Analysis and Functional Annotation of Expressed Sequence Tags from Coconut (*Cocos nucifera* L.) Mature Embryo

H. D. D. Bandupriya[a b *], and J. M. Dunwell[b]

[a] - *Tissue Culture Division, Coconut Research Institute, Lunuwila, Sri Lanka*
[b] - *School of Biological Sciences, University of Reading, Reading, United Kingdom*

*- Corresponding author: dbandupriya@yahoo.com

ABSTRACT

The coconut *Cocos nucifera* L. a member of Arecaceae is one of the major economically important woody palms, which is cultivated mainly for edible oil. The advancement in molecular tools such as expressed sequence tag (EST) library construction provide a novel opportunity for insight into the physiology of this palm species. In this paper we report a rich EST collection from coconut mature zygotic embryos using the cost effective 454 GS-FLX pyrosequencingtechnology. The unigenes of the newly built library were compared to publicly available protein databases and Gene Ontology (GO) terms were determined. After clustering and assembly using the Newbler software, 155872 of high quality reads were assembled in to 5332 contigs and 36337 singletons. The contigs were search against the National Center for Biotechnology Information (NCBI) non-redundant (NR) protein database using the blastx algorithm. Of theses 54% contigs showed positive hits with the NCBI database. Of the embryo tissue ESTs among the top most highly expressed transcripts; there were some ESTs encoded for previously described genes which we can assume having functions during somatic and zygotic embryogenesis. Detailed functional annotation of the unigenes obtained by GO slim terms by using BLAST2GO software revealed assigning annotation classes to 2448 (46%) of the 5332 contigs in mature embryo library which generated 10675 GO terms. Based on these data, this large scale genomic resource will lay a foundation and facilitate comparative genomic studies between coconut and other monocotyledonous and dicotyledonous plants by facilitating gene discovery and functional studies.

Key words: *Coconut, expressed sequence tag, gene ontology, mature zygotic embryo*

INTRODUCTION

Even though there is an increase in the molecular data available for model species and many crops, several agronomically important crop species still do not have genomic or expressed sequence data available. For such species transcriptome data mining has been carried out by cDNA library construction. With the improvement of DNA-sequencing technology, high speed, high-throughput methods for sequencing have resulted in the replacement of these methods by much more optimized modern methods. Moreover, the sequencing technology is rapidly changing owing to the invention and commercial introduction of new approaches which are popularly known as next generation sequencing technologies (Mochida and Shinozaki, 2010). These high technology instruments became available commercially in 2004. A next generation high-throughput sequencing method based on the Roche 454 Genome Sequencer (GS) FLX platform has emerged (Margulies *et al.*, 2005) as a cost effective and most widely used *de novo* EST sequencing which has been used so far for the successful construction of EST libraries from different plant species including the model plant *Arabidopsis thaliana* (Weber *et al.*, 2007), maize (Vega-Arreguin *et al.*, 2009), olive (Alagna *et al.*, 2009), chestnut (Barakat *et al.*, 2009), Eucalyptus (Novaes *et al.*, 2008), as well as fish species (Kristiansson *et al.*, 2009)¾ ', coral species *A. millepora* (Meyer *et al.*, 2009), worms (Shin *et al.*, 2008) and insects (Hahn *et al.*, 2009). The 454 sequencing technology can identify a large number of expressed sequences and the huge amount of data generated from this technique enables the sequencing of large genome species which were inaccessible using traditional sequencing methods. The 454 sequencing technology not only identify a large number of expressed sequences, but it can also discover new genes via deep sequencing thus an effective method to revealing the expression of many rare transcripts. The sequenced cDNA show direct information on the mature transcripts for coding part of the genome, so EST databases are very useful tools for the discovery of novel genes, investigation of genes of unknown function, comparative genomic studies, gene mapping, and functional studies. This intensive growth of nucleotide sequences availability in public databases allows us to use them as reference data to discover new genes and compare them between species.

