ONLINE RESOURCES

Expression screening and annotation of *Sepiella maindroni* ink sac cDNA library

SONG WEI-WEI[1,2]*, MU CHANG-KAO[1,2], SHAO CHU[1,2], ZHANG YU-YU[1,2], ZHAN PING-PING[1,2] and WANG CHUN-LIN[1,2]

[1]*Key Laboratory of Applied Marine Biotechnology, Ministry of Education, Ningbo University, Ningbo 315211, Zhejiang Province, People's Republic of China*

[2]*Faculty of Life Science and Biotechnology of Ningbo University, Ningbo 315211, People's Republic of China*

[Wei-Wei S., Chang-Kao M., Chu S., Yu-Yu Z., Ping-Ping Z. and Chun-Lin W. 2012 Expression screening and annotation of *Sepiella maindroni* ink sac cDNA library. *J. Genet.* **91**, e103–e108. Online only: http://www.ias.ac.in/jgenet/OnlineResources/91/e103.pdf]

Introduction

Sepiella maindroni is a high-valued cuttlefish, which is one of the most favorite Chinese seafoods. When stimulated, the cuttlefish squirts out ink which is of interest as it offers an excellent model for understanding melanin formation. The peculiar ink defence system of the cuttlefish *S. maindroni* is based on the activity of the highly specialized ink gland which continuously produces insoluble black melanin pigment that is stored in the ink sac (Palumbo *et al.* 1994; Palumbo 2003; Fiore *et al.* 2004).

Construction of cDNA library provides basic information for finding relevant genes and identifying their functions. A cDNA library can be used not only to screen the target genes required, but also to express the genes (Clifton and Mitreva 2009; Frentiu *et al.* 2009; Luo *et al.* 2010). Expressed sequence tags (ESTs) are necessary resources for gene discovery and for the development of polymorphic DNA markers, such as microsatellites and single nucleotide polymorphisms (SNP). (D'Agostino *et al.* 2007; Malde and Jonassen 2008). Significant EST resources have been developed for many aquaculture species including mollusks such as *Crassostrea gigas*, *C. virginica*, *Argopecten irradians*, *Tegillarca granosa* and *Pinctada martensii*. However, there is no information available for cephalopods in ESTs database. For organisms with poor genetic background, the ease and relative economy of partial sequencing of cDNA library can be a highly productive way to generate a great deal of information about genetic and functional genes (Rudd 2003; Bao and Lin 2010). The aim of this study was to create a cDNA library for the analysis of ESTs and to indentify genes related to melanogenesis.

*For correspondence. E-mail: songweiwei@nbu.edu.cn.

Keywords. expressed sequence tags; cDNA library; *Sepiella maindroni*.

Materials and methods

Healthy *S. maindroni* were collected from Liuheng (Zhoushan island, Zhejiang, China) and their ink sacs were frozen in liquid nitrogen for RNA extraction. Total RNA from ink sac was extracted using Trizol reagent (Invitrogen, Carlsbad, USA) according to the manufacturer's instruction. mRNA was obtained from the total RNA using Oligotex mRNA kits (Qiagen). The quality and concentration of mRNA samples were examined using EB-strained agarose gel electrophoresis and spectrophotometer analysis.

The Creator Smart cDNA Library Construction kit (Clontech, Mountain View, USA) was used for the construction of the cDNA library. Total RNA concentration was determined using a spectrophotometer and 1 μg RNA was used with 1 μL of SMART IV oligonucleotide (Clontech) and 1 μL of CD SIII/3′-PCR primer for first-strand cDNA synthesis. The first strand cDNA was initially amplified by long-distance PCR (LD-PCR) using hot-start amplification. Further the LD-PCR products were treated with proteinase K (Clontech). All the subsequent procedures including proteinase K treatment, restriction digestion with SfiI, size fractionation, and ligation were followed as per the manufacturer's instructions (Clontech). The cDNA was then ligated to the pBluescript II SK vector. The transformation of recombined plasmids was performed using DH-5α competent cells (Sangon, Shanghai, China) following the instructions.

Independent colonies were picked up and grown for 16 h at 37°C in 1.2 mL LB broth containing 30 μg/mL ampicillin Apr-IPTG/x-gal (Sangon). Plasmid DNA was isolated using Prep 96 Plasmid kit (Qiagen, Dusseldorf, Germany) from bacterial cultures. The plasmids were stored at −20°C until usage. The cDNA inserts were directionally sequenced from 5′-end of the cDNAs using 3130XL Genetic Analyzer (Applied Biosystems, Foster City, USA).

The ESTs were first processed to screen vector and linker sequences using DNA tools at the quality of QC10 (min. match 14, penalty 2 and min. score 30). The poly (A) tails were trimmed wherever necessary and sequences less than 100 bases were removed. Further the principle of 100 bp and 90% was used to merge the unigenes. All unique sequences were compared to the GenBank database using BLASTX and BLASTN in a nonredundant (NR) program (http://www.ncbi.nlm.nih.gov/BLAST). The *E* value scores less than 1.0×10^{-5} were considered to be significant. The BLAST search results were used to obtain further information on function, motif and relativity with other species through the databases on GenBank (http://www.ncbi.nlm.nih.gov), Expasy (http://www.expasy.org) and Ensemble (http://www.ensemble.org). Gene ontology (GO) annotations were assigned using the program GoPipe. (http://www.geneontology.org). Cluster of orthologous groups of proteins (COG) categories were also processed (http://www.ncbi.nlm.nih.gov/COG/).

