Electronic Journal of Biotechnology ISSN: 0717-3458
Vol.13 No.1, Issue of January 15, 2010
© 2010 by Pontificia Universidad Católica de Valparaíso -- Chile
Received December 16, 2008 / Accepted August 2, 2009
DOI: 10.2225/vol13-issue1-fulltext-14
RESEARCH ARTICLE

Normalized embryoid cDNA library of oil palm (*Elaeis guineensis*)

Pek-Lan Chan[1]
Advanced Biotechnology and Breeding Centre
Malaysian Palm Oil Board
No. 6, Persiaran Institusi
Bandar Baru Bangi
43000 Kajang, Selangor, Malaysia

Lay-Sun Ma[1]
Academia Sinica, 128 Sec. 2
Academia Rd, Nankang
Taipei 115, Taiwan-Republic of China

Eng-Ti Leslie Low[1]
Advanced Biotechnology and Breeding Centre
Malaysian Palm Oil Board
No. 6, Persiaran Institusi
Bandar Baru Bangi
43000 Kajang, Selangor, Malaysia

Elyana M. Shariff
Myagri Associates Sdn Bhd
25-2, Jalan Seri Putra ½
Bandar Seri Putra Bangi
43000 Kajang, Selangor, Malaysia

Leslie Cheng-Li Ooi
Advanced Biotechnology and Breeding Centre
Malaysian Palm Oil Board
No. 6, Persiaran Institusi
Bandar Baru Bangi
43000 Kajang, Selangor, Malaysia

Suan-Choo Cheah
Asiatic Centre for Genome Technology
Lot L3-I-1, Enterprise 4
Technology Park Malaysia, Bukit Jalil
57000 Kuala Lumpur, Malaysia

Rajinder Singh*
Advanced Biotechnology and Breeding Centre
Malaysian Palm Oil Board
No. 6, Persiaran Institusi
Bandar Baru Bangi
43000 Kajang, Selangor, Malaysia
E-mail: rajinder@mpob.gov.my

Financial support: Ministry of Science, Technology and Innovation (MOSTI), Malaysia under the Malaysia-MIT Biotechnology Partnership Programme (MMBPP).

Keywords: auxin, expressed sequence tags, gene expression, low abundance genes, tissue culture.

Note: The sequences have been deposited in GenBank. Accession numbers: DW247764-DW248770.

[1] **These authors contributed equally to this work.**

Abbreviations:
EC: embryogenic callus
EMB: embryoid
EO library: standard embryoid cDNA library
EON: normalized embryoid cDNA library
ESTs: expressed sequence tags
GBSS: granule-bound starch synthase

*Corresponding author

GOI: genes of interest
HAP: hydroxyapatite
MPOB: Malaysian Palm Oil Board
NEC: non-embryogenic callus
RPL2: ribosomal L2
SDS: Sodium Dodecyl Sulfate
WAA: weeks after anthesis

A normalized embryoid cDNA library (EON) was constructed based on reassociation kinetics reaction. Results from dot blot hybridization and sequencing of EON cDNA clones clearly indicated that the normalization process reduced the frequency of high abundance transcripts and increased the frequency of low abundance gene transcripts. A total of 553 non-redundant expressed sequence tags (ESTs) were identified, 325 of these were not observed in the standard oil palm cDNA libraries sequenced previously. A total of 10 EON cDNA clones were chosen for expression profiling across samples from different stages of the tissue culture process. Two of the genes exhibited promising expression patterns for predicting the embryogenic potential in callus. Some of these genes were also differentially expressed in the various tissues of oil palm. This study showed that normalization of the existing embryoid library improved the chances of identifying transcripts not captured in the standard libraries, some of which could be associated with embryogenesis. This collection of ESTs is particularly well suited for use as candidate genes for development of an oil palm DNA chip, which can be used to obtain a more comprehensive view of the molecular mechanism associated with oil palm tissue culture.

Oil palm is the most important source of vegetable oil globally. Its propagation by tissue culture was first described in the 1970s (Jones, 1974). Since then, the commercial advantage of tissue culture planting materials over conventional seedlings has been well established. However, the tissue culture process remains fraught with difficulties. The rate of callogenesis of oil palm explants remains low, at about 19% (Corley and Tinker, 2003), while the rate of embryogenesis from proliferating callus cultures is only 6 percent (Wooi, 1995). Attempts to overcome the problem *via* culture media and environmental modifications (Jones, 1974; Wong et al. 1996; Muniran et al. 2008), development of suspension culture (Tarmizi et al. 2004) and bioreactor system (Gorret et al. 2004) have somewhat improved the efficiency rate of oil palm tissue culture. However, low embryogenesis rate remains a recalcitrant problem. Therefore, understanding the molecular basis of oil palm tissue culture could provide the necessary information needed to improve the efficiency of the process further. The first step towards this would be to isolate and evaluate the genes that are expressed during oil palm tissue culture.

The number of genes expressed in plants has been estimated at between 26,500 and 45,000 (Sterck et al. 2007). Typically, messenger RNAs are divided into three frequency classes: super-prevalent (10-15 mRNA species, each with about 5,000 copies, representing 10-20% of the total mRNA mass); intermediate (1,000-2,000 mRNA species, 225 copies per species, 40-45% of the total mass) and complex (15,000-20,000 mRNA species, 15 copies per species, 40-45% of the total mass) (Bishop et al. 1974; Zhang et al. 2005). However, the distribution of the expressed genes among the different tissues is highly variable. The massive presence of super-prevalent mRNAs in a tissue often hampers large-scale expressed sequence tag (EST) sequencing (Chu et al. 2003). As EST sequencing continues, the remaining unidentified genes become progressively harder to find because they are of lower abundance and restricted to only certain cell types. The most scarce and tissue-specific genes need to be catalogued as these genes play an essential role in plant development (Reddy et al. 2002; Shary and Guha-Mukherjee, 2004). Sequencing from standard cDNA libraries is ineffective for discovering rarely expressed genes, as the intermediate and highly expressed genes are sequenced more frequently. Therefore, a sequencing programme for expressed genes needs to be strategised in order to isolate all the classes of genes. This can be achieved by constructing a normalized cDNA library containing equal representations of genes expressed in a particular cell, tissue or organ.

At least two main approaches for constructing normalized cDNA libraries have been reported (Weissman, 1987; Bonaldo et al. 1996). Weissman (1987) proposed to normalize a cDNA library through saturation hybridization of cDNAs with genomic DNA. This technique depends on the presence of relatively even copies of most genes in a genome. Although normalization can be achieved by this approach, it has been suggested that obtaining sufficient rare cDNAs to achieve saturation in the hybridization process is difficult (Soares et al. 1994). The rare cDNAs may also not hybridize to genomic DNA efficiently based on their reassociation kinetics.

The other approach depends on reannealing of cDNAs following the principle of second-order kinetics. It is postulated that rare species anneal less rapidly and that the single-stranded fraction of the cDNA becomes progressively more normalized during the course of the hybridization (Soares et al. 1994; Kopczynski et al. 1998).

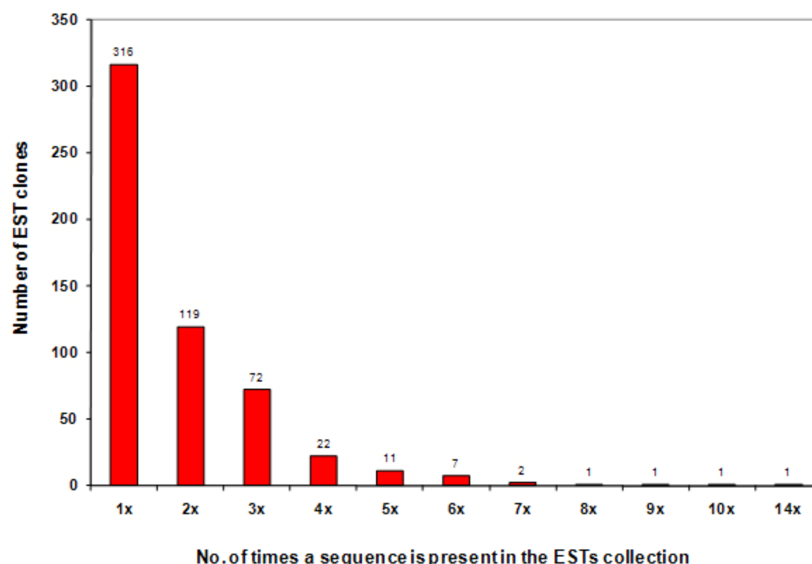

Figure 1. Frequency of redundant clones among ESTs from the normalized embryoid library (EON).

For example, when 50% of the very rare species are annealed, the most abundant species in the single-stranded fraction would not be more than twice as abundant as the rarest species. Although this reassociation-kinetics based approach is considerably more complex, libraries can be readily produced in which the minor mRNA species from a small fraction of the cells of an organ are nearly as abundantly represented as the most abundant RNA species from the predominant cell population.

Sequencing of tissue culture-related genes from oil palm standard cDNA libraries was described previously by Ho et al. (2007) and Low et al. (2008). A total of 1,723 ESTs were obtained from suspension cell culture (Ho et al. 2007), while another 17,599 ESTs were generated through the sequencing of standard cDNA libraries from leaf-derived embryogenic callus (EC), non-embryogenic callus (NEC) and embryoid (EMB) (Low et al. 2008). In order to capture rarely expressed genes associated with somatic embryogenesis, the existing EMB library was chosen for normalization following the method described by Bonaldo et al. (1996). The method is simple and has certain advantages over other procedures. It is based on the reassociation-kinetics principle involving hybridization of a 10-fold excess of driver over tracer. The DNA driver is generated by PCR amplification of the tracer (single-stranded DNA of the starting cDNA library) produced *in vivo* using the helper phage M13K07. The driver hybridizes with the tracer in the form of single-stranded circles, followed by hydroxyapatite (HAP) column purification of the remaining unbound single-stranded plasmids. The single-stranded plasmids are converted to double-strands and electroporated into bacterial host cells. This technique preserves the longest cDNAs and minimizes redundancies while increasing the representation of rare cDNAs by about three-fold on average (Bonaldo et al. 1996). The discovery of rarely expressed genes from the normalized embryoid cDNA library (EON) will enrich the existing EST collection of genes expressed during oil palm tissue culture. Availability of these non-redundant ESTs by the normalization process will aid in the development of a comprehensive cDNA microarray for identification of candidate genes that may be regulated during oil palm tissue culture.

MATERIALS AND METHODS

Plant materials

Different stages of oil palm (*Elaeis guineensis* Jacq.) tissue culture materials (NEC, EC, EMB) were provided by FELDA Agricultural Services Sdn. Bhd. Various tissues of oil palm (spear leaves, kernel at 15 weeks after anthesis (WAA), mesocarp at 15WAA, roots from seedling palms and young inflorescences) were harvested from *tenera* palms at Malaysian Palm Oil Board (MPOB)-UKM Research Station, Bangi, Selangor, Malaysia. All the samples were frozen at -80ºC prior to RNA extraction.

Construction of standard embryoid cDNA library (EO library)

Total RNA was isolated from oil palm embryoid cultures as described by Rochester et al. (1986). Poly $(A)^+$ RNA was purified using oligo-dT cellulose chromatography according to Singh and Cheah (2000). The cDNA library was constructed using the ZAP-cDNA® Gigapack® III Gold Cloning kit (Stratagene). Mass excision of cDNA clones from an amplified library was performed using the ExAssist Helper phage following the manufacturer's instructions (Stratagene). The excised phagemids were transformed into SOLR cells and grown at 37ºC overnight on LB agar plates

Table 1. Examples of newly identified non-redundant ESTs with potential involvement in the tissue culture process based on their putative functions.

EST Acc. No[a]	Putative function	Identical registration[b]	E-Value	Organism	Reference
DW247791	Fertilization-Independent Endosperm (FIE)	AAN85568.1	$6.0e^{-87}$	*Eucalyptus grandis*	Luo et al. 2000
DW247830	Putative transcription factor Myb 1	XP_480121.1	$2.0e^{-39}$	*Oryza sativa*	Shin et al. 2007
DW247837	Putative germin A	XP_480451.1	$1.0e^{-50}$	*Oryza sativa*	Neutelings et al. 1998
DW247889	Crumpled Leaf (CRL)	BAD12566.1	$3.0e^{-74}$	*Arabidopsis thaliana*	Asano et al. 2004
DW247987	F-box family protein	NP_198741.1	$1.0e^{-42}$	*Arabidopsis thaliana*	Ruegger et al. 1998
DW248114	Putative copine III	XP_466003.1	$6.0e^{-41}$	*Oryza sativa*	Hua et al. 2001
DW248206	Putative late embryogenesis abundant (LEA) protein	XP_470376.1	$8.0e^{-64}$	*Oryza sativa*	Che et al. 2006
DW248632	Homeodomain protein HOX3	AAU12247.1	$5.0e^{-62}$	*Gossypium hirsutum*	Ingouff et al. 2003
DW248696	Granule-bound starch synthase (GBSS)	AAF89270.1	$9.0e^{-108}$	*Vauquelinia californica*	Miyazawa et al. 1999
DW248718	Putative c-Myc binding protein	BAC43721.1	$2.0e^{-33}$	*Arabidopsis thaliana*	Ben-Porath et al. 1999

[a]GenBank accession number of ESTs from EON library.
[b]GenBank accession number of registered sequences identical to our ESTs.

containing 100 µg/ml ampicillin. Double-stranded DNA (dsDNA) from the SOLR cells was extracted and transformed into XL1-Blue MRF' competent cells (Hanahan, 1985).