The coconut (*Cocos nucifera* L.) a member of Arecaceae is one of the major economically important woody palms, which is cultivated mainly for edible oil. 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 worldwide (Fernando and Gamage, 2000; Hornung, 1995; Karunaratne and Periyapperuma, 1989; Perera *et al.*, 2008; Perera *et al.*, 2007; Verdeil *et al.*, 1994). However, the tissue culture process remains fraught with several constraints. Low embryogenesis rate hampers the establishment of this process as commercially viable protocol. Therefore, understanding of the molecular basis of coconut tissue culture would provide necessary information needed for the improvement of the *in vitro* propagation protocol. Transcriptomic approaches have successfully been used to catalogue genes that are expressed during embryogenesis in economically important species (Cairney *et*

al., 2006; van Zyl*et et al.*, 2003; Sharma *et al.*, 2008). However, information on the genome sequence and transcript profiles for coconut which is of great economical and social importance in tropical countriesas a plantation crop is completely lacking. Thus it is of utmost important to initiate transcriptome analysis to gain a molecular insight of different biological processes during embryogenesis.

By April 2012 there were around 124 ESTs available in GenBankdatabases for coconut. In this paper we report a rich EST collection (more than 5000) from coconut mature zygotic embryos using the cost effective 454 GS-FLX pyrosequencingtechnology. The unigenes of the newly built library were compared to publicly available protein databases and Gene Ontology (GO) terms were determined. The availability of these EST sequences will lay a foundation and facilitate comparative genomic studies between coconut and other monocotyledonous and dicotyledonous plants, facilitate gene discovery and functional studies, support development of cDNAmicroarrays.

MATERIALS AND METHODS

Plant material

The coconut variety Sri Lanka Tall was used as the plant material in this study. Plant materials were obtained from Bandirippuwa Estate, Coconut Research Institute, Sri Lanka. Mature embryos at the age of 12 months after pollination (12ME) were collected for the analysis. After dissecting the embryos from the endosperm they were immediately frozen in liquid nitrogen and stored at - 80 °C until further processing.

RNA extraction, cDNA synthesis and sequencing

Total RNA was isolated using the RNeasy PlantMini kit (Qiagen) using manufacturer's protocol. The RNA quality was tested using a 1% ethidium bromide-stained (EtBr-stained) agarose gel, and the concentration was assessed using a NanoDrop™ ND-1000 spectrophotometer (ThermoScientific, UK) before processing. The RNA samples were treated with DNase I (RNase free DNase set; Qiagen) prior to cDNA synthesis.

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 GCCd(T)20–3') to synthesize the firststrand cDNA in the presence of PowerScript Reverse Trascriptase (BD Biosciences Clontech). For double-stranded cDNA (ds cDNA) synthesis, the cDNA was diluted and amplified using PCR Advantage II polymerase (BD Biosciences Clontech) in the presence of SMART PCR primer (5'-AAG CAG TGG TAT CAA CGC AGA GT- 3'), and the following the thermal profile: 1 min at 95 °C followed by 25 cycles of 95 °C for 7 s, 65 °C for 20 s, and 72 °C for 3 min. 5 µL 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 ng/µL. A total yield of 3 µg of cDNA was prepared for each tissue

type by conducting several long distance PCR reactions. Samples were sent to the Centre for Genomic Research, University of Liverpool for further sample preparation and sequencing using the 454-GS FLX Genome Sequencer.

EST assembly and annotation

High quality sequences were selected for further processing and assembly using the GS FLX software v2.0.01 by using a series of normalization and quality filtering techniques for the screening of weak and low quality sequences.Adapter trimming and poly(A/T) and short sequence (< 50 bp) removal were performed by in-house Perl scripts to obtain clean ESTs. The Newblersoftware which is provided with the Roche GS FLX sequencer was used for the assembly of these high quality sequences into consensus contigs. The edited EST was translated into six reading frames and compared with the non-redundant protein database at the National Center for Biotechnology Information (NCBI) using the default setting of BLASTX program. BLASTX results with E-values equal or less than 10^{-6} were treated as 'significant matches', whereas ESTs with no hits or matches with E-values more than 10^{-6} to proteins in NCBI were classified as 'no significant matches'. The ESTs were mapped to Gene Ontology (GO) by using Blast2GO (Conesa*et et al., 2005*) and summarized according to their molecular functions, biological processes and cellular components. The B2GO software implements blast, mapping (retrieving GO terms associated with each blast hit), GO annotation (giving functional terms to each query) based on their function, statistical testing and InterProScan tools.