Table 1. Similarity analysis of the 560 unique EST sequences based on the results from Blastx against NCBI's Nr database.

Similarity	Number	Percentage (%)
$10^{-150} < E \leq 10^{-100}$	14	2.5
$10^{-100} < E \leq 10^{-50}$	135	24.11
$10^{-50} < E \leq 10^{-20}$	144	25.71
$10^{-20} < E \leq 10^{-5}$	75	13.39
Total meaningful match ($E \leq 10^{-5}$)	368	65.71
Less meaningful match ($E > 10^{-5}$)	88	15.71
No match	104	18.58
Total	560	100

Results and discussion

The quality of the cDNA libraries for an EST sequencing project is a major consideration. Usually, the quality is examined through several factors such as redundancy and complexity of transcriptomes, frequencies of nonoverlapping genes, the number of full-length clones, internal priming and chimerical clones, and the level of contamination. These factors depend on the method of constructing the library (Perez *et al.* 2005). We created the cDNA library using the SMART cDNA library technique, which made a high-quality cDNA library. It obtained 560 unique genes from 979 ESTs, corresponding to a discovery rate of 57.2%. Cluster analysis of these ESTs identified 560 unique sequences containing 84 contigs and 476 singlets (table 1). The high percentage of new genes without significant GenBank hits is common for EST collections in marine organisms.

With the majority of *S. maindroni* genes having no significant homology to known genes, it is difficult to understand the functional significance of the transcriptome. Nevertheless, some genes are identified and their putative function can be inferred. It is not surprising to see that most of the identified genes are related to cellular components, metabolic pathways, and protein binding. (table 2; figure 1).

Some of the genes are involved in ink defence system. Sixty-eight unigenes (14.29%) were similar to *Euprymna scolopes* which were used to code a family of unusual proteins named reflectins, structural platelets in reflective tissues of the squid (table 3). These proteins are encoded by at least six genes in three subfamilies and have no reported homologous outside of squids. The composition of these proteins are unusual, with four relatively rare residues (tyrosine, methionine, arginine and tryptophan) comprising >57% of a reflectin. These protein-based reflectors in squids provide

Table 2. KEGG classifications of ink sac unigenes.

Category	No. of unigenes	Percentage (%)	No. of KEGG pathways mapped
Metabolism		64.29	
Carbohydrate metabolism	18		16
Energy metabolism	5		4
Lipid metabolism	2		2
Nucleotide metabolism	1		1
Amino acid metabolism	17		1
Glycan biosynthesis and metabolism	1		1
Xenobiotics biodegradation and metabolism	1		1
Genetic information processing		8.57	
Translation	3		3
Folding, sorting and degration	3		3
Cellular processes		25.71	
Cell motility	12		1
Cell growth and death	2		1
Cell communication	3		3
Immune system	1		3
Human disease		1.43	
Immune disorders	1		1

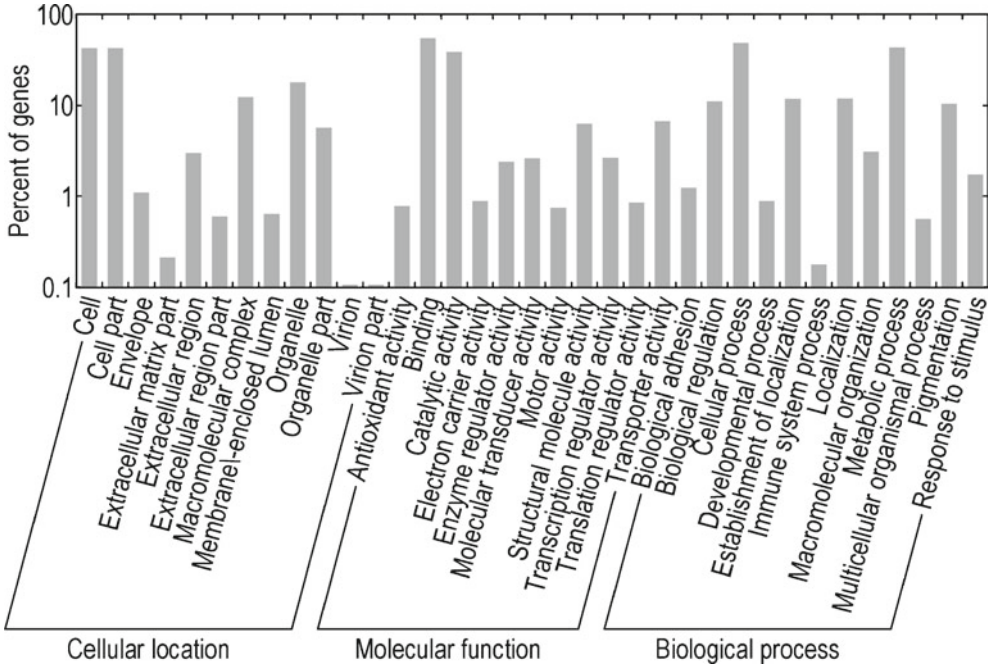

Figure 1. The gene ontology categories of genes from the *Sepiella maindroni* cDNA library according to molecular function, biological process and cellular components. The matched unigenes from the cDNA library are classified.