Preparation of single-stranded DNA (ssDNA)

Conversion of dsDNA into ssDNA was carried out by super infection of XL1-Blue MRF' transformants with the helper phage M13K07 (Vieira and Messing, 1987). Transformants (white colonies) were pooled and grown at 37ºC overnight in LB broth containing 100 µg/ml ampicillin. The overnight bacterial culture was diluted to 1:100 and allowed to grow further to an OD_{600} of 0.2. The helper phage M13K07 was added to a multiplicity of infection (M.O.I) of 10- to 20-fold excess and incubated for 1 hr at 37ºC with gentle shaking (200 rpm). A final concentration of 70 µg/ml kanamycin was then added to select the infected bacterial cells, and the growth was continued overnight with vigorous shaking (300 rpm). After centrifugation (4,000 rpm, 10 min, Sorvall RC-5C PLUS), 20% PEG/2.5 M NaCl solution was added to the supernatant for precipitation of phage particles at 4ºC overnight. Phage pellet obtained from centrifugation (12,000 rpm, 20 min, Sorvall RC-5C PLUS) was dissolved in TE buffer (pH 8.0) and extracted with phenol: chloroform: isoamyl alcohol (25:24:1) for protein removal. The ssDNA phagemids were ethanol-precipitated (at least 1 hr, -20ºC) and the pellet was dissolved in TE buffer (pH 8.0). Digestion of the ssDNA with *Pvu*II enzyme was performed at 37ºC for 4 hrs prior to purification using a Bio-Gel HTP HAP column (Bio-Rad) as described previously (Soares et al. 1994; Ali et al. 2000). *Pvu*II digestion was used to eliminate double-stranded plasmid contaminants in the ssDNA preparation. The eluted ssDNA from HAP column was then concentrated by extracting twice with water-saturated butanol, once with

Table 2. Comparison of standard and normalized embryoid cDNA libraries.

	Embryoid Tissue	
	Standard (EO)	Normalized (EON)
Number of ESTs generated	3,463	1,007
Average insert size (bp)	1,400	1,500
Average read length (bp)	500	500

dry-butanol and once with water-saturated ether. The sample was desalted by passing through a spin column from QIAquick PCR Purification Kit (Qiagen), followed by purification using a Microcon YM-100 column (Millipore Inc.).

Preparation of driver DNA

The driver DNA was prepared by PCR amplification of ssDNA using the SK primer (5'-CGCTCTAGAACTAGTGGATC -3') and R primer (5'-ATAGGGCGAATTGGGTAC -3'). 10 ng of ssDNA was amplified for 35 cycles using a step cycle program (95ºC for 30 sec, 63ºC for 30 sec and 72ºC for 2 min) in 100 µl reaction mixture containing 1 µM of both primers, 200 µM dNTPs, 1.5 mM $MgCl_2$, 1 x PCR buffer (Invitrogen) and 1 U Taq DNA polymerase (Invitrogen). The amplified PCR products (driver DNA) were purified using the QIAquick PCR Purification Kit (Qiagen).

Normalization of standard embryoid cDNA library

Reassociation method described by Bonaldo et al. (1996) was used to normalize the EO library. The reaction mixture containing driver (10 µg), tracer (1 µg of single-stranded plasmid DNA), 50% deionized formamide, 200 µg each of 5'-blocking oligonucleotide (5'-CGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCACGAGG-3') and 3'-blocking oligonucleotide (5'-A_{18} CTCGAGGGGGGGCCCGGTACCC

Table 3. Predominant transcripts in the normalized embryoid library.

Putative ID	# ESTs in consensus
Ribosomal protein L23	14
Ribosomal protein S3	10
Putative polyprotein*	9
Putative pectin methylesterase*	8
Putative formamidase*	7
Unknown protein*	7
Putative RNA-binding protein RNP1 precursor*	6
MYB transcription factor	6
Extensin-like protein	6
Unknown protein	6
Pyruvate dehydrogenase kinase isoform 1; PDK1*	6

*ESTs identified in the normalized embryoid library (EON) and not observed in the sequencing of the standard embryoid library (EO).

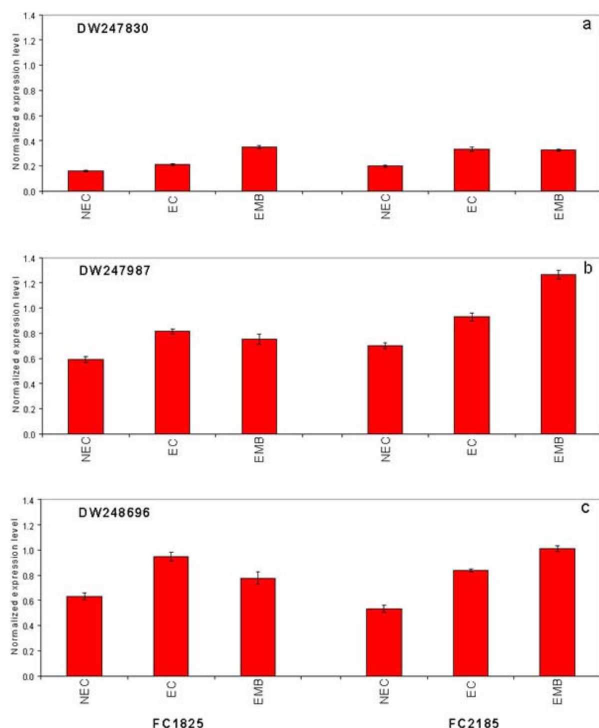

Figure 2. Three genes with up-regulated expression in EC and EMB as compared to NEC in both clonal sets. Two sets of tissue culture clones (FC1825 and FC2185) were used to determine the expression patterns of these genes using quantitative real-time PCR. Normalized expression levels were calculated from the mean values of 2 amplification reactions and the error bars represent ± standard deviation. At p-value of 0.15 generated by the ANOVA, only DW247830 and DW248696 showed significant differential gene expression between EC/EMB and NEC.
(a) DW247830 (Putative transcription factor Myb 1).
(b) DW247987 (F-box family protein).
(c) DW248696 (Granule-bound starch synthase).

AATTCGCCCTAT-3') was heated to 80ºC for 3 min under mineral oil. This was followed by addition of 20 µl of 10 x hybridization buffer (1.2 M NaCl, 0.1 M Tris pH 8.0 and 50 mM EDTA) and 20 µl of 10% SDS. Hybridization was then carried out at 35ºC for 20 hrs (Calculated C_0t ~5; C_0t value = initial DNA concentration in moles nucleotides per liter X the reassociation time in second). The remaining ssDNA was purified by HAP chromatography and converted into dsDNA using Sequenase Version 2.0 (USB Corp.) prior to electroporation into competent cells of *E. coli* DH5α'. The entire normalized cDNA library was plated onto LB agar containing 100 µg/ml ampicillin.

Sequencing and analysis of ESTs

Randomly selected clones from the normalized library were sequenced from the 5' end on an ABI PRISM 377 (Applied Biosystems). Sequence analysis was carried out using the PHRED (Ewing and Green, 1998; Ewing et al. 1998) and StackPACK programs (Miller et al. 1999). BLAST search was carried out against the GenBank non-redundant protein database. Functional classification of the non-redundant ESTs was performed using Blast2GO (Conesa et al. 2005).

Dot blot hybridization

A different batch of cDNA clones (1,002) was randomly selected from the EON library for dot blot hybridization. For this purpose, 200 ng of plasmid from each of the 1,002 cDNA clones were transferred onto Hybond-N+ nylon membranes (GE Healthcare, formerly Amersham Biosciences). The radioactive probes used for screening were gel purified cDNA fragments of pectinesterase, ribosomal L23A (60S), ribosomal L2 (50S), cys-peroxiredoxin, lipid transfer protein homolog and metallothionein-like protein. The membranes were pre-hybridized in 5 x SSPE, 5 x Denhardt's (1 x Denhardt's solution is 0.02% each Ficoll 400, bovine serum albumin and polyvinylpyrrolidone), 0.5% SDS and 100 µg/ml herring sperm DNA at 65ºC for 4 hrs. Hybridization was carried out in a buffer containing 5 x SSPE, 0.5% SDS, 100 µg/ml herring sperm DNA and ^{32}P-labelled probe (1-2 x 10^6 cpm/ml) at 65ºC overnight. The membranes were washed at 65ºC in 2 x SSC/0.1% SDS (30 min), 1 x SSC/0.1% SDS (30 min) and 0.5 x SSC/0.1% SDS (30 min). Positive clones were manually counted after exposure of the membranes to X-ray film (Kodak) for 4-8 hrs.

Quantitative real-time PCR

Total RNA was isolated from tissue culture materials and various tissues of oil palm according to McCarty (1986). The total RNA was purified using RNeasy Mini Kit (Qiagen), and on-column digestion with RNase-free DNase I (Qiagen) was performed in the RNeasy Mini Spin column according to the manufacturer's instructions. The quality of the total RNA was investigated by electrophoretic fractionation on an Agilent 2100 Bioanalyzer (Agilent Technologies) and a RNA 6000 Nano LabChip® (Caliper Technologies Corp.). Reverse transcription of total RNA to first-strand cDNA was carried out using High-capacity cDNA Reverse Transcription Kits (Applied Biosystems). A total of 2 µg of total RNA was reverse transcribed in the presence of 50 units of MultiScribe™ Reverse Transcriptase in a total volume of 20 µl. This step was performed under the following temperature scheme: 10 min at 25ºC, 120 min at 37ºC, 5 sec at 85ºC. The single-stranded cDNA was used as template for quantitative real-time PCR using ABI PRISM® 7000 Sequence Detection System (Applied Biosystems) equipped with ABI Prism 7000 SDS Software (Applied Biosystems). Specific primers and TaqMan® probes for all the genes listed in Table 1 were designed using the Primer Express Software (Applied Biosystems). For each PCR reaction, 40 or 80 ng of cDNA was added to 12.5 µl TaqMan® Universal PCR Master Mix (Applied Biosystems) and 1.25 µl 20 x TaqMan® Gene Expression Assays (Applied Biosystems) in a total volume of 25 µl. PCR was carried out at 50ºC for 2 min, 95ºC for 10 min, followed by 40 cycles at 95ºC for 15 sec and 60ºC for 1 min. Following amplification, threshold cycle (C_t) values were determined by adjusting the threshold setting

within the exponential amplification region. The Ct values were accepted if the standard deviation between two replications was smaller than 0.5. Delta-Ct method was then used to transform the Ct values to raw expression values. Three housekeeping genes with the most stable expression in oil palm tissues were used in normalization of raw expression values of the genes of interest (GOI). Geometric mean of the three housekeeping genes in each tissue, which is known as normalization factor, was calculated according to Vandesompele et al. (2002). Normalized expression levels of GOI in each tissue were obtained by dividing the raw expression values of GOI in each tissue with the appropriate normalization factor. The three housekeeping genes encode for superoxide dismutase, DNA polymerase III and unknown protein (unpublished data).

Statistics

Analysis of variance (ANOVA) was performed across normalized expression values obtained from the two different sets of tissue culture clones. This was carried out for the ESTs that were used to screen for the embryogenic potential of callus. Values of $p < 0.15$ were considered as significant differential gene expression between NEC and EC/EMB. The statistical analysis was performed using TIGR Multiexperiment Viewer version 4.1 (Saeed et al. 2003).

Table 4. Comparison of frequencies of known genes in the EO and EON libraries of oil palm.

	Percentage (Frequency)[a]		
	EO library[b]	EON library[b]	EON library[c]
Lipid transfer protein homolog	0.840 (29/3,463)	0.000 (0/1,007)	0.099 (1/1,002)
Cys-peroxiredoxin	0.430 (15/3,463)	0.000 (0/1,007)	0.000 (0/1,002)
Pectinesterase	0.000 (0/3,463)	0.794 (8/1,0 07)	0.699 (7/1,002)
Ribosomal L23A (60S)	0.116 (4/3,463)	1.390 (14/1,0 07)	0.898 (9/1,002)
Metallothionein-like protein	0.982 (34/3,463)	0.397 (4/1,007)	0.499 (5/1,002)
Ribosomal L2 (50S)	0.375 (13/3,463)	0.000 (0/1,007)	0.699 (7/1,002)
Ribosomal protein (total)	7.768 (269/3,463)	5.859 (59/1,007)	
Lectin	0.375 (13/3,463)	0.099 (1/1,0 07)	
PVR3- like protein	0.693 (24/3,463)	0.198 (2/1,007)	
PBS lyase HEAT-like repeat-containing protein	0.000 (0/3,463)	0.496 (5/1,007)	
Pyruvate dehydrogenase kinase isoform I	0.000 (0/3,463)	0.596 (6/1,0 07)	
Early nodulin	0.029 (1/3,463)	0.596 (6/1,0 07)	
Crumpled leaf	0.000 (0/3,463)	0.298 (3/1,007)	

EO: standard embryoid cDNA library.
EON: normalized embryoid cDNA library.
[a]Frequency = No. of positive clones identified/ total no. of clones analysed.
[a]Percentage (%) = frequency x 100.
[b]Data based on sequencing analysis.
[c]Data based on dot blot analysis.