RESULTS AND DISCUSSION

EST generation

A total of 207624 ESTs with an average length 201bp were generated. Cleaning of the raw sequences (removal of primer, polyA tail, etc) resulted in a total of 155872 high quality reads. After clustering and assembly using the Newbler software, these reads were assembled in to 5332contigs and 36337singletons. These 5332 contigs were assembled using a total number of 119535 reads. The average contig length was 418bp. The length distribution of contigs and their component reads are summarized in Table 1 and Table 2. It demonstrated that majority of contigs of the library fall between 200-500 bp (Table 1). This encountered 3478 contigs (65.2%) from the total number of contigs. Of the assembled contigs, 23.2% contained less than 200 bp where as 3.0% of had more than 1000 bp. When considering the number of reads per contig, majority of them had reads less than 10. However, 24% of contigs had more than 10 ESTs (Table 2).

The contigs were search against the NCBI non-redundant (NR) protein database using the blastx algorithm. 54% (2870 contigs out of 5332) contigs showed positive hits with the NCBI database. The representation of majority of the ESTs as unclassified and with no BLAST hits may be due one of the following reasons: 1) advantage of deep sequencing ability in 454 sequencing may have produced novel genes which are specific to coconut. 2) The contigs may represent 3' or 5' UTR non coding regions which do not have protein matches 3) the short sequence

lengths of the contigs may have resulted in the low annotation efficiency 4) some of them may have incomplete coverage by the genome. The observation from a previous study has confirmed that the EST length was a significant co-relater of the presence or absence of a significant BLAST match (Lokanathan *et al.*, 2010). The presence of unknown or unclassified ESTs in large proportions is always associated with transcriptome analysis studies (Bettencourt *et al.*, 2010; Low *et al.*, 2008; Sun *et al.*, 2010). The top hits for BLASTX results are shown in Fig. 2. It showed that the majority of sequences of 12ME have protein similarity with *Oryza sativa* followed by *Vitisvinifera* and *Zea mays*.

Table 1 - Length distribution of assembled contigs

Nucleotides length (bp)	Contigs
93- 100	11
101- 200	649
201- 300	978
301- 400	1358
401- 500	1142
501- 600	417
601-700	248
701- 800	159
801- 900	115
901- 1000	93
1001- 1500	140
1501- 2000	21
>2000	1
Total	**5332**
Maximum length	**2148 bp**
Average length	**418 bp**

Table 2 - Summary of component reads per assembly

Number of reads	Number of contigs
1-10	4059
11-20	649
21-30	240
31- 40	110
41-50	71
51-100	126
101-150	55
151-200	12
201-250	8
>250	2
Total	**5332**

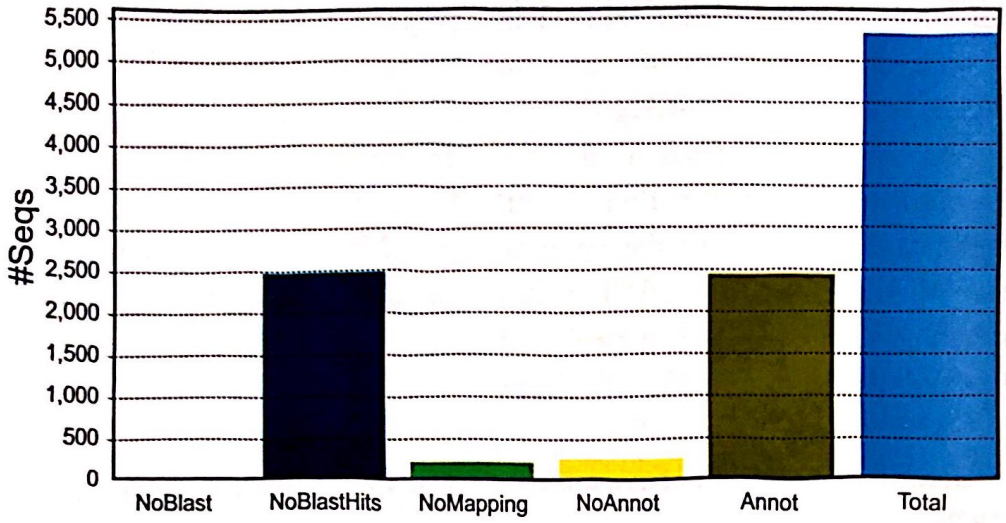

Fig. 1 - Gene ontology annotation process results in 12ME library of coconut

Putative identity of abundant genes in mature zygotic embryo tissues

In this study, EST data were used to identify the genes which are expressed abundantly assuming that higher number of reads in a particular contig represent higher number of mRNA molecules encoding that gene in a given EST library as described in previous studies (Costa *et al.*, 2010; Ho *et al.*, 2007). The most highly expressed top 20 genes in 12ME library are given in Table 3. This table was generated by comparing the BLAST2GO data after the BLASTX step with the manually generated BLASTX data for the top 100 highly abundant contigs in each library. The annotations were generated manually such that if the top BLAST hit was not informative the second or third hit was used for putative gene annotation. In the 12ME library, the most highly expressed gene (259 ESTs; contig 5259) was an unnamed protein and the second most abundant transcript was encoded for Elongation factor 1 gene which is considered as one of the genes involving in the housekeeping functions.

