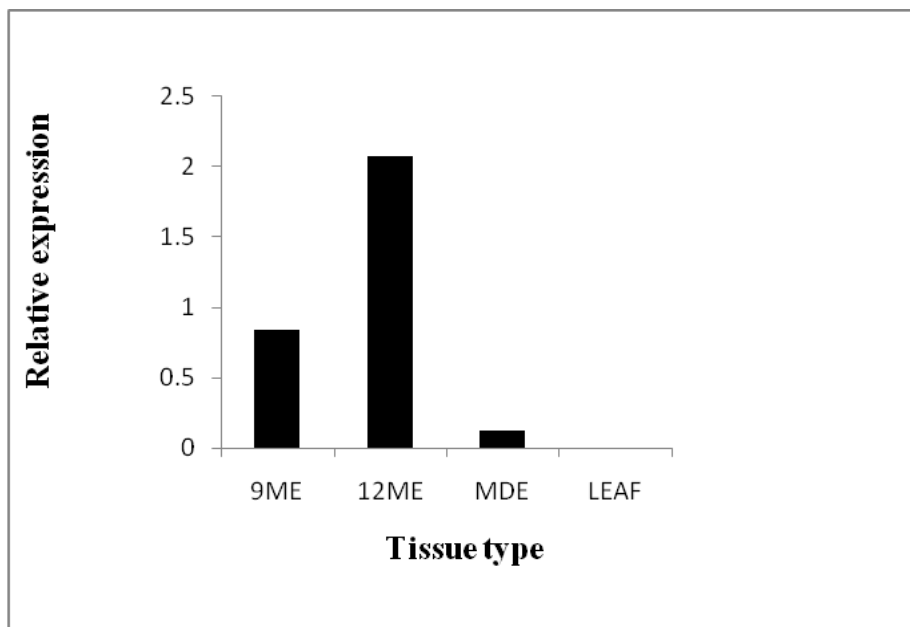

Figure 3. Relative expression levels of CnANT accumulation in the four different tissue types determined by RT-qPCR analysis. CnANT transcripts were compared relative to the elongation factor gene. LEAF: leaf, 9ME: embryo, 9 month after pollination, 12ME: embryo, 12 month after pollination MDE: microspore derived embryo

EgAP2-1 (Morcillo *et al.* 2007) when subject to a BLASTX homology search. The EST abundance of this gene was high in the 12ME library. Even though the abundance was not as high in the 12ME library, a considerable number of ESTs was encoded by the contigs in 9ME and MDE. In 9ME, the contig was present 19 times while in MDE it appeared 34 times. However no transcripts encoding the CnANT gene were found in LEAF library. These observations are comparable with the results obtained from the quantitative real time PCR where the highest relative expression was observed in 12ME while LEAF tissues showed negligible relative expression (Figure 3). Furthermore, the comparison of putative gene annotations, species and GI values of the contigs in different libraries revealed that a few more genes from the AP2 family are common to embryo tissue libraries (Table 1).

The present study provides evidence that CnANT is strongly induced in late embryogenesis and plays a role during zygotic embryo maturation phase. First, a qPCR experiment showed higher relative expression at the most mature stage of zygotic embryo (12 months after pollination). Secondly, global transcriptome analysis carried out using different stages of embryo maturity showed a higher accumulation of CnANT transcripts in the 12ME library compared to the 9ME library and the LEAF library based on the BLASTX analysis (sequence cut-off $1.0e^{-6}$). The expression data and global transcriptome data are therefore consistent across the embryo maturity stage. Higher expression of AIL5 and BBM genes in later stages of embryo development has been reported previously (Boutilier *et al.* 2002; Morcillo *et al.* 2007; Tsuwamoto *et al.* 2010). Unlike in BBM which was expressed in stages as early as the globular stage, AIL gene expression has been analysed after laser capture microdissection and *in situ* hybridization and shown to occur only in embryos older than the heart stage (Casson *et al.* 2005; Tsuwamoto *et al.* 2010).

Based on the results, presence of AIL gene in coconut is reported and findings supports the fact that this gene is highly expressed in maturing stages of embryo. CnANT could be used as embryogenesis marker, since its expression changed with progression of embryo development. These findings may offer a valuable contribution to the evaluation of embryogenic culture responses in coconut.

References

Alagna F., D'Agostino N., Torchia L., Servili M., Rao R., Pietrella M., Giuliano G., Chiusano M.L., Baldoni L. and Perrotta G (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* **10**: 399

Banno H., Ikeda Y., Niu Q-W. and Chua N-H. (2001) Overespression of *Arabidopsis ESR1* induces initiation of shoot regeneration. *Plant Cell* **13**: 2609-2618

Barakat A., DiLoreto D.S., Zhang Y., Smith C., Baier K., Powell W.A., Wheeler N., Sederoff R. and Carlson J.E. (2009) Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biol* **9**: 51

Boutilier K., Offringa R., Sharma V.K., Kieft H., Ouellet T., Zhang L., Hattori J., Liu C-M., van Lammeren A.A.M., Miki B.L.A., Clusters J.B.M. and van Lookeren Campagne M.M. (2002) Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**: 1737-1749

Branton R.L. and Blake J. (1983) Development of organized structures in callus derived from explants of *Cocos nucifera* L. *Ann Bot* **52**:673-678

Brautigam A. and Gowik U. (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol (Stuttg)* **12**(6):831–841