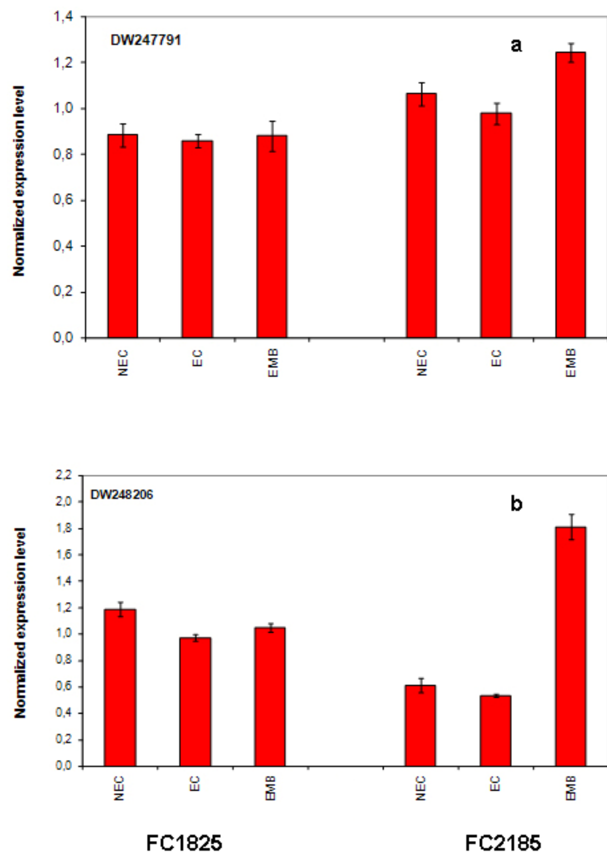

Figure 3. Genotype-specific expression patterns of two genes as validated by quantitative real-time PCR. Transcripts of DW247791 and DW248206 were only detected in the EMB tissues from FC2185. Normalized expression levels were calculated from the mean values of 2 amplification reactions and the error bars represent ± standard deviation. (a) DW247791 (Fertilization-independent endosperm).
(b) DW248206 (Putative late embryogenesis abundant protein).

RESULTS

The standard embryoid cDNA library (EO) and its normalized library (EON)

The EO library used for normalization was described by Low et al. (2008). The EO library had a titer of 10^9 pfu/ml and the average insert size of 1,400 bp as determined by PCR amplification of 3,463 randomly selected clones from the EO library using vector-specific primers. Sequences from this library have been deposited in the dbEST division of GenBank (Accession numbers: EY408451-EY413718).

A normalized library was constructed from the standard cDNA library in order to reduce the frequency of high abundance cDNAs and increase the chances of identifying the low abundance genes. The normalization involved hybridization of a large excess of cDNA inserts with the single-stranded cDNA of the EO library until a C_0t value of 5 was reached. The conditions used allowed the high and moderately abundance cDNAs to anneal to form dsDNA. The remaining ssDNA was purified by HAP chromatography, yielding the normalized EON library with at least 10,000 colonies. PCR amplification of 1,007 randomly selected cDNA clones from this library revealed that the average insert size is 1,500 bp (Table 2).

Cluster analysis of ESTs from the normalized library

The cDNA clones from the EON library were randomly selected and sequenced from the 5' end. Similarity search of 1,007 high quality ESTs with at least 400 bp insert size was performed against the GenBank protein database using BLASTX. With an E-value cut off at 10^{-6}, approximately 68% of the ESTs showed significant similarity to known sequences from plants. The other 32% of the ESTs had no matches with any known sequences. These sequences are either unique to oil palm (Tu et al. 2007) or could represent sequences containing non-coding regions such as the 5' untranslated region (5' UTR) which are less conserved across species (Jouannic et al. 2005). The 1,007 nucleotide sequences reported have been deposited in GenBank (Accession numbers: DW247764-DW248770).

StackPACK clustering analysis of the 1,007 ESTs resulted in the identification of 316 singletons (groups that contain only one EST) and 237 clusters, which were assembled from two or more ESTs in the EON library (Figure 1). A total of 553 non-redundant ESTs (singletons + a representative of each consensus) were thus identified. Of these 553 ESTs, 399 showed significantly similarity to known sequences in the non-redundant protein database at a cut off E-value of 10^{-6} (397 exhibited significant similarity to plant sequences). The remaining 154 (28%) did not show significant similarity to any known sequences in the public databases. Some of these ESTs may be specific to oil palm.

The list of EST transcripts predominant in the normalized library is shown in Table 3. Ribosomal protein L23A with a cluster of 14 sequences was predominant. The other abundant transcripts were ribosomal protein S3, putative polyprotein, putative pectin methylesterase, putative formamidase, extensin-like protein, etc. Among these transcripts, some were not identified in the standard EO library. This group of genes is also indicated in Table 3.

Evaluation of normalization by sequencing analysis

Frequency of redundant clones in EO and EON library were examined to verify the degree of normalization. The results showed that normalization had reduced the frequency of highly expressed genes, such as ribosomal protein, metallothionein-like protein, lipid transfer protein homolog, cys-peroxiredoxin and PVR3-like protein. Conversely, it increased the frequency of low abundance genes like pectinesterase, early nodulin, PBS lyase HEAT-

Table 5. Clustering result. The 553 non-redundant ESTs identified from the normalized library were clustered with MPOB's EST collection and the available published sets of oil palm ESTs.

Number of non-redundant ESTs in normalized library	553
Number of non-redundant ESTs similar to MPOB's EST collection	228
Number of non-redundant ESTs similar to the published sets of oil palm ESTs*	88
Total number of non-redundant ESTs benefited from the normalization process	237

*Published sets of oil palm ESTs by Jouannic et al. (2005) and Ho et al. (2007).

like repeat-containing protein and crumpled leaf. Table 4 summarizes the frequency of the cDNA clones observed in the sequencing of the two libraries (EO and EON). The results suggest that normalization process minimized the differential abundance of the various cDNA clones in the EO library.

Evaluation of normalization by dot blot hybridization

A total of 1,002 cDNA clones from the EON library were screened by dot blot hybridization. These clones were different from those used in the sequencing analysis. To evaluate the extent of the normalization process, six probes representing low, medium and high abundance genes in the EO library were chosen for the hybridization experiments (Table 4). Positive signals generated from hybridization between the selected probes and cDNA clones were recorded. Comparing the dot blot data of the EON library with the sequencing data of the EO library revealed that normalization reduced the frequency of highly expressed genes and simultaneously increased the frequency of the genes not captured in conventional cDNA libraries. The comparison is summarized in Table 4. As expected, the frequencies of highly abundant clones such as lipid transfer protein homolog and metallothionein-like protein were reduced 8-fold and 2-fold, respectively in the EON library. For the medium abundance clones, the frequency of cys-peroxiredoxin was significantly decreased while a 2-fold increase in ribosomal L2 (RPL2) was observed. Promising results were also obtained for the lowly abundant clones, pectinesterase and ribosomal L23A (60S). The frequency of these genes increased significantly in the EON library as compared to the EO library.

New ESTs identified by normalization

Normalization increased the frequency of identifying new ESTs and this in return increased the size of the EST collection. Clustering of the ESTs from the EO and EON libraries revealed that 76% (762 out of 1,007; 423 non-redundant ESTs) of the ESTs in the EON library were not previously sequenced. However, when the clustering was performed with all the EST collection at MPOB (17,599 ESTs, GenBank accession numbers: EY396120 to EY413718), the percentage of ESTs specific to the EON library was reduced to approximately 57% (570 out of 1,007). From these numbers, 325 non-redundant ESTs (188 singletons and 137 consensus sequences) were obtained (Electronic supplementary materials, Table S1) with 205 exhibiting significant homology to known sequences in the public databases. When the 325 non-redundant ESTs were further compared to the published sets of oil palm ESTs by Jouannic et al. (2005) and Ho et al. (2007), 237 ESTs were found unique to the EON library (Table 5). This revealed that the normalization process greatly increased the discovery of new ESTs. Examples of the ESTs identified from the normalization process together with their biological functions are given in Table 6.

Functional classification of ESTs from the EON library

Blast2GO was used to assign 325 non-redundant ESTs from the EON library to three main functional categories: Cellular Component, Biological Process and Molecular Function. It was found that only 152 ESTs (47%) were successfully mapped to the Gene Ontology Consortium structure under one or more ontologies. The summary of Gene Ontology classification at Level 2 is shown in Table 7. The ESTs were mostly assigned to the Cellular Component (349) category, with 65% involved in functions related to cell and cell part. The second highest category was assigned to the Biological Process, where about 28% were mapped to metabolic process, followed by cellular process (27%) and response to stimulus (10%). With respect to Molecular Function, only 162 sequences were classified under this category. The predominant functional classes in this category are assigned under binding (43%) and catalytic activity (42%).

Expression studies of genes associated with oil palm tissue culture using quantitative real-time PCR

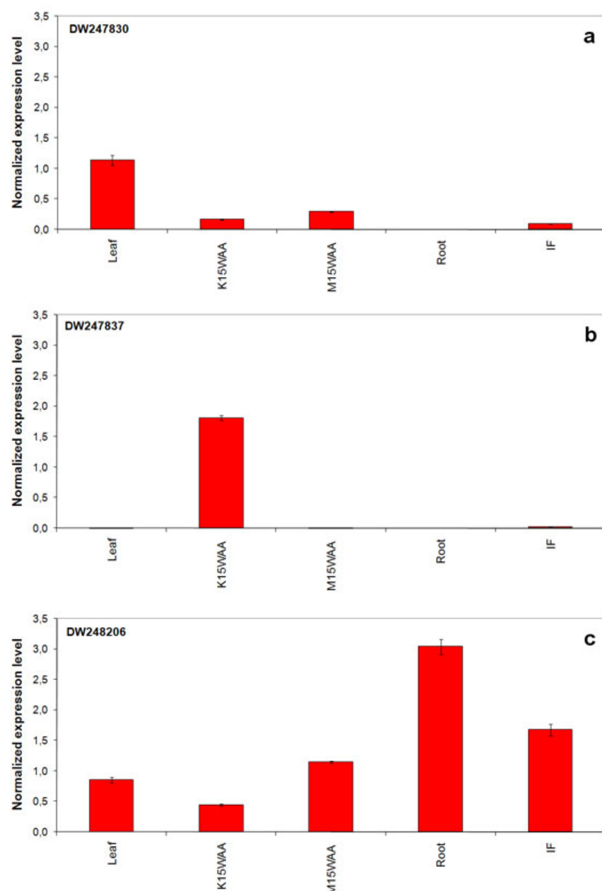

Figure 4. Group of genes with tissue-specific expression in various tissues of oil palm. Expression profiles of these genes were validated across five tissues using quantitative real-time PCR. Normalized expression level was calculated from the mean values of 2 amplification reactions and the error bars represent ± standard deviation. Leaf (spear leaves), K15WAA (kernel at 15 weeks after anthesis), M15WAA (mesocarp at 15WAA), root (roots from seedling palms) and IF (young inflorescences).
(a) DW247830 (Putative transcription factor Myb 1).
(b) DW247837 (Putative germin A).
(c) DW248206 (Putative late embryogenesis abundant protein).

Many of the unique ESTs identified encoded for genes which, based on their putative functions, may be involved in the tissue culture process. Examples of these genes are shown in Table 1. This group of genes may provide further insight into the molecular mechanism associated with oil palm tissue culture. With this in mind, quantitative real-time PCR was used to profile the expression of these genes across two sets of tissue culture clones (FC1825 and FC2185, each comprising the three main developmental stages in oil palm tissue culture: NEC, EC, EMB). The main purpose of this study was to determine if the genes are differentially expressed in NEC, EC or EMB. To ensure that the transcript levels observed were not due to differences in their genetic backgrounds, each set (NEC, EC and EMB) was sampled from a single genotype. The use of two genotypes (FC1825 and FC2185) can help determine if there is potential for the differences in transcript level to be reproducible across different genetic backgrounds. The inclusion of tissues such as spear leaves, kernel at 15 WAA, mesocarp at 15WAA, roots from seedling palms and young inflorescences (with unopened flowers) was also to determine expression of the selected genes in the various tissues of whole plant.

Based on the analysed real-time PCR data, only five of the ESTs demonstrated informative expression profiles in the tissue culture samples. The remaining five gene clones (data not shown) were expressed through out the different tissue culture developmental stages, however no significant expression profiles were observed. The three ESTs in Figure 2-DW247830 (putative transcription factor Myb1), DW247987 (F-box family protein) and DW248696 (granule-bound starch synthase)- were up-regulated in EC and EMB as compared to NEC. However, at p-value of 0.15 generated by the ANOVA, only DW247830 and DW248696 showed significant differential gene expression between EC/EMB and NEC. The expression patterns of these two transcripts were reproducible in both sets of the tissue culture clones, where they were consistently down-regulated in NEC. Up-regulation of these genes is an indicator that the callus has potential to produce embryoids. Another group of two ESTs (Figure 3) -DW247791 (fertilization-independent endosperm) and DW248206 (putative late embryogenesis abundant protein)- exhibited higher expression in EMB of the genotype FC2185. The expression patterns of these transcripts are considered genotype-specific. It may be interesting to further explore their expression profiles across a wider range of tissue culture materials.

To determine the expression of these genes in the whole plant, the 10 selected TaqMan® probes were also screened across various tissues of oil palm. The results are presented in Figure 4. Interestingly, DW247837 (putative germin A) was exclusively expressed at high levels in the kernel at 15WAA. A few other genes also showed tissue-specific expression with at least 0.8-fold up-regulation in specific tissues compared to other tissues tested. Clones DW247830 and DW248206 were abundantly expressed in the spear leaf and root tissues, respectively. The genes associated with tissue culture also appear to play a vital role in the development and maintenance of highly differentiated plant organs.