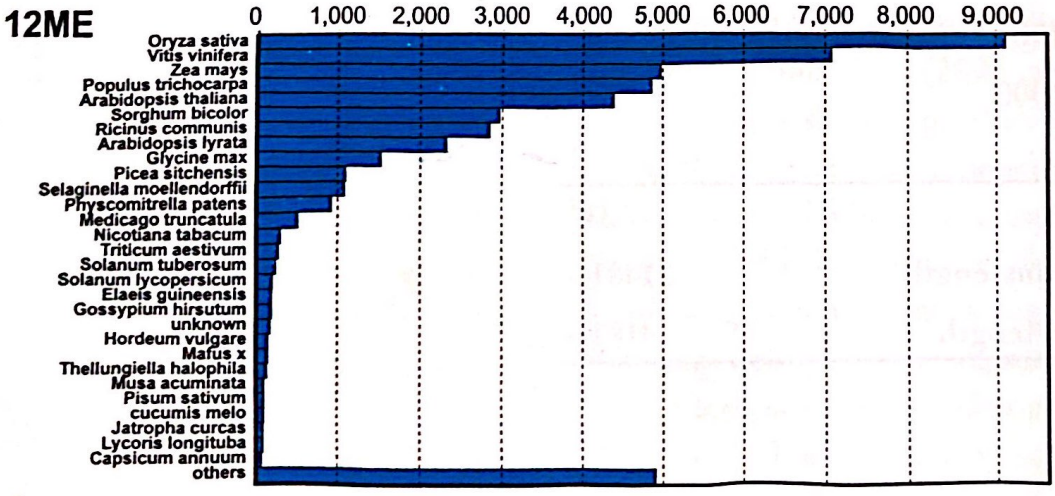

Fig. 2 - Species distribution chart of BLASTX similarity matches in mature zygotic embryo EST library.

Table 3 - Putative identity of the 20 most abundant sequences in mature zygotic embryos of coconut

Contig Number	No of EST	Putative identity	Species	GI Number	E value
5259	249	unnamed protein	*Vitisvinifera*	157356287	4.13E-87
3106	235	Elongation factor 1	*Populustrichocarpa*	7489318	2.0E-5
4343	213	auxin-repressed kda protein	*Zea mays*	195612466	1.15E-11
41	211	AP2 protein (AINTEGUMENTA-like)	*Elaeisguineensis*	56567285	1.66E-166
226	187	24-sterol C-methyltransferase	*Gossypiumhirsutum*	73761691	8.76E-31
4331	172	pyrophosphate-dependent phosphofructo-1-kinase	*Elaeisoleifera*	55296628	7.54E-48
4292	162	udp arabinose mutase	*Pisumsativum*	2130521	3.81E-28
4412	161	alcohol dehydrogenase 1	*Coixlacryma-jobi*	217069784	5.53E-154
4291	156	hexokinase 2	*Nicotianatabacum*	45387409	2.76E-10
5088	153	phosphofructokinase, putative	*Ricinuscommunis*	223544315	1.0E-16
4917	151	alcohol dehydrogenases	*Populustrichocarpa*	224060281	2.26E-52
4616	150	sequence-specific dna binding transcription factor	*Oryzasativa*	223542458	2.40E-58
159	147	atp synthase cf0 subunit i	*Medicagotruncatula*	153012207	2.94E-12
4288	147	reversibly glycosylated polypeptide	*Oryzasativa*	4158232	1.14E-35
227	146	sterol 24-c-methyltransferase	*Oryzasativa*	3560531	1.11E-62
4845	145	class ichitinase	*Pisumsativum*	1705807	1.40E-33
5243	143	actin depolymerizing	*Elaeisguineensis*	7330254	5.91E-42
4334	137	cycloartenol c-24 methyltransferace	*Dioscoreazingiberensis*	30881481	1.60E-10
4713	132	hexokinase 3	*Nicotianasylvestris*	50512102	9.80E-11
4708	131	chitinase	*Nicotianatabacum*	116323	2.50E-14