Cairney J. and Pullman G.S. (2007) The cellular and molecular biology of conifer

embryogenesis. *New Phytologist* **176**: 511–536

Casson S., Spencer M., Walker K. and Lindsey K. (2005) Laser capture microdissection for the analysis of gene expression during embryogenesis of *Arabidopsis*. *Plant J* **42**: 111-123

Fernando S.C. and Gamage C.K. (2000) Abscisic acid induced somatic embryogenesis in immature embryo explants of coconut (*Cocos nucifera* L.). *Plant Sci.* **151**: 193-198

George E.F. and Sherrington P.D. (1984) Plant Propagation by Tissue Culture-Handbook and Directory of Commercial Laboratories. Exegetics Ltd, Edington, UK

Graham I.A., Besser K., Blumer S., Branigan C.A., Czechowski T., Elias L., Guterman I., Harvey D., Isaac P.G., Khan A.M., Larson T.R., Li Y., Pawson T., Penfield T., Rae A.M., Rathbone D.A., Reid S., Ross J., Smallwood M.F., Segura V., Townsend T., Vyas D., Winzer T. and Bowles D. (2010) The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327**:328–331

Hornung R. (1995) Micropropagation of *Cocos nucifera* (L). from plumular tissues excised from mature zygotic embryos. *Plant Rech Dev* **2**: 38-41

Karunaratne S. and Periyapperuma K. (1989) Culture of immature embryos of coconut (*Cocos nucifera* L.): callus proliferation and somatic embryogenesis. *Plant Sci* **62**: 247-253

Kim S., Soltis P.M., Wall K. and Soltis D.E. (2006) Phylogeny and domain evolution in the APETALA2-like gene family. *Mol Biol Evol* **23**: 107-120

Li Y., Luo H.M., Sun C., Song J.Y., Sun Y.Z., Wu Q., Wang N., Yao H., Steinmetz A. and Chen S.L. (2010) EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. *BMC Genome* **11**:268

Luo H., Li Y., Sun C., Wu Q., Song J., Sun Y., Steinmetz A. and Chen S. (2010) Comparison of 454-ESTs from *Huperzia serrata* and *Phlegmariurus carinatus* reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. *BMC Plant Biol* **10**:209

Morcillo F., Gallard A., Pillot M., Jouannic S., Aberlenc-Bertossi F., Collin M., Verdeil J-L. and Tregear J.W. (2007) EgAP2-1, an *AINTEGUMENTA-like* (*AIL*) gene expressed in meristematic and proliferating tissues of embryos in oil palm. *Planta* **226**: 1353-1362

Morozova O., Hirst M. and Marra M.A. (2009) Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* **10**:135–151

Nole-Wilson S., Tranby T.L. and Kreizek A. (2005) *AINTEGUMENTA-like* (*AIL*) genes are expressed in young tissues and may specify meristematic or division-competent states. *Plant Mol Biol* **57**: 613-628

Novaes E., Drost D.R., Farmerie W.G., Pappas G.J., Grattapaglia D., Sederoff R.R. and Kirst M. (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* **9**: 312

Okamuro J.K., Caster B., Villarroel R., VanMontagu M. and Jofuku K.D. (1997) The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc Natl Acad Sci USA* **94**: 7076-7081

Olsvik P.A., Lie K.K., Jordal A.E., Nilsen T.O. and Hordvik I. (2005) Evaluation of potential reference genes in real time RT-PCR studies of Atlantic salmon. *BMC Mol Biol* **6**: 21

Ouakfaoui S.E., Schnell J., Abdeen A., Colville A., Labbe H., Han S., Baum B., Laberge S. and Miki B. (2010) Control of somatic

embryogenesis and embryo development by AP2 transcription factors. *Plant Mol Biol* **74**: 313-326

Palovaara J., Hallberg H., Stasolla C. and Hakman I. (2010) Comparative expression pattern analysis of WUSCHEL related homeobox 2 (WOX2) and WOX8⁄9 in developing seeds and somatic embryos of the gymnosperm *Picea abies. New Phytologist* **188**: 122–135

Perera P.I.P., Hocher V., Verdeil J-L., Bandupriya H.D.D., Yakandawala D.M.Y., Weerakoon L.K. (2008) Androgenic potential of coconut (*Cocos nucifera* L.). *Plant Cell Tiss Organ Cul* **92**: 293-302

Perera P.I.P., Hocher V., Verdeil J-L., Doulbeau S., Yakandawala D.M.Y., Weerakoon L.K. (2007) Unfertilized ovary: a novel explant for coconut (*Cocos nucifera* L.) somatic embryogenesis. *Plant Cell Rep* **26**: 21-28

Srinivasan C., Liu Z., Heidmann I. and Supena E.D.J. (2007) Heterologous expression of the *BABY BOOM* AP2/ERF transcription factor enhances the regeneration capacity of tobacco (*Nicotiana tabacum* L.). *Planta* **225**: 341-351

Tsuwamoto R., Yokoi S. and Takahata Y. (2010) *Arabidopsis EMBRYOMAKER* encoding an AP2 domain transcription factor plays a key role in develpomental change from vegetative to embryonic phase. *Plant Mol Biol* **73**: 481-492

Verdeil J-L., Huet C., Grosdemange F. and Buffard-Morel J. (1994) Plant regeneration from cultured immature inflorescence of coconut (*Cocos nucifera* L.): evidence for somatic embryogenesis. *Plant Cell Rep* **13**: 218-221

Wang W., Wang Y., Zhang Q., Qi Y. and Guo D. (2009) Global characterization of Artemisia annua glandular trichome transcriptome using 454 pyrosequencing. *BMC Genom* **10**:465