DISCUSSION

Experimental strategy of the normalized process

The main objective of this study was to construct a normalized library with approximately equal representation of all the mRNA sequences in order to increase the chances of identifying rare genes in embryoid tissues. The normalized cDNA library denoted EON was constructed based on the reassociation kinetics reaction. Effectiveness

of the normalization process was confirmed through direct sequencing and dot blot analysis. Direct sequencing of 1,007 randomly picked ESTs demonstrated changes in the redundancy level of certain genes in EON library compared to the standard EO library. Frequency of genes, such as ribosomal protein, metallothionein-like protein, lipid transfer protein homolog and cys-peroxiredoxin, which were reported to be highly abundant in the standard EO library were reduced in the EON library. The frequency of low abundance genes like pectinesterase, early nodulin, PBS lyase HEAT-like repeat-containing protein and crumpled leaf were increased. The same pattern was also observed in the dot blot hybridization experiment using a different set of EST clones (1,002). It was found that the number of lowly abundant clones such as pectinesterase and ribosomal L23A (60S) were significantly increased in the EON library. The dot blot hybridization method demonstrated here has not been used as a screening method in previous studies on normalized libraries. The method is time-consuming, since plasmids of each cDNA clone have to be prepared for spotting on the membranes. Nevertheless, it is a simple and effective method to complement the sequencing results in validating the normalization process.

Differences in frequency of ESTs between ribosomal protein L2 and L23A were observed in EON library. While ribosomal protein L2 and total ribosomal protein declined with normalization, frequency of ribosomal protein L23A increased. The results are unusual for ribosomal protein genes, as it is generally assumed that most of the ribosomal protein genes share a common promoter which leads to a unified control mechanism of gene expression. However, in the work carried out by Barakat et al. (2001) using *Arabidopsis* ESTs, it was found that the frequency of ESTs among different families of ribosomal protein genes varies greatly and the expression was also regulated differently. Furthermore, it was also observed that although RPL23A-1 and RPL23A-2 genes were 94% identical at the amino acid level, differences in the arrangement and number of predicted motifs at the 5' regulatory regions had probably contributed to the differential transcriptional regulation of these genes (McIntosh and Bonham-Smith, 2005). The analysis of ribosomal protein genes in human also concluded that small numbers of ribosomal protein genes were regulated in a tissue-specific manner and each of them is controlled by different regulators (Ishii et al. 2006). These findings will shed light on the study of ribosomal protein genes across plant species.

The efficiency of the normalization method was further supported by the clustering results of the ESTs. Comprehensive analysis revealed that 325 non-redundant ESTs from the EON library had not been previously identified in the EST collection at MPOB. About 47% of these ESTs were mapped to the GO terms, where majority of the genes appeared to be related to component of cells and organelles. This is expected as the cells at embryoid stage are actively dividing to form differentiated cells containing high storage lipid content and starch reserve (Kanchanapoom and Domyoas, 1999). When the comparison was extended to other published sets of oil palm ESTs (Jouannic et al. 2005; Ho et al. 2007), 237 ESTs were still specific to the EON library. The normalization process, as expected, increased the chances of identifying the low abundance transcripts, not isolated using conventional EST sequencing or any other routine molecular biology techniques. These genes are also listed in the Electronic supplementary material (Table S1).

Table 6. List of newly identified normalized library ESTs not found in MPOB's oil palm EST collection.

New EST	E-Value	Function
Putative pectin methylesterase	$9e^{-122}$	Cell wall modification
Glycoside hydrolase family 28 protein	$1e^{-121}$	Carbohydrate metabolism
Protein kinase family protein	$1e^{-116}$	Signal transduction
Fimbrin-like protein 2	$1e^{-108}$	Cell division
Granule-bound starch synthase I	$1e^{-103}$	Starch biosynthesis
Formamidase-like protein	$1e^{-102}$	Metabolism
Protein kinase AKINbetagamma-1	$1e^{-101}$	Signal transduction
OSJNBa0058K23.13	$1e^{-100}$	Unknown
Putative beta-ketoacyl-CoA synthase	$4e^{-99}$	Lipid biosynthesis
Putative 3-oxoacyl-[acyl-carrier-protein] synthase I	$5e^{-99}$	Fatty acid biosynthesis
Beta-adaptin-like protein A	$2e^{-96}$	Intracellular protein transport

Table 7. Gene Ontology classification (Level 2) of non-redundant ESTs from normalized embryoid library (EON).

GO Classification	No. of sequences
Cellular Component	
Cell	113
Cell part	113
Organelle	87
Organelle part	14
Macromolecular complex	13
Membrane-enclosed lumen	4
Envelope	2
Extracellular region	1
Virion	1
Virion part	1
	***349**
Biological Process	
Metabolic process	66
Cellular process	64
Response to stimulus	23
Establishment of localization	14
Localization	14
Biological regulation	13
Developmental process	12
Multicellular organismal process	10
Reproductive process	5
Reproduction	5
Immune system process	4
Multi-organism process	3
Growth	2
	***235**
Molecular Function	
Binding	69
Catalytic activity	68
Molecular transducer activity	8
Structural molecule activity	4
Transport activity	4
Antioxidant activity	4
Transcription regulator activity	4
Motor activity	1
	***162**

*Asterisk represents the total number of sequences in each category.

Identification of potential somatic embryogenesis-related genes

Regeneration of plants through the tissue culture process is important for mass propagation of elite planting materials (Sharry et al. 2006) and production of transgenic plants through genetic engineering (Masani and Parveez, 2008). In oil palm, the estimated time taken for the tissue culture process (from sampling of immature leaf explant to field planting) is around 58 months (Rohani et al. 2000). Major limitation of this process is the low embryogenesis rate. Therefore, identification of genes associated with somatic embryogenesis can help improve the current embryogenesis rate, which remains at 6%.

Initially, researchers used the candidate gene approach to identify potential genes. This led to the association of auxin-inducible genes and stress response genes with the somatic embryogenesis process in oil palm (Meilina and Ooi, 2006). Another interesting gene, serine kinase, which is a classical embryogenesis marker, is also postulated to be involved in the signal transduction pathway of oil palm somatic embryogenesis (Ooi et al. 2008). This gene is one of the non-redundant EST (325) found in the EON library and was not observed in the previously sequenced standard cDNA libraries reported by Low et al. (2008). Studies by Ooi et al. (2008) also showed that the transcript levels of these kinases were low in embryogenic calli and suspension cultures. This could be the main reason why serine kinase was not captured in the standard library and not predominantly found in EON library.

The advancement of technology has now made it possible to explore the expression profiles of large number of genes simultaneously, instead of relying on the candidate gene approach. Such technology, like microarray, has been implemented to profile the gene expression patterns during somatic embryogenesis in maize (Che et al. 2006) and soybean (Thibaud-Nissen et al. 2003). Using an alternative strategy, Low et al. (2008) successfully carried out digital northern analysis of 17,599 tissue culture ESTs to identify genes associated with oil palm embryogenesis. The genes identified by Low et al. (2008) were the lipid-transfer protein family, catalase 2, defensin EGAD1 and dehydrin-like protein. However, none of the 10 ESTs short listed for quantitative real-time PCR analysis in this study were coding for these genes. The 10 ESTs were specific to the EON library and were short listed based on their similarity to genes with potential involvement in tissue culture process. These ESTs were not identified previously probably due to the limitation of standard libraries in capturing low abundance transcripts. The presence of these unique ESTs will increase the chances of identifying additional somatic embryogenesis markers, which are required for accurate prediction of embryogenic potential across tissue culture materials.

The response of a leaf explant on a tissue culture medium is highly genotype-dependent (Thibaud-Nissen et al. 2003). Therefore, the search for molecular markers with reproducible expression patterns across tissue culture materials derived from different genetic backgrounds is a challenging task. It is for this reason that two genotypes were used to test the expression profiles of selected genes. Based on the quantitative real-time PCR analysis, this study identified two ESTs that can be used to screen for the embryogenic potential of callus (Figure 2a and Figure 2c). The first EST encodes for the granule-bound starch synthase, GBSS (DW248696), involved in the synthesis of starch in the amyloplast organelle (Denyer et al. 1999; Miyazawa et al. 1999). Histological studies by Kanchanapoom and Domyoas (1999) showed the accumulation of starch grains during induction of oil palm somatic embryogenesis. This may explain the up-regulation of the GBSS transcript in EC/EMB of both the genotypes tested. Another EST, putative transcription factor Myb1 (DW247830), is a DNA-binding protein. Involvement of this protein in regulation of auxin-inducible genes has been reported previously in *Arabidopsis* (Ruegger et al. 1998; Shin et al. 2007). As auxin is used in the induction of oil palm embryogenic cell types, it was proposed by Meilina and Ooi (2006) that this protein is likely to trigger the same pathway during oil palm somatic embryogenesis.

Genotype-specific gene expression patterns

Two ESTs demonstrated genotype-specific regulation of gene expression where they were expressed at higher levels in the EMB tissues from FC2185 (Figure 3). These ESTs are fertilization-independent endosperm (DW247791) and putative late embryogenesis abundant protein (DW248206), which encode for the proteins that have been implicated in the regulation of embryo development (Luo et al. 2000; Che et al. 2006). Spatial expression of these genes may be dependent on the developmental stage of the embryo. A recent microarray analysis by Che et al. (2006) showed a number of genes to be differentially expressed during the transition from mid to late embryo development in maize. In oil palm, embryoids at different developmental stages have been observed in a single solid culture medium (Rohani et al. 2000). Therefore, to observe a similar expression profile for the genes identified in this study, the embryos tested may have to be sampled at exactly the same developmental stage.

Identification of differentially expressed genes across various tissues of oil palm

The genes listed in Table 1 were also profiled across various tissues of oil palm. Differential expression of these genes were observed in the spear leaves (putative transcription factor Myb1), kernel at 15WAA (putative germin A) and the roots of seedling palms (putative late embryogenesis abundant protein). Among these genes, germin A demonstrated the most interesting expression pattern. This gene was detected exclusively in 15WAA kernel with minimal expression in other tissues tested (Figure 4b). The expression profile of this gene is similar to

what has been reported in conifers, where high accumulation of a germin-like protein was observed in the zygotic embryos of *Pinus caribaea* (Neutelings et al. 1998). The putative transcription factor Myb1 was found to be predominantly expressed in the spear leaves (Figure 4a). This gene has been shown to be expressed in barley leaves containing undifferentiated plastids (Churin et al. 2003). Another gene, putative late embryogenesis abundant protein exhibited a tissue-specific expression in the root tissues sampled at the actively dividing stage (Figure 4c). The localized expression of a late embryogenesis abundant protein in the root meristemic tissues has previously been demonstrated in *Phaseolus vulgaris* (Colmenero-Florez et al. 1999).

Importance of normalized embryoid library

Using the constructed normalized embryoid cDNA library, we were able to study genes related to oil palm tissue culture not previously identified in the standard library. The usefulness of the normalization process in discovering rare mRNA species is thus obvious. In other crops, such as cassava (Lokko et al. 2007), *Brassica napus* (Malik et al. 2007), cotton (Tu et al. 2007), citrus (Terol et al. 2007) and watermelon (Levi et al. 2006), genes related to economically important traits have been isolated from normalized cDNA libraries and characterized. Furthermore, the high number of unique genes available through the normalized library also enhances the assembly of an EST database. This database would be useful for high throughput assay systems, such as DNA microarray, for the analysis of gene expression profiles.

Efforts are on-going to further characterize additional genes associated with oil palm tissue culture. The collection of ESTs from the normalized embryoid library will serve as a suitable pool for mining candidate genes related to somatic embryogenesis. These candidate genes can also be used as probes for genetic mapping to determine the genomic loci associated with embryogenesis. Combining data from genome and transcriptome analysis will help enhance our understanding on the molecular mechanism associated with the tissue culture process. This will assist in the development of diagnostic tools for predicting tissue culture amenity in oil palm.

ACKNOWLEDGMENTS

The authors would like to thank the Director-General of MPOB for permission to publish this paper. We are indebted to FELDA Agricultural Services Sdn. Bhd. for providing the tissue culture materials. Special thanks to Mr. Andy Chang Kwong Choong for his valuable comments on the manuscript.

REFERENCES

ALI, Shahjahan; HOLLOWAY, Bethwyn and TAYLOR, William C. Normalization of cereal endosperm EST libraries for structural and functional genomic analysis. *Plant Molecular Biology Reporter*, June 2000, vol. 18, no. 2, p. 123-132.

ASANO, Tomoya; YOSHIOKA, Yasushi; KUREI, Shunsuke; SAKAMOTO, Wataru; SODMERGEN and MACHIDA, Yasunori. A mutation of the CRUMPLED LEAF gene that encodes a protein localized in the outer envelope membrane of plastids affects the pattern of cell divison, cell differentiation, and plastid division in *Arabidopsis*. *The Plant Journal*, May 2004, vol. 38, no. 3, p. 448-459.

BARAKAT, Abdelali; SZICK-MIRANDA, Kathleen; CHANG, Ing-Feng; GUYOT, Romain; BLANC, Guillaume; COOKE, Richard; DELSENY, Michel and BAILEY-SERRES, Julia. The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiology*, October 2001, vol. 127, no. 2, p. 398-415.