Of the embryo tissue ESTs among the top 20 most highly expressed transcripts; there were some ESTs encoded for previously described genes which we can assume having functions during somatic and zygotic embryogenesis. Auxin-repressed-kda protein coding gene was at the third place among the highly expressed genes in the mature zygotic embryo library. It was found that PvIAP1, kda protein coding gene isolated from bean seeds (Walz *et al.*, 2002) has very high homology to other plant seed storage proteins such as some late embryogenesis abundant (LEA) proteins in *Arabidopsis* (Walz *et al.*, 2008). It can therefore suggest as a gene with a role during embryogenesis. The next most abundant EST has a very high sequence similarity with the EgAP2-*1* encoding AP2 transcription factor identified in *Elaeisguineensis* Jacq. (Oil palm) during embryogenesis. A homologue (*CnANT*) gene to *EgAP2-1* was identified in coconut and its expression during embryogenesis has been determined (Bandupriya and Dunwell, 2010). *EgAP2-1* and *CnANT* shared very

high sequence similarities in their conserved regions with *BABY BOOM* (*BBM*) a well charachterized embryogenesis related gene. Ectopic expression of *BBM* was sufficient to induce somatic embryogenesis and shoot development in *Brassica napus*, *Arabidopsis* and *Nicotiana tabacum*when over-expressed (Boutilier *et al.*, 2002; Srinivasan *et al.*, 2007). This suggests the activity of *BBM* in cell proliferation during embryogenesis and similar functions were suggested for homologue palm genes (Morcillo *et al.*, 2007). Furthermore, 24-sterol C-methyltransferase coding EST which is among the top five most highly expressed transcripts has shown functions during embryogenesis (Schrick *et al.*, 2002). For an example, in *Arabidopsis* it has been identified three sterol methyltransferasegenes namely *sterol methyltransferase 1 (SMT1), SMT2 and SMT3*.All three SMT genes have shown strong and spatially distinct expression in developing embryos, as detected by the *in situ* RNA hybridization (Diener *et al.*, 2000). Moreover, the *smt1* plants have shown pleiotropic defects such as poor growth and fertility, sensitivity of the root to calcium, and a loss of proper embryo morphogenesis (Diener *et al.*, 2000). *smt1* has an altered sterol content: it accumulates cholesterol and has less C-24 alkylated sterols content (Diener et al., 2000). Among the 12ME ESTs, there were few more ESTs which showed increased level of transcripts encoded for genes such as ATP synthase CF0 subunit andchitinase which have been previously described during somatic and zygotic embryogenesis of oil palm (Ho *et al.*, 2007).

Gene ontology annotation

Detailed functional annotation of the unigenes was obtained by Gene Ontology (GO) slim terms by using BLAST2GO software (Conesa *et al.*, 2005). This software performs BLASTX similarity search against the non-redundant protein sequence database with entries from GenPept, Swissprot, PIR, PDF, PDB, and NCBI RefSeq and retrieves the GO terms for the top 20 BLAST results.It was possible to assign annotation classes to 2448 (46%) of the 5332 contigs in 12ME library which generated 10675 GO terms for the contigs in this library (Fig. 2). Of the Of the sequences with significant BLAST hits, 184 sequences failed in GO term assignment due to multiple factors, i.e. failure of the Blast outputs mapping with GO terms (NoMapping), and the missing of reliable GO annotation for the query sequences in the final annotation step (No Annotation). The annotations in each contig were used to assign coconut contigs to one of the three major GO annotation categories for Molecular Function, Biological Process and Cellular Component.