BEN-PORATH, Ittai; YANUKA, Ofra and BENVENISTY, Nissim. The *Tmp* gene, encoding a membrane protein, is a c-Myc target with a tumorigenic activity. *Molecular and Cellular Biology*, May 1999, vol. 19, no. 5, p. 3529-3539.

BISHOP, J.O.; MORTON, J.G.; ROSBASH, M. and RICHARDSON, M. Three abundance classes in HeLa cell messenger RNA. *Nature*, July 1974, vol. 250, no. 5463, p. 199-204.

BONALDO, Maria F.; LENNON, Gregory and SOARES, Marcelo B. Normalization and subtraction: two approaches to facilitate gene recovery. *Genome Research*, September 1996, vol. 6, no. 9, p. 791-806.

CHE, Ping; LOVE, Tanzy M.; FRAME, Bronwyn R.; WANG, Kan; CARRIQUIRY, Alicia L. and HOWELL, Stephen H. Gene expression patterns during somatic embryo development and germination in maize Hi II callus cultures. *Plant Molecular Biology*, September 2006, vol. 62, no. 1-2, p. 1-14.

CHU, Zhaohui; PENG, Kaiman; ZHANG, Lida; ZHOU, Bin; WEI, Jun and WANG, Shiping. Construction and characterization of a normalized whole-life-cycle cDNA library of rice. *Chinese Science Bulletin*, February 2003, vol. 48, no. 3, p. 229-235.

CHURIN, Yuri; ADAM, Eva; KOZMA-BOGNA, Laszlo; NAGY, Ferenc and BÖRNER, Thomas. Characterization of two Myb-like transcription factors binding to CAB promoters in wheat and barley. *Plant Molecular Biology*, May 2003, vol. 52, no. 2, p. 447-462.

COLMENERO-FLOREZ, José M.; MORENO, Liz P.; SMITH, Claudia E. and COVARRUBIAS, Alejandra A. Pvlea-18, a member of a new late-embryogenesis-abundant protein family that accumulates during water stress and in

the growing regions of well-irrigated bean seedlings. *Plant Physiology*, May 1999, vol. 120, no. 1, p. 93-103.

CONESA, Ana; GÖTZ, Stefan; GARCÍA-GÓMEZ, Juan M.; TEROL, Javier; TALÖN, Manuel and ROBLES, Montserrat. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, September 2005, vol. 21, no. 18, p. 3674-3676.

CORLEY, R.H.V. and TINKER, P.B. *The oil palm*. 4th ed. Oxford, Blackwell Science Ltd., 2003. 562 p. ISBN 0-632-05212-0.

DENYER, Kay; WAITE, Darren; MOTAWIA, Saddik; MØLLER, Birger L. and SMITH, Alison M. Granule-bound starch synthase I in isolated starch granules elongates malto-oligosaccharides processively. *Biochemical Journal*, May 1999, vol. 340, no. 1, p. 183-191.

EWING, Brent and GREEN, Phil. Base-calling of automated sequencer traces using *phred.* II. Error probabilities. *Genome Research*, March 1998, vol. 8, no. 3, p. 186-194.

EWING, Brent; HILLIER, LaDeana; WENDL, Michael C. and GREEN, Phil. Base-calling of automated sequencer traces using phred. I. accuracy assessment. *Genome Research*, March 1998, vol. 8, no. 3, p. 175-185.

GORRET, Nathalie; BIN ROSLI, Samsul Kamal; OPPENHEIM, Sheldon F.; WILLIS, Laura B.; LESSARD, Philip A.; RHA, ChoKyun and SINSKEY, Anthony J. Bioreactor culture of oil palm (*Elaeis guineensis*) and effects of nitrogen source, inoculum size, and conditioned medium on biomass production. *Journal of Biotechnology*, March 2004, vol. 108, no. 3, p. 253-263.

HANAHAN, D. Techniques for transformation of *E. coli*. In: GLOVER, D.M. ed. *DNA Cloning: a practical approach*. Washington, IRL Press, 1985, p. 109-135.

HO, Chai-Ling; KWAN, Yen-Yen; CHOI, Mei-Chooi; TEE, Sue-Sean; NG, Wai-Har; LIM, Kok-Ang; LEE, Yan-Ping; OOI, Siew-Eng; LEE, Weng-Wah; TEE, Jin-Ming; TAN, Siang-Hee; KULAVEERASINGAM, Harikrishna; SYED ALWEE, Sharifah S.R. and MEILINA, Ong A. Analysis and functional annotation of expressed sequence tags (ESTs) from multiple tissues of oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics*, October 2007, vol. 8, p. 381.

HUA, Jian; GRISAFI, Paula; CHENG, Shu-Hua and FINK, Gerald R. Plant growth homeostasis is controlled by the *Arabidopsis BON1* and *BAP1* genes. *Genes and Development*, September 2001, vol. 15, no. 17, p. 2263-2272.

INGOUFF, Mathieu; FARBOS, Isabelle; WIWEGER, Malgorzata and VON ARNOLD, Sara. The molecular characterization of *PaHB2*, a homeobox gene of the HD-GL2 family expressed during embryo development in Norway spruce. *Journal of Experimental Botany*, May 2003, vol. 54, no. 386, p. 1343-1350.

ISHII, Kyota; WASHIO, Takanori; UECHI, Tamayo; YOSHIHAMA, Maki; KENMOCHI, Naoya and TOMITA, Masaru. Characteristics and clustering of human ribosomal protein genes. *BMC Genomics*, February 2006, vol. 7, p. 37.

JONES, L.H. Propagation of clonal oil palms by tissue culture. *The Planter*, 1974, vol. 50, no. 585, p. 374-381.

JOUANNIC, Stefan; ARGOUT, Xavier; LECHAUVE, Frédéric; FIZAMES, Cécile; BORGEL, Alain; MORCILLO, Fabienne; ABERLENC-BERTOSSI, Frédérique; DUVAL, Yves and TREGEAR, James. Analysis of expressed sequences tags from oil palm (*Elaeis guineensis*). *FEBS Letters*, May 2005, vol. 579, no. 12, p. 2709-2714.

KANCHANAPOOM, Kamnoon and DOMYOAS, Preamrudee. The origin and development of embryoids in oil palm (*Elaeis guineensis* Jacq) embryo culture. *ScienceAsia*, December 1999, vol. 25, no. 4, p. 195-202.

KOPCZYNSKI, Casey C.; NOORDERMEER, Jasprina N.; SERANO, Thomas L.; CHEN, Wei-Yu; PENDLETON, John D.; LEWIS, Suzanna; GOODMAN, Corey S. and RUBIN, Gerald M. A high throughput screen to identify secreted and transmembrane proteins involved in *Drosophila* embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, August 1998, vol. 95, no. 17, p. 9973-9978.

LEVI, A.; DAVIS, A.; HERNANDEZ, A.; WECHTER, P.; THIMMAPURAM, J.; TREBITSH, T.; TADMOR, Y.; KATZIR, N.; PORTNOY, V. and KING, S. Gene expressed during the development and ripening of watermelon fruit. *Plant Cell Report*, November 2006, vol. 25, no. 11, p. 1233-1245.

LOKKO, Y.; ANDERSON, J.V.; RUDD, S.; Raji, A.; HORVATH, D.; MIKEL, M.A.; KIM, R.; LIU, L.; HERNANDEZ, A.; DIXON, A.G.O. and INGELBRECHT, I.L. Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Report*, September 2007, vol. 26, no. 9, p. 1605-1618.

LOW, Eng-Ti L.; ALIAS, Halimah; BOON, Soo-Heong; SHARIFF, Elyana M.; TAN, Chi-Yee A.; OOI, Leslie C.L.; CHEAH, Suan-Choo; RAHA, Abdul-Rahim; WAN, Kiew-Lian and SINGH, Rajinder. Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. *BMC Plant Biology*, May 2008, vol. 8, p. 62.

LUO, Ming; BILODEAU, Pierre; DENNIS, Elizabeth S.; PEACOCK, James W. and CHAUDHURY, Abed. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing *Arabidopsis* seeds. *Proceedings of the National Academy of Sciences of the United States of America*, September 2000, vol. 97, no. 19, p. 10637-10642.

MALIK, Meghna R.; WANG, Feng; DIRPAUL, Joan M.; ZHOU, Ning; POLOWICK, Patricia L.; FERRIE, Alison M.R. and KROCHKO, Joan E. Transcript profiling and identification of molecular markers for early microspore embryogenesis in *Brassica napus*. *Plant Physiology*, May 2007, vol. 144, no. 1, p. 134-154.

MASANI, Abdul M.Y. and PARVEEZ, Ghulam K.A. Development of transformation vectors for the production of potentially high oleate transgenic oil palm. *Electronic Journal of Biotechnology*, July 2008, vol. 11, no. 3, p. 1-9.

MCCARTY, Donald R. A simple method for extraction of RNA from maize tissues. *Maize Newsletter*, 1986, vol. 60, p. 61.

MCINTOSH, Kerri B. and BONHAM-SMITH, Peta C. The two ribosomal protein *L23A* genes are differentially transcribed in *Arabidopsis thaliana*. *Genome*, June 2005, vol. 48, no. 3, p. 443-454.

MEILINA, Ong A. and OOI, Siew-Eng. Biomarkers: finding a niche in oil palm tissue culture. Part 1-Laying the foundation. *Oil Palm Bulletin*, November 2006, vol. 53, p. 36-48.

MILLER, Robert T.; CHRISTOFFELS, Alan G.; GOPALAKRISHNAN, Chella; BURKE, John; PTITSYN, Andrey A.; BROVEAK, Tania R. and HIDE, Winston A. A comprehensive approach to clustering of expressed human gene sequence: The sequence tag alignment and consensus knowledge base. *Genome Research*, November 1999, vol. 9, no. 11, p. 1143-1155.

MIYAZAWA, Yutaka; SAKAI, Atsushi; MIYAGISHIMA, Shin-Ya; TAKANO, Hiroyoshi; KAWANO, Shigeyuki and KUROIWA, Tsuneyoshi. Auxin and cytokinin have opposite effects on amyloplast development and the expression of starch synthase genes in cultured bright yellow-2 tobacco cells. *Plant Physiology*, October 1999, vol. 121, no. 2, p. 461-470.

MUNIRAN, F.; BHORE, S.J. and SHAH, F.H. Micropropagation of *Elaeis guineensis* Jacq. ‘Dura’: comparison of three basal media for efficient regeneration. *Indian Journal of Experimental Biology*, January 2008, vol. 46, no. 1, p. 79-82.

NEUTELINGS, Godfrey; DOMON, Jean M.; MEMBRÉ, Nathalie; BERNIER, François; MEYER, Yves; DAVID, Alain and DAVID, Hélène. Characterization of a germin-like protein gene expressed in somatic and zygotic embryo of pine (*Pinus caribaea* Morelet). *Plant Molecular Biology*, December 1998, vol. 38, no. 6, p. 1179-1190.

OOI, S.E.; HARIKRISHNA, K and ONG-ABDULLAH, M. Isolation and characterization of a putative serine/threonine kinase expressed during oil palm tissue culture. *Journal of Oil Palm Research*, April 2008, vol. 1, special issue, p. 14-22.

REDDY, A.R.; RAMAKRISHNA, W.; CHANDRA SEKHAR, A.; ITHAL, N.; RAVINDRA BABU, P.; BONALDO, M.F.; SOARES, M.B. and BENNETZEN, J.L. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. *indica* cv. Nagina 22). *Genome*, February 2002, vol. 45, no. 1, p. 204-211.

ROCHESTER, Dean E.; WINER, Jill A. and SHAH, Dilip M. The structure and expression of maize genes encoding the major heat shock protein, hsp 70. *The Embo Journal*, March 1986, vol. 5, no.3, p. 451-458.

ROHANI, O.; SHARIFAH, S.A.; MOHD Rafii, Y.; ONG, M.; Tarmizi, A.H. and Zamzuri, I. Tissue culture of oil palm. In: YUSOF, B.; JALANI, B.S. and CHAN, K.W. eds. *Advances in oil palm research*. Malaysia, Malaysian Palm Oil Board, 2000, vol. 1, p. 238-283.

RUEGGER, Max; DEWEY, Elizabeth; GRAY, William M.; HOBBIE, Lawrence; TURNER, Jocelyn and ESTELLE, Mark. The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p. *Genes and Development*, January 1998, vol. 12, no. 2, p. 198-207.

SAEED, A.J.; SHARON, V.; WHITE, J.; LI, J.; LIANG, W.; BHAGABATI, N.; BRAISTED, J.; KLAPA, M.; CURRIER, T.; THIAGARAJAH, M.; STURN, A.; SNUFFIN, M.; REZANTSEV, A.; POPOR, D.; RYLTSOV, A.; KOSTUKORICH, E.; BORISOVSKY, I.; LIU, Z.; VINSARICH, A; TRUSH, V. and QUACKENBUSH, J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, February 2003, vol. 34, no. 2, p. 374-378.

SHARRY, Sandra; CABRERA PONCE, José L.; HERRERA ESTRELLA, Luis; RANGEL CANO, Rosa M.; LEDE, Silvia and ABEDINI, Walter. An alternative pathway for plant in vitro regeneration of chinaberry-tree *Melia azadarach* L. derived from the induction of somatic embryogenesis. *Electronic Journal of Biotechnology*, 2006, vol. 9, no. 3, p. 187-194.

SHARY, Semarjit and GUHA-MUKHERJEE, Sipra. Isolation and expression studies of differentiation-specific genes in tobacco dihaploids using PCR-based subtractive hybridization method. *Plant Science*, February 2004, vol. 166, no. 2, p. 317-322.

SHIN, Ryoung; BURCH, Adrien Y.; HUPPERT, Kari A.; TIWARI, Shiv B.; MURPHY, Angus S.; GUILFOYLE, Tom J. and SCHACHTMAN, Daniel P. The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. *The Plant Cell*, August 2007, vol. 19, no. 8, p. 2440-2453.

SINGH, Rajinder and CHEAH, Suan-Choo. Identification of flower specific proteins in the oil palm (*Elaeis guineensis*) by two-dimensional gel electrophoresis. *Journal of Oil Palm Research*, December 2000, vol.12, no. 2, p. 46-54.

SOARES, Marcelo B.; BONALDO, Maria F.; JELENE, Pierre; SU, Long; LAWTON, Lee and EFSTRATIADIS, Argiris. Construction and characterization of a normalized cDNA library. *Proceedings of the National Academy of Sciences of the United States of America*, September 1994, vol. 91, no. 20, p. 9228-9232.

STERCK, Lieven; ROMBAUTS, Stephane; VANDEPOELE, Klaas; ROUZé, Pierre and VAN DE PEER, Yves. How many genes are there in plants (...and why are they there?). *Current Opinion in Plant Biology*, April 2007, vol. 10, no. 2, p. 199-203.

TARMIZI, A.H.; NORJIHAN, M.A. and ZAITON, R. Multiplication of oil palm suspension cultures in a bench-top (2 litre) bioreactor. *Journal of Oil Palm Research*, December 2004, vol. 16, no. 2, p. 44-49.

TEROL, Javier; CONESA, Ana; COLMENERO, Jose M.; CERCOS, Manuel; TADEO, Francisco; AGUSTÍ, Javier; ALÓS, Enriqueta; ANDRES, Fernando; SOLER, Guillermo; BRUMOS, Javier; IGLESIAS, Domingo J.; GÖTZ, Stefan; LEGAZ, Francisco; ARGOUT, Xavier; COURTOIS, Brigitte; OLLITRAULT, Patrick; DOSSAT, Carole; WINCKER, Patrick; MORILLON, Raphael and TALON, Manuel. Analysis of 13000 unique *Citrus* clusters associated with fruit quality, production and salinity tolerance. *BMC Genomics*, January 2007, vol. 8, p. 31.

THIBAUD-NISSEN, Françoise; SHEALY, Robin T.; KHANNA, Anupama and VODKIN, Lila O. Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. *Plant Physiology*, May 2003, vol. 132, no. 1, p. 118-136.

TU, Li-Li; ZHANG, Xian-Long; LIANG, Shao-Guang; LIU, Di-Qiu; ZHU, Long-Fu; ZENG, Fan-Chang; NIE, Yi-Chun; GUO, Xiao-Ping; DENG, Feng-Ling; TAN, Jia-Fu and XU, Li. Gene expression analyses of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Reports*, August 2007, vol. 26, no. 8, p. 1309-1320.

VANDESOMPELE, Jo; DE PRETER, Katleen; PATTYN, Filip; POPPE, Bruce; VAN ROY, Nadine; DE PAEPE, Anne and SPELEMAN, Frank. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, June 2002, vol. 3, no. 7, p. research0034.1- 0034.11.

VIEIRA, Jeffrey and MESSING, Joachim. Production of single-stranded plasmid DNA. *Methods in Enzymology*, 1987, vol. 153, p. 3-11.

WEISSMAN, S.M. Molecular genetic techniques for mapping the human genome. *Molecular Biology and Medicine*, June 1987, vol. 4, no. 3, p. 133-143.

WONG, G.; TAN, C.C. and SOH, A.C. Large scale propagation of oil palm clones - experiences to date. In: *Third International Symposium on In Vitro Cultures and Horticultural Breeding*. (Jerusalem, Isreal). 1996, p. 9.

WOOI, K.C. Oil palm tissue culture-current practice and constraints. In: RAO, V.; HENSON, I.E. and RAJANAIDU, N. eds. *Recent developments in oil palm tissue culture and biotechnology*. Malaysia, Malaysian Palm Oil Board, 1995, p. 21-32.

ZHANG, Z.X.; ZHANG, F.D.; TANG, W.H.; PI, Y.J. and ZHENG, Y.L. Construction and characterization of normalized cDNA library of maize inbred MO17 from multiple tissues and developmental stages. *Molecular Biology*, March 2005, vol. 39, no. 2, p. 177-184.

Table S1. List of 325 non-redundant ESTs identified from the normalized embryoid cDNA library (EON).

EST_GenBank Accession number	Putative ID	Accession Number for Putative ID	Organism	E-Value
DW247869,DW247952,DW248094,DW248119,DW248158,DW248316,DW248682,DW248688	Putative pectin methylesterase	NP_921409.1	*Oryza sativa*	9,E-122
DW248175	Polygalacturonase (pectinase) family protein	NP_173351.1	*Arabidopsis thaliana*	3,E-121
DW248165	Protein kinase family protein	NP_199758.2	*Arabidopsis thaliana*	5,E-116
DW248163	Putative fimbrin protein	AAN13139.1	*Arabidopsis thaliana*	3,E-108
DW248661,DW248696	Granule-bound starch synthase	AAF89270.1	*Vauquelinia californica*	9,E-108
DW248097,DW248389	Formamidase-like protein	AAM64380.1	*Arabidopsis thaliana*	1,E-101
DW247956	Protein kinase akinbetagamma-1	AAG31751.1	*Zea mays*	1,E-100
DW248031	Beta-ketoacyl-coa synthase, putative	BAD32939.1	*Oryza sativa*	5,E-99
DW248456	Putative 3-oxoacyl-[acyl-carrier-protein] synthase I	BAD35225.1	*Oryza sativa*	6,E-99
DW248088	Osjnba0058k23.13	XP_473915.1	*Oryza sativa*	1,E-97
DW248129	Beta-adaptin-like protein A	NP_917777.1	*Oryza sativa*	3,E-96
DW247972,DW248331	Putative decoy	BAD53489.1	*Oryza sativa*	3,E-95
DW248249	Methionine synthase	AAL33589.1	*Zea mays*	7,E-95
DW248134	Serine/threonine protein kinase (MHK)	AAL07181.1	*Arabidopsis thaliana*	9,E-95
DW247896,DW248608	Phosphate transport protein G7, mitochondrial	T05707	*Soybean*	6,E-94
DW247911	Nitrite reductase	1908371C	*Nicotiana tabacum*	2,E-93
DW247862,DW248292	Alpha-L-arabinofuranosidase	AAP97437.1	*Malus x domestica*	4,E-92
DW247822,DW248604	Unknown protein	E96775	*Arabidopsis thaliana*	4,E-92
DW248328	Beta-tubulin	BAA06382.1	*Oryza sativa*	5,E-92
DW248393	Xyloglucan endotransglycosylase XET2	AAF80591.1	*Asparagus officinalis*	2,E-91
DW247845	Glycosyl transferase family 8 protein-like	BAD46265.1	*Oryza sativa*	9,E-89
DW248579,DW248707	Ankyrin-like protein	BAD34362.1	*Oryza sativa*	2,E-88
DW247791	Fertilization independent endosperm development protein	AAN85568.1	*Eucalyptus grandis*	6,E-87
DW248356,DW248561	26S proteasome subunit RPN9b	AAP86669.1	*Arabidopsis thaliana*	1,E-83
DW247947	Pyruvate dehydrogenase E1 beta subunit	AAC32149.1	*Picea mariana]*	4,E-83
DW248481	O-linked N-acetyl glucosamine transferase, putative	AAF26789.1	*Arabidopsis thaliana*	3,E-81
DW248264	Ubiquitin conjugating enzyme	AAB88617.1	*Zea mays*	4,E-81
DW247959,DW248099	Monodehydroascorbate reductase	AAZ66138.1	*Lycopersicon esculentum*	3,E-78
DW247785,DW247800,DW247847,DW248083,DW248708,DW248733	Putative RNA-binding protein RNP1 precursor	XP_468382.1	*Oryza sativa*	3,E-78
DW247891,DW248565	Nodulin-like protein/	NP_915846.1	*Oryza sativa*	4,E-78
DW248058	Unknown protein	AAP54296.1	*Oryza sativa*	1,E-77
DW248090	Catalase isozyme 1	P17598	*Gossypium hirsutum*	2,E-77
DW247833	Osjnba0088a01.23	CAD41383.2	*Oryza sativa*	3,E-76
DW247810	Guanine nucleotide-binding protein, putative	AAR07072.1	*Oryza sativa*	7,E-76
DW248089,DW248662,DW248683	Putative chloroplast outer membrane protein	AAV32207.1	*Oryza sativa*	2,E-74
DW247889,DW247936,DW248525	Crumpled leaf	BAD12566.1	*Arabidopsis thaliana*	3,E-74
DW248140	Cytochrome P450	AAR13307.1	*Phaseolus vulgaris*	3,E-74

DW248572	Flavoprotein alpha-subunit, having alternative splicing products, putative	XP_470330.1	*Oryza sativa*	2,E-73
DW247920	26S proteasome AAA-atpase subunit (RPT3)	BAA96920.1	*Arabidopsis thaliana*	7,E-73
DW248348	Ribosomal protein S1	AAW47294.1	*Eichhornia crassipes*	6,E-72
DW248263,DW248694	Putative S-like Rnase	BAD82178.1	*Oryza sativa*	3,E-71
DW248229	Peroxidase ATP6a	CAA67310.1	*Arabidopsis thaliana*	4,E-71
DW248689	Nucleic acid binding protein, putative	XP_470715.1	*Oryza sativa*	5,E-70
DW248239	RNA helicase, putative	AAD30595.1	*Arabidopsis thaliana*	2,E-68
DW248061,DW248626	Putative ubiquitin protein	XP_470175.1	*Oryza sativa*	3,E-67
DW248531	Peroxidase	CAD67479.1	*Asparagus officinalis*	2,E-65
DW248749	Sucrose transporter	AAV41028.1	*Saccharum hybrid cultivar*	6,E-65
DW248465	Putative enolase	AAL33814.1	*Arabidopsis thaliana*	9,E-65
DW247883	Putative NAM	BAB03447.1	*Oryza sativa*	2,E-64
DW248716	Cytosolic malate dehydrogenase	AAS18241.1	*Glycine max*	7,E-64
DW248206	Late embryogenesis abundant protein, putative	XP_470376.1	*Oryza sativa*	8,E-64
DW247818,DW248518,DW248628	Myosin XI	AAD34597.1	*Zea mays*	2,E-63
DW247781,DW248215	Osjnba0043l09.29	XP_474033.1	*Oryza sativa*	6,E-63
DW247838,DW248583	Unknown protein	XP_480138.1	*Oryza sativa*	8,E-63
DW248161,DW248373	Unknown protein	BAD28473.1	*Oryza sativa*	4,E-62
DW247922,DW248133	Peroxisomal protein PEX19 family protein	NP_568351.1	*Arabidopsis thaliana*	5,E-62
DW248632	Homeodomain protein HOX3	AAU12247.1	*Gossypium hirsutum*	5,E-62
DW247913	Expressed protein	AAN15483.1	*Arabidopsis thaliana*	2,E-61
DW247946	Peroxisomal membrane protein 22 kda, putative	NP_567940.1	*Arabidopsis thaliana*	2,E-61
DW248287,DW248295	Putative cinnamoyl-coa reductase I	XP_482628.1	*Oryza sativa*	2,E-60
DW248735,DW248767	Putative receptor kinase	AAT40550.1	*solanum demissum*	4,E-60
DW248005,DW248063	Putative 0-deacetylbaccatin III-10-O-acetyl transferase-like	AAL73122.1	*Musa acuminata*	1,E-59
DW248549	Unknown protein	XP_550260.1	*Oryza sativa*	1,E-59
DW248143	Unknown protein	AAT85152.1	*Oryza sativa*	2,E-58
DW248504,DW248763	Type A response regulator 9	CAI79413.1	*Oryza sativa*	1,E-57
DW248195	Ntgp4	AAD09518.1	*Nicotiana tabacum*	2,E-57
DW248177,DW248327	Putative receptor kinase	XP_469524.1	*Oryza sativa*	7,E-56
DW247772,DW248034,DW248514	Putative ribosomal protein S14	XP_464199.1	*Oryza sativa*	2,E-55
DW248584,DW248728	Expressed protein	XP_469186.1	*Oryza sativa*	4,E-55
DW248306,DW248645	Ferredoxin III, chloroplast precursor (Fd III)	FER3_MAIZE	*Zea mays*	5,E-55
DW248769	Unknown protein	AAO42438.1	*Arabidopsis thaliana*	7,E-55
DW247872	Unknown protein	XP_464701.1	*Oryza sativa*	4,E-54
DW248413	Alcohol dehydrogenase homolog, putative	XP_483291.1	*Oryza sativa*	4,E-53
DW248019	Helicase, putative	XP_469747.1	*Oryza sativa*	7,E-53
DW247926,DW248190,DW248697	Unknown protein	BAD44156.1	*Arabidopsis thaliana*	9,E-53
DW248399	Peroxidase	BAD97435.1	*Pisum sativum*	7,E-52
DW248074,DW248208	Putative phospatase 2A inhibitor	AAF27100.1	*Arabidopsis thaliana*	8,E-52
DW247874	Microtubial binding protein, putative	XP_480477.1	*Oryza sativa*	9,E-52
DW247852,DW248265	Osjnba0072k14.5	XP_472312.1	*Oryza sativa*	5,E-51
DW248209	Ring-box protein-like	AAM48038.1	*Arabidopsis thaliana*	5,E-51
DW247864,DW248317	Mitochondrial transcription termination factor-related	NP_195529.1	*Arabidopsis thaliana*	1,E-50
DW247837	Germin A, putative	XP_480451.1	*Oryza sativa*	1,E-50
DW248015	PK12 protein kinase	AAC04324.1	*Nicotiana tabacum*	3,E-49