Fig. 3a shows the functional classification of coconut contigs into different GO categories within the molecular function main category. The GO level 3 was used for the annotation and construction of the pie chart. Total of 1668 GO accessions were assigned for 13 different types of molecular functions in this library. A large proportion of GO assigned sequences in themolecular function category fell into nucleotide binding,protein binding and kinase activity sub categories. The comprising of the majority of the of GO terms in molecular function ontology with proteins involved in

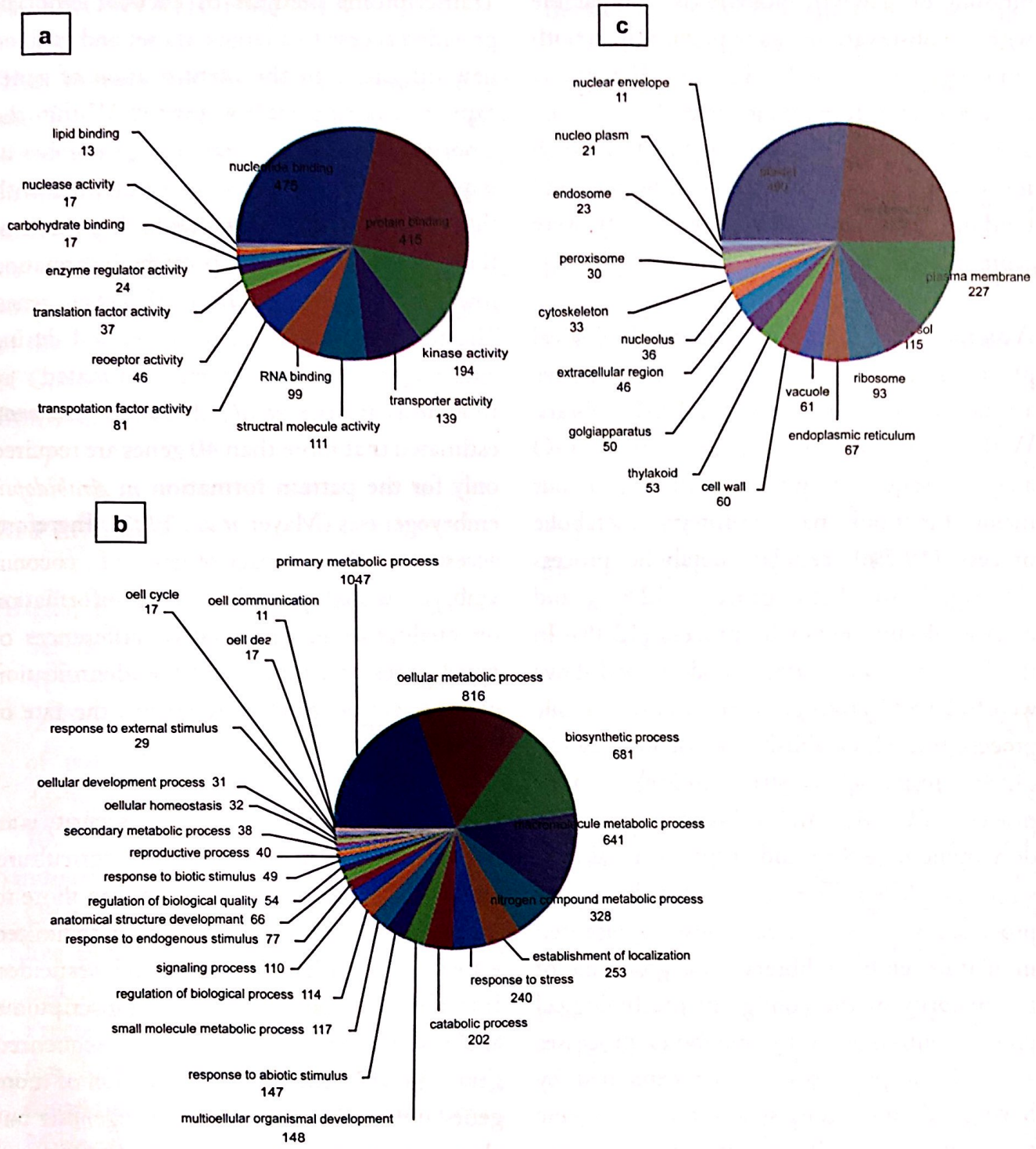

Fig. 3 - GO term distribution of ESTs in mature zygotic embryo library within the moleular function (a), biologicalprocess and cellular component(c)categories. The GO Level 3 pie chart was constructed with the default settings after appling GO slim.

binding or catalytic activity is comparable with the observations made previously in both embryogenesis related EST data (Rensing *et al.*, 2005) and developing seeds (Costa *et al.*, 2010). Some of the molecular functions such nuclease activity, carbohydrate binding, lipid binding and enzyme regulator activity were represented at low levels.