DW247857,DW248159,DW248491	Putative aldo/keto reductase	AAO42123.1	*Arabidopsis thaliana*	2,E-48
DW247884,DW248166,DW248376	Germin-like protein	AAQ95582.1	*Zea mays*	3,E-48
DW248147	Dihydrodipicolinate reductase-like	BAD37640.1	*Oryza sativa*	3,E-48
DW248224	Loricrin-like protein	XP_466391.1	*Oryza sativa*	6,E-48
DW248047,DW248513	Unknown protein	XP_482104.1	*Oryza sativa*	8,E-48
DW247814	Unknown protein	AAP46254.1	*Oryza sativa*	3,E-47
DW248342	Hypothetical protein	AAP54923.1	*Oryza sativa*	1,E-46
DW248230	Unknown protein	AAM47581.1	*Sorghum bicolor*	5,E-46
DW247881,DW248712	Chloroplast stromal ascorbate peroxidase	AAS55853.1	*Vigna unguiculata*	1,E-44
DW248370	Endonuclease/exonuclease/phosphatase family protein	NP_566904.2	*Arabidopsis thaliana*	2,E-44
DW248459	Tyrosine-rich hydroxyproline-rich glycoprotein	AAA98492.1	*Petroselinum crispum*	2,E-44
DW248614	Dipeptidyl peptidase IV, putative	XP_465026.1	*Oryza sativa*	3,E-44
DW247848,DW248007,DW248017,DW248021	Putative serine carboxypeptidase	AAP51746.1	*Oryza sativa*	8,E-44
DW247976	Vesicle-associated membrane protein, putative	NP_567101.1	*Arabidopsis thaliana*	1,E-43
DW248201	Steroid sulfotransferase, putative	BAD27869.1	*Oryza sativa*	6,E-43
DW248311,DW248492	At2g01275	AAT06458.1	*Arabidopsis thaliana*	1,E-42
DW247987	F-box family protein	NP_198741.1	*Arabidopsis thaliana*	1,E-42
DW248427	At1g80190	AAQ65107.1	*Arabidopsis thaliana*	4,E-42
DW248454,DW248535	Hypothetical protein	AAN64476.1	*Oryza sativa*	7,E-42
DW248060,DW248571	Sucrose-phosphate synthase	CAA72491.1	*Craterostigma plantagineum*	2,E-41
DW247796	Histone H2A	P25469	*Lycopersicon esculentum*	2,E-41
DW247840	Laccase (diphenol oxidase)	BAB11074.1	*Arabidopsis thaliana*	4,E-41
DW247968	Expressed protein	NP_175979.1	*Arabidopsis thaliana*	5,E-41
DW248114	Copine III, putative	XP_466003.1	*Oryza sativa*	6,E-41
DW248368	Fiber protein-like	XP_482707.1	*Oryza sativa*	7,E-41
DW248576	Hypothetical protein	AAC16077.1	*Arabidopsis thaliana*	7,E-41
DW248384	Unknown protein	AAP50951.1	*Oryza sativa*	8,E-41
DW248332	Unknown protein	BAD81289.1	*Oryza sativa*	1,E-40
DW248686,DW248757	Hypothetical protein	XP_469175.1	*Oryza sativa*	2,E-40
DW248353	Unknown protein	AAP40494.1	*Arabidopsis thaliana*	2,E-40
DW248030	Erwinia induced protein 2	XP_467660.1	*Oryza sativa*	3,E-40
DW247794,DW247830	Putative transcription factor Myb1	XP_480121.1	*Oryza sativa*	2,E-39
DW247799,DW247945,DW247995,DW248307,DW248488	Putative branched-chain amino acid aminotransferase protein	NP_913581.1	*Oryza sativa*	2,E-39
DW247931,DW248355,DW248369	Tubulin alpha-6 chain, putative	AAG50105.1	*Arabidopsis thaliana*	5,E-39
DW248446	10-deacetylbaccatin III-10-O-acetyl transferase-like	BAB17109.1	*Oryza sativa*	9,E-39
DW248542,DW248633	Putative AMP-binding protein	AAN05507.1	*Oryza sativa*	3,E-38
DW247933	Unknown protein	XP_469606.1	*Oryza sativa*	5,E-38
DW248117,DW248191,DW248379	Zinc-finger protein C60910-like	XP_483536.1	*Oryza sativa*	2,E-37
DW247827	Glutamate dehydrogenase mutant	AAB51595.1	*Zea mays*	1,E-35
DW247992	Uracil phosphoribosyltransferase	CAA72093.1	*Nicotiana tabacum*	5,E-35
DW247842,DW247860,DW248762	Diacylglycerol kinase	AAS57722.1	*Elaeis oleifera]*	1,E-34
DW248217	Unknown protein	BAD68600.1	*Oryza sativa*	2,E-34
DW247899	Hypothetical protein	AAC98056.1	*Arabidopsis thaliana*	3,E-34

DW247971,DW247974,DW248197,DW248536	Phototropic-responsive NPH3 family protein	NP_199624.1	*Arabidopsis thaliana*	7,E-34
DW247878,DW248718	Putative c-myc binding protein	BAC43721.1	*Arabidopsis thaliana*	2,E-33
DW247821	Expressed protein	AAC16265.1	*Arabidopsis thaliana*	4,E-33
DW248207	Forkhead-associated domain-containing protein	NP_179748.2	*Arabidopsis thaliana*	8,E-33
DW247803,DW247975	Hypothetical protein	XP_476781.1	*Oryza sativa*	6,E-31
DW248029,DW248236,DW248415,DW248506,DW248516,DW248540,DW248545	Unknown protein	BAD81134.1	*Oryza sativa*	1,E-29
DW248467,DW248524	Beta Galactosidase-like protein	NP_914661.1	*Oryza sativa*	2,E-29
DW248739	Expressed protein	AAX92759.1	*Oryza sativa*	2,E-29
DW248123	Unknown protein	XP_470859.1	*Oryza sativa*	3,E-29
DW248050,DW248198	Unknown protein	AAF23194.1	*Arabidopsis thaliana*	8,E-28
DW248322	At5g57100	AAO24585.1	*Arabidopsis thaliana*	5,E-27
DW247914	Osjnba0084k01.15	CAE04843.2	*Oryza sativa*	9,E-27
DW248405	Unknown protein	XP_475058.1	*Oryza sativa*	5,E-26
DW247777	Auxin binding protein 1 beta1, putative	BAC56118.1	*Matricaria recutita*	1,E-25
DW248664	Small nuclear ribonucleoprotein-like	BAD81919.1	*Oryza sativa*	1,E-25
DW248711	Unknown protein	XP_467275.1	*Oryza sativa*	3,E-25
DW247769	Hypothetical protein	BAC78587.1	*Oryza sativa*	2,E-24
DW247805	Acetyl transferase, putative	XP_469881.1	*Oryza sativa*	3,E-24
DW248225,DW248381	Unknown protein	BAD68785.1	*Oryza sativa*	2,E-23
DW248092	Unknown protein	XP_478980.1	*Oryza sativa*	2,E-23
DW247887	Bzip transcription factor	AAN61914.1	*Capsicum chinense*	4,E-23
DW248103	At1g80890	AAR24692.1	*Arabidopsis thaliana*	5,E-23
DW248233	S-receptor kinase KIK1 precursor, putative	XP_478649.1	*Oryza sativa*	6,E-23
DW248120,DW248447	MET-10+related protein-like	XP_478402.1	*Oryza sativa*	3,E-22
DW248194,DW248603,DW248665	Glycine-rich protein	NP_192563.2	*Arabidopsis thaliana*	1,E-21
DW248248	Putative glyceraldehyde-3-phosphate dehydrogenase	BAD45405.1	*Oryza sativa*	1,E-21
DW248243,DW248282,DW248293	Oj000126_13.5	XP_472406.1	*Oryza sativa*	2,E-21
DW247977	Putative protein	CAB40770.1	*Arabidopsis thaliana*	3,E-21
DW247856,DW248091,DW248167,DW248255,DW248402,DW248616,DW248673	Formamidase, putative	AAN12921.1	*Arabidopsis thaliana*	2,E-20
DW248750	Stilbenecarboxylate synthase	AAN76183.1	*Hydrangea macrophylla*	3,E-20
DW247955	SET domain protein 113, putative	NP_915934.1	*Oryza sativa*	7,E-20
DW248676	At3g54290	AAP68273.1	*Arabidopsis thaliana*	1,E-19
DW247928	Cbxx protein, putative	BAD27899.1	*Oryza sativa*	3,E-19
DW248196	Vesicle transport v-SNARE 11/ vesicle soluble NSF attachment protein receptor VTI1a (VTI1A)	NP_198767.1	*Arabidopsis thaliana*	4,E-19
DW247898	SAR DNA binding protein	BAA31260.1	*Oryza sativa*	2,E-18
DW248483	Pyrrolidone carboxyl peptidase-like protein	BAD28772.1	*Oryza sativa*	1,E-17
DW248204,DW248391,DW248395,DW248420,DW248476,DW248657	Pyruvate dehydrogenase kinase isoform 1; PDK1	AAC63961.1	*Zea mays*	3,E-17
DW247912,DW248171,DW248751	Putative GDP-fucose protein-O-fucosyltransferase 2	XP_478789.1	*Oryza sativa*	6,E-17
DW248301	AGO1 homologous protein	BAB96814.1	*Oryza sativa*	1,E-16
DW248205	Ran1b	CAA98188.1	*Lotus corniculatus*	2,E-16

DW248211,DW248588	Putative polyprotein (87)	AAU90298.1	*Solanum demissum*	4,E-16
DW247942	Unknown protein	XP_450623.1	*Oryza sativa*	4,E-16
DW248122	Osjnba0011j08.3	XP_473607.1	*Oryza sativa*	6,E-16
DW247886,DW247906,DW248692	Unknown protein	AAD56321.1	*Arabidopsis thaliana*	9,E-16
DW248403,DW248517	Ran-binding protein, putative	NP_196230.2	*Arabidopsis thaliana*	1,E-15
DW248333	Unknown protein	XP_470140.1	*Oryza sativa*	1,E-15
DW248551	Hypothetical protein	XP_450124.1	*Oryza sativa*	3,E-15
DW247893,DW248086	Putative UDP-glucuronate decarboxylase 2	T40108.1	*Nicotiana tabacum*	1,E-14
DW248238	SEU3A protein	CAF18249.1	*Antirrhinum majus*	1,E-14
DW247820	Putative protein	CAB80268.1	*Arabidopsis thaliana*	2,E-14
DW248003,DW248533	Osjnba0081l15.17	XP_472941.1	*Oryza sativa*	6,E-14
DW248519	At4g27020/f10m23_360	AAL09800.1	*Arabidopsis thaliana*	6,E-14
DW248388	Unknown protein	BAC42179.1	*Arabidopsis thaliana*	2,E-13
DW247789,DW247963,DW248526	Zinc finger protein-like	BAD87039.1	*Oryza sativa*	4,E-13
DW248335,DW248470,DW248641	Putative AMP-binding protein	XP_470183.1	*Oryza sativa*	1,E-12
DW248245	Zinc finger (CCCH-type) protein-like	BAD37519.1	*Oryza sativa*	1,E-12
DW247905,DW247944,DW248080	Transcription activator	AAM52877.1	*Arabidopsis thaliana*	2,E-12
DW247909	Unknown protein	XP_464377.1	*Oryza sativa*	2,E-12
DW248312	Tryptophanyl-trna synthetase, putative	BAD87909.1	*Oryza sativa*	2,E-12
DW248468,DW248562,DW248740	Unknown protein	XP_478743.1	*Oryza sativa*	3,E-12
DW248013,DW248142,DW248406	Putative lysyl-trna synthetase	XP_466819.1	*Oryza sativa*	7,E-12
DW248111	Unknown protein	AAK43955.1	*Arabidopsis thaliana*	3,E-11
DW247895	DNA binding zinc finger protein, putative	BAD53902.1	*Oryza sativa*	6,E-11
DW248226	Zinc finger (C3HC4-type RING finger) family protein	NP_196200.2	*Arabidopsis thaliana*	9,E-11
DW247970	Guanine nucleotide-binding protein beta subunit, putative	XP_475866.1	*Oryza sativa*	3,E-10
DW248378,DW248442,DW248760	Putative transthyretin, having alternative splicing products	XP_470808.1	*Oryza sativa*	6,E-10
DW248710	Osjnbb0022f23.5	XP_472837.1	*Oryza sativa*	6,E-10
DW248546	Unknown	AAM66029.1	*Arabidopsis thaliana*	7,E-10
DW247903,DW248135,DW248329	Putative ATP-dependent Clp protease ATP-binding subunit clpx1	XP_465056.1	*Oryza sativa*	7,E-10
DW248170,DW248575	Hypothetical protein At2g26200	F84657	*Arabidopsis thaliana*	2,E-09
DW248494	Chlorophyll synthase	CAB85464.1	*Avena sativa*	4,E-09
DW247846	Oj1174_d05.12	NP_914441.1	*Oryza sativa*	2,E-08
DW248489	Membrane acyl-coa binding protein	AAT81164.1	*Agave americana*	6,E-08
DW248392	Osjnba0072d21.9	XP_472246.1	*Oryza sativa*	1,E-07
DW247824,DW247828,DW247957,DW247978,DW248172,DW248530,DW248640,DW248685,DW248714	Putative polyprotein	AAT40504.1	*Solanum demissum*	3,E-07
DW247941,DW247943	Hypothetical protein	XP_550378.1	*Oryza sativa*	5,E-07
DW248330	LIM domain containing protein-like	BAD72536.1	*Oryza sativa*	6,E-07
DW248385	Osjnba0029h02.21	XP_473065.1	*Oryza sativa*	1,E-06
DW247825	Hypothetical protein	AAP54796.1	*Oryza sativa*	2,E-06
DW248433	Lipid transfer protein (LTP) family protein	NP_566712.1	*Arabidopsis thaliana*	2,E-06

DW247798,DW247764,DW248719	Putative LYST-interacting protein LIP5	BAD62317.1	*Oryza sativa*	1,E-05
DW248547,DW248704	Similar to auxin-independent growth promoter	AAF18531.1	*Arabidopsis thaliana*	2,E-05
DW248266	DNA-binding protein Gt-2	CAA48328.1	*Oryza sativa*	2,E-05
DW248352	Glycosyl hydrolase family protein 17	NP_201547.1	*Arabidopsis thaliana*	2,E-05
DW247948,DW247993	Unknown protein	BAD37520.1	*Oryza sativa*	3,E-05
DW248232,DW248281,DW248687	Unknown protein	XP_464016.1	*Oryza sativa*	4,E-05
DW248643,DW248675,DW248726	At4g02725	AAU15166.1	*Arabidopsis thaliana*	1,E-04
DW248691	Ensangp00000013283	EAA10147.3	*Anopheles gambiae*	2,E-04
DW248453	Root border cell-specific protein, putative	BAD30223.1	*Oryza sativa*	3,E-04
DW248308	Hypothetical protein	XP_483030.1	*Oryza sativa*	6,E-04
DW248679	Cg4030-pa	NP_611545.1	*Drosophila melanogaster*	6,E-04
DW248601,DW248656	EREBP-4 like protein	CAB10530.1	*Arabidopsis thaliana*	1,E-03
DW247861,DW247925,DW248667	DNA-directed RNA polymerase	CAA36734.1	*Glycine max*	2,E-03
DW247919	Acidic ribosomal protein (60S)	AAS20966.1	*Hyacinthus orientalis*	5,E-03
DW248068	Acyl-coa oxidase	AAL01888.1	*Glycine max*	7,E-03
DW248130	4-alpha-glucanotransferase, putative	NP_201291.1	*Arabidopsis thaliana*	1,E-02
DW248503	P70 protein	CAC84774.1	*Nicotiana tabacum*	2,E-02
DW248383	Unknown protein	NP_912587.1	*Oryza sativa*	2,E-02
DW248011	Wiskott-Aldrich syndrome protein family member 4	CAD48858.1	*Homo sapiens*	3,E-02
DW248192	Homeodomain-leucine zipper, putative	NP_917179.1	*Oryza sativa*	5,E-02
DW247863,DW248361,DW248474,DW248619	Uorf	AAZ72652.1	*Craterostigma p.*	6,E-02
DW247850,DW248035,DW248069	F5o11.11	AAF79633.1	*Arabidopsis thaliana*	7,E-02
DW248484,DW248592	Protein kinase family protein	NP_850128.1	*Arabidopsis thaliana*	2,E-01
DW247949	Putative protein	CAB80056.1	*Arabidopsis thaliana*	2,E-01
DW247817,DW248452	Hypothetical protein	CAH13218.1	*Legionella pneumophila*	2,E-01
DW247915,DW248555	Arginine decarboxylase	XP_493706.1	*Oryza sativa*	3,E-01
DW248437	Serine carboxypeptidase II, CP-MII	CAA70815.1	*Hordeum vulgare*	3,E-01
DW247865,DW247877,DW248612	Spore germination B3 gerac like, C-terminal	ZP_00663356.1	*Syntrophomonas wolfei*	3,E-01
DW247927	Unknown protein	CAG13100.1	*Tetraodon nigroviridis*	3,E-01
DW248669	Chaperone grpe, putative	XP_481754.1	*Oryza sativa*	4,E-01
DW248054,DW248125,DW248699,DW248747	Response regulator 5	BAB20580.1	*Zea mays*	5,E-01
DW247997,DW248213	Unknown	AAW26857.1	*Schistosoma japonicum*	5,E-01
DW248101,DW248581	Polyketide synthase/peptide synthetase	CAG28797.1	*Magnaporthe grisea*	6,E-01
DW247768	Hypothetical protein MG05340.4	EAA54548.1	*Magnaporthe grisea*	6,E-01
DW248004,DW248638	Hypothetical protein Tb927.2.1530	XP_340252.1	*Trypanosoma brucei*	8,E-01
DW248460,DW248505,DW248630,DW248642	Unnamed protein product	BAC86178.1	*Homo sapiens*	9,E-01
DW248121,DW248432	Amino acid transporter	NP_111781.1	*Thermoplasma volcanium*	1,E+00
DW247776	D71945 hypothetical protein jhp0316	NP_223035.1	*Helicobacter pylori J99*	1,E+00
DW248520	MGC84669 protein	AAH80108.1	*Xenopus laevis*	1,E+00

DW248469	Hypothetical protein	NP_945241.1	*Bacteriophage EJ-1*	1,E+00
DW248668,DW248705,DW248715	Hypothetical protein Ecan03000892	210728,1	*Ehrlichia canis*	1,E+00
DW247792,DW247880,DW248371,DW248674	Hypothetical protein PF14_0123	NP_702011.1	*Plasmodium falciparum*	2,E+00
DW248234	Tpa: hdc06273	DAA02493.1	*Drosophila melanogaster*	2,E+00
DW248001	Myosin heavy chain, similar to Entamoeba histolytica.	AAO52540.1	*Dictyostelium discoideum*	2,E+00
DW248045,DW248438	Ankyrin containing protein	YP_142420.1	*Acanthamoeba polyph.*	2,E+00
DW248351	Probable acyltransferase	NP_925955.1	*Gloeobacter violaceus*	2,E+00
DW248408,DW248479,DW248487,DW248501	Putative nodulin mtn21	XP_463858.1	*Oryza sativa*	2,E+00
DW248020,DW248151,DW248417	Transcriptional regulator, tetr family	NP_346290.1	*Streptococcus pneumo.*	2,E+00
DW248629	Hypothetical protein efaedraft_1975	EAN10436.1	*Enterococcus faecium*	2,E+00
DW248291,DW248738	Hypothetical protein	CAH11521.1	*Legionella pneumophila*	3,E+00
DW247834,DW248462,DW248493,DW248527	Cytochrome c oxidase subunit III	AAT08574.1	*Bipes tridactylus*	3,E+00
DW247901,DW248077,DW248212	Hyothetical protein	NP_702385.1	*Plasmodium falciparum*	3,E+00
DW248315	Hypothetical protein cao19.9402	EAK91833.1	*Candida albicans*	3,E+00
DW247829,DW248009	Predicted protein	XP_324247.1	*Neurospora crassa*	3,E+00
DW248128,DW248671,DW248690	Seven transmembrane helix receptor	BAC05810.1	*Homo sapiens*	3,E+00
DW248274	Multidrug resistance protein norm	YP_269616.1	*Colwellia psychrerythraea*	3,E+00
DW247835,DW247967,DW248294,DW248725,DW248761	Atpase subunit 6	AAZ31259.1	*Cantharellus cibarius*	3,E+00
DW248269	Sugar transferase	ABA04673.1	*Nitrobacter winogradskyi*	3,E+00
DW247932	Organelle processing peptidase, putative	NP_704868.1	*Plasmodium falciparum*	3,E+00
DW248100,DW248594,DW248622,DW248644	Unnamed protein product	BAC85241.1	*Homo sapiens*	3,E+00
DW248072,DW248102	Putative ferrichrome ABC transporter	NP_802841.1	*Streptococcus pyogenes*	4,E+00
DW247951,DW248590	Hypothetical protein PF11_0484	NP_701340.1	*Plasmodium falciparum*	4,E+00
DW247826,DW248186,DW248748	PREDICTED: similar to interferon, alpha 2; alpha-2a interferon	XP_528568.1	*Pan troglodytes*	4,E+00
DW248221,DW248410	Hypothetical protein, conserved	AAX81006.1	*Trypanosoma brucei*	4,E+00
DW247767,DW248095,DW248261	Ga13283-pa	EAL30756.1	*Drosophila pseudo.*	4,E+00
DW248409,DW248700	Conserved hypothetical protein	ZP_00573606.1	*Frankia sp. EAN1pec*	4,E+00
DW248599	Unnamed protein product	CAG00062.1	*Tetraodon nigroviridis*	4,E+00
DW248340	Hypothetical protein alr1263	NP_485306.1	*Nostoc sp.*	4,E+00
DW247999,DW248000,DW248449	Kif24 protein	AAH67395.1	*Mus musculus*	5,E+00
DW248241,DW248455	Hydrolase, haloacid dehalogenase-like family	YP_037943.1	*Bacillus thuringiensis*	5,E+00
DW248372	MADS-box transcription factor, putative	EAL65726.1	*Dictyostelium discoideum*	5,E+00

DW247854	Hypothetical protein BSU30390	NP_390917.1	*Bacillus subtilis*	5,E+00
DW247765,DW247790,DW248081	Predicted protein	EAA50518.1	*Magnaporthe grisea*	5,E+00
DW248377	Hypothetical protein PC106132.00.0	CAH79475.1	*Plasmodium chabaudi*	5,E+00
DW248042	Hypothetical protein PC302539.00.0	CAH87592.1	*Plasmodium chabaudi*	6,E+00
DW248253	Hypothetical protein DDB0184540	XP_629480.1	*Dictyostelium discoideum*	6,E+00
DW248457	Hypothetical protein STH1268	YP_075097.1	*Symbiobacterium thermophilum*	6,E+00
DW248346,DW248508,DW248553	Bh1907	C83888	*Bacillus halodurans*	6,E+00
DW247989,DW248358,DW248445	Ubiquitin carboxyl-terminal hydrolase, putative	NP_703615.1	*Plasmodium falciparum*	6,E+00
DW247879,DW248477	Unknown	AAW27053.1	*Schistosoma japonicum*	6,E+00
DW248345,DW248367,DW248615	Hypothetical protein AN1104.2	EAA66222.1	*Aspergillus nidulans*	6,E+00
DW247853,DW248164,DW248262,DW248416,DW248759	Phosphoribosylglycinamide formyltransferase 1	NP_929995.1	*Photorhabdus luminescens*	6,E+00
DW248127	Kinetoplast DNA-associated protein, putative	XP_808911.1	*Trypanosoma cruzi*	7,E+00
DW248729	Glp_149_5305_3950	EAA36721.1	*Giardia lamblia*	7,E+00
DW248443	Hypothetical protein CBG11204	CAE66012.1	*Caenorhabditis briggsae*	8,E+00
DW247808,DW248284	Hypothetical protein cao19.6050	XP_719012.1	*Candida albicans*	8,E+00
DW247866,DW248577	PREDICTED: similar to mkiaa0978 protein	XP_230750.2	*Rattus norvegicus*	8,E+00
DW248324	Ptprt	CAI19869.1	*Homo sapiens*	8,E+00
DW248016	Hypothetical protein Tb09.v1.0370	XP_827123.1	*Trypanosoma brucei*	9,E+00
DW248444	Unnamed protein product	CAG11692.1	*Tetraodon nigroviridis*	9,E+00
DW248754	ORF MSV009 leucine rich repeat gene family protein, similar to Amsacta moorei entomopoxvirus Q3 ORF SW:P28854	NP_048080.1	*Melanoplus sanguinipes*	9,E+00
DW247807,DW247983,DW248586	Hypothetical protein lpg0285	YP_094339.1	*Legionella pneumophila*	9,E+00
DW248185,DW248730	Hypothetical protein tll1570	NP_682360.1	*Thermosynechococcus*	1,E+01
DW248235	Phosphatase, putative	YP_013450.1	*Listeria monocytogenes*	1,E+01
DW247990,DW248070	No hit			
DW248223,DW248296,DW248310,DW248344	No hit			
DW247802,DW248183,DW248199,DW248448	No hit			
DW247836,DW248573	No hit			
DW248458,DW248471,DW248611	No hit			
DW247815,DW248485,DW248623	No hit			
DW248509,DW248652	No hit			
DW247935,DW248181,DW248655	No hit			
DW248244,DW248286,DW248461,DW248731	No hit			

DW247841,DW248033,DW248482,DW248560,DW248737	No hit			
DW248654,DW248742	No hit			
DW247873	No hit			
DW247894	No hit			
DW247940	No hit			
DW248132	No hit			
DW248180	No hit			
DW248220	No hit			
DW248275	No hit			
DW248298	No hit			
DW248418	No hit			
DW248475	No hit			
DW248507	No hit			
DW248539	No hit			
DW248566	No hit			
DW248678	No hit			