Assignment of GO accessions for biological process at level 3 is shown in Fig. 3b. This gave a total of 5305 GO terms for 12ME library. Within this category, the majority of GO assigned sequences were classified into four major functions namelyprimary metabolic process (19.7%), cellular metabolic process (15.3%), biosynthetic process (12.8%) and macromolecule metabolic process (12.1%).In this library, the four categories described above were followed by nitrogen compound metabolic process (6.2%), eatablishment of localization (4.8%), response to stress (4.5%),catabolic process (3.8%), multicellular organismal development (2.8%) and response to abiotic stimulus (2.8%). The rest of the biological processes were numerically less represented in mature embryo library. The grouping of the majority of the contigs in the biological process ontology in to metabolic processes and cellular processes can be explained by having cells at dividing stages in embryogenic tissues involving cell cycle thus using energy for cell maintenance as explained for a ciliate species (Lokanathan *et al.*, 2010).A total of 1617 annotations were observed for cellular component in 17 sub categories (Fig. 3c). The GO analysis identified well represented categories within cellular components including, plastid (30.0%), mitochondrial (26.4%) and plasma membrane (14.0%).

Transcriptome analysis of coconut embryos provided access to a large data set and enabled new insights into the identification of genes expressed during embryogenesis. Within the embryo tissue library were a large number of sequences that do not have similarities with the available data in Genbank. A portion of these sequences might provide information towards the identification of novel genes. The total number of genes expressed during embryogenesis has been estimated in thousands (Girke *et al.*, 2000). It has been estimated that more than 40 genes are required only for the pattern formation in *Arabidopsis* embryogenesis (Mayer *et al.*, 1991). Therefore, access to the transcriptome of coconut embryos would provide crucial information on qualitative or quantitative differences of novel genes and thus allow the identification of more genes which may govern the fate of the embryo.

Furthermore, food and bioenergy security is an important issue in contemporary agriculture. Crop yields must be doubled or even more to meet the demands on poorer soils with limited water and with less fertilizers and pesticides. Information gained from transcriptome analysis projects coupled with sequenced genomes will allow the identification of more genes not only related to embryogenesis but also associated with other quantitative and qualitative traits that can accelerate food production to meet the demand, especially in the tropical regions where we find expanding populations but with limited agricultural lands since large areas of forests must be preserved as conservation areas due to their high conservation value.

ACKNOWLEDGEMENTS

The authors are gratefully acknowledged the Tissue Culture Division of the Coconut Research Institute, Sri Lanka for the assistance in providing coconut samples. We are grateful to Dr. Andrew Meade for his kind assistance during Blast2GO analysis. This project was supported by the Commonwealth Commission of the United Kingdom and the Association of Commonwealth Universities, through the British Council.

REFERENCES

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

Bandupriya, H.D.D., and J.M. Dunwell. (2010). Characterization of *AINTEGUMENTA* like gene in Coconut (*Cocos nucifera* L.) and its expression during embryogenesis, Proceedings of 12th World Congress of the International Association for Plant Biotechnology. 6-11 June, St. Louis, Missouri, USA.

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

Bettencourt, R., M. Pinheiro, C. Egas, P. Gomes, M. Afonso, T. Shank, and R.S. Santos. (2010). High-throughput sequencing and analysis of the gill tissue transcriptome from the deep-sea hydrothermal vent mussel *Bathymodiolus azoricus*. *BMC Genomics* 11: 559.

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

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

Cairney, J., L. Zheng, A. Cowels, J. Hsiao, V. Zismann, J. Liu, S. Ouyang, F. Thibaud-Nissen, J. Hamilton, K. Childs, G.S. Pullman, Y. Zhang, T. Oh, and C.R. Buell. (2006). Expressed sequence tags from loblollypine embryos reveal similarities with angiosperm embryogenesis. *Plant Mol. Biol.* 62: 485-501.

Conesa, A., S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Costa, G.G.L., K.C. Cardoso, L.E.V. Del Bem, A.C. Lima, M.A.S. Cunha, L. de Campos-Leite, R. Vicentini, F. Papes, R.C. Moreira, J.A. Yunes, F.A.P. Campos, and M.J. Da Silva. (2010). Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L. *BMC Genomics* 11: 462.

Diener, A.C., H. Li, W.X. Zhou, W.J. Whoriskey, W.D. Nes, and G.R. Fink. (2000). *STEROL METHYLTRANSFERASE 1* controls the level of cholesterol in plants. *The Plant Cell* 12: 853–870.

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

Girke, T., J. Todd, S. Ruuska, J. White, C. Benning, and J. Ohlrogge. (2000). Microarray analysis of developing Arabidopsis seeds. *Plant Physiol.* 124: 1570-1581.

Hahn, D.A., G.J. Ragland, D.D. Shoemaker, and D.L. Denlinger. (2009). Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*. *BMC Genomics* 10: 234.

Ho, C-L., Y-Y. Kwan, M-C. Choi, S-S. Tee, W-H. Ng, K-A. Lim, Y-P. Lee, S-E. Ooi, W-W. Lee, J-M. Tee, S-H. Tan, H. Kulaveerasingam, S.S.R.C. Alwee, and M.O. Abdullah. (2007). Analysis and functional anotation of expressed sequence tags (ESTs) from multiple tissues of oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics* 8: 381.

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

Kristiansson, E., N. Asker, L. Forlin, and D.G.J. Larsson. (2009). Characterization of the *Zoarcesviviparus* liver transcriptome using massively parallel pyrosequencing. *BMC Genomics* 10: 345.

Lokanathan, Y., A. Mohd-Adnan, K.L. Wan, and S. Nathan. (2010). Transcriptome analysis of the *Cryptocaryon irritans* tomont stage identifies potential genes for the detection and control of cryptocaryonosis. *BMC Genomics* 11: 76.

Low, E.L., H. Alias, S. Boon, E.M. Shariff, C.A. Tan, L.C.L. Ooi, S. Cheah, A. Raha, K. Wan, and R. Singh. (2008). Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biology* 8: 62.

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.

Meyer, E., G.V. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J.K. Colbourne, B.L. Willis, and M.V. Matz. (2009). Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics* 10: 219.

Mochida, K., and K. Shinozaki. (2010). Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.* 51: 497-523.

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

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

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

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

Rensing, S.A., D. Lang, E. Schumann, R. Reski, and A. Hohe. (2005). EST sequencing from embryogenic *Cyclamen persicum* cell cultures identifies a high proportion of transcripts homologous to plant genes involved in somatic embryogenesis. *J. Plant Growth Regul.* 24: 102-115.

Schrick K, U. Mayer, G. Martin, C. Bellini, C. Kuhnt, J. Schmidt, G. Jurgens. (2002). Interactions between sterol biosynthesis genes in embryonic development of *Arabidopsis*. *Plant J.* 31: 61-73.

Sharma, S.K., S. Millam, P.E. Hedley, J. McNicol, and G.J. Bryan. (2008). Molecular regulation of somatic embryogenesis in potato: an auxin led perspective. *Plant Mol. Biol.* 68: 185-201.

Shin, H., M. Hirst, M.N. Bainbridge, V. Magrini, E. Mardis, D.G. Moerman, M.A. Marra, D.L. Baillie, and S.J.M. Jones. (2008). Transcriptome analysis for *Caenorhabditis elegans* based on novel expressed sequence tags. *BMC Biol.* 6: 30.

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

Sun, C., Y. Li, Q. Wu, H. Luo, Y. Sun, J. Song, E.M.K. Lui, and S. Chen. (2010). *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics* 11: 262.

van Zyl, L., P.V. Bozhkov, D.H. Clapham, R.R. Sederoff, and S. von Arnold. (2003). Up, down and up again is a signature global gene expression patternat the beginning of gymnosperm embryogenesis. *Gene ExprPatterns* 3: 83-91.

Vega-Arreguin, J.C., E. Ibarra-Laclette, B. Jimenez-Moraila, O. Martinez, J.P. Vielle-Calzada, L. Herrera-Estrella, and A. Herrera-Estrella. (2009). Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics* 10: 299.

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

Walz, A., S. Park, J.P. Slovin, J. Ludwig-Muller, Y. Momonoki, and J.D. Cohen. (2002). A gene encoding a protein modified by the phytohormone indoleacetic acid. *Proc. Natl. Acad. Sci.* USA 99. pp. 1718-1723.

Walz, A., C. Seidel, G. Rusak, S. Park, J.D. Cohen, and J. Ludwig-Muller. (2008). Heterologous expression of IAP1, a seed protein from bean modified by indole-3-acetic acid, in *Arabidopsisthaliana* and *Medicago truncatula*. *Planta* 227: 1047-1061.

Weber, A.P.M., K.L. Weber, K. Carr, C. Wilkerson, and J.B. Ohlrogge. (2007). Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiol.* 144: 32-42.