a marked example of nanofabrication in animal systems (Crookes *et al.* 2004). As previously reported, the tyrosine (Y) is the main raw material of ink pigment, and this protein is highly expressed in ink sac which was not a reflector, reflectins were related to the ink production. Arginine kinase involved in melanogensis was also highly expressed in this library (Palumbo *et al.* 2000; Palumbo 2003; Fiore *et al.* 2004; Di Cristo *et al.* 2007). Arginine kinase catalyzes the reversible transfer of a phosphoryl group and plays an important role in cellular energy metabolism in invertebrates.

The peculiar ink defence system of the cuttlefish *Sepiella maindroni* is based on the activity of the highly specialized ink glands which continuously produces black insoluble melanin pigment that is stored in the ink sac. COG categories results showed that abundance of ESTs were classified into these major functional categories, cytoskeleton, energy production and conversion, and signal transduction mechanisms (table 4). Actins participate in many important cellular processes including muscle contraction, cell motility, cell division and cytokinesis, vesicle and organelle movement, cell signalling, and the establishment and maintenance of cell junctions and cell shape. PDZ/LIM protein plays very important and diverse biological roles. They have been implicated in numerous vital processes, such as cytoskeleton organization, neuronal signalling, cell lineage specification, organ development, and oncogenesis. Functionally, all PDZ and LIM domain proteins can associate with and/or influence the actin cytoskeleton. The rootlet is found in diverse organisms across wide phylogenetic distances, from green algae, mollusks, amphibians to mammals. Rootletin is the long-sought structural component of the ciliary rootlet. Transgelin-2 (SM22) is a transformation and shape-change sensitive actin cross-linking/gelling protein found in fibroblasts and smooth muscle.

Table 3. The comparative with other species (Blastx).

Species	No. of unigenes	Percentage (%)	*E* value $\leq 10^{-5}$
Euprymna scolopes	68	14.29	68
Sepioteuthis lessoniana	59	12.39	59
Branchiostoma floridae	42	8.82	42
Amblyomma americanum	5	1.05	5
Lycosa singoriensis	5	1.05	5
Xenopus tropicalis	4	0.84	4
Tetraodon nigroviridis	4	0.84	4
Sepia officinalis	3	0.63	3
Loligo pealei	2	0.42	2
Drosophila	2	0.42	2
Others	282	59.24	194

The majority of the unique EST sequences had similarities to known genes, making them more significant for functional analysis. The EST sequences should enhance the effectiveness of molecular studies, especially for expression profiling and the analysis of genes involved in development and reproduction (Meng *et al.* 2010; Priyanka *et al.* 2010; Sathiyamoorthy *et al.* 2010). The EST information could be used to design and microarray, and analyse the proteomics

Table 4. Classification of ink sac unique sequences according to COG and gene annotation results.

Gene annotation	ESTs	GI number	Reference organism	*E* value
Amino acid transport and metabolism				
Argininosuccinate synthase	1	47204754	*Xenopus tropicalis*	2.00E-47
Glutamine synthetase	1	57283127	*Panulirus argus*	4.00E-47
Selenide, water dikinase	2	112983862	*Drosophila melanogaster*	3.00E-76
Spermidine synthase	1	210108597	*Mus musculus*	9.00E-63
Tryptophan synthase	1	15239755	*Arabidopsis thaliana*	7.00E-37
Carbohydrate transport and metabolism				
Alpha-adducin	2	30959104	*Rattus norvegicus*	5.00E-30
Fructose-1,6-bisphosphatase	2	196001309	*M. musculus*	5.00E-47
Fructose-bisphosphate aldolase	6	126697434	*Echinococcus multilocularis*	4.00E-51
Glyceraldehyde-3-phosphate dehydrogenase	8	124264155	*Coturnix coturnix*	6.00E-62
Probable pectate lyase 4	1	7019647	*A. thaliana*	3.00E-25
Probable transaldolase	1	114052613	*D. melanogaste*	4.00E-70
Pyruvate kinase isozymes M1/M2	3	149612407	*M. musculus*	4.00E-83
Triosephosphate isomerase B	1	47271422	*Danio rerio*	7.00E-80
Uncharacterized protein C12orf5	1	187608286	*Homo sapiens*	4.00E-24
Cell cycle control, cell division, chromosome partitioning				
Microtubule-associated protein	1	47271390	*R. norvegicus*	6.00E-38
Cell wall/membrane/envelope biogenesis				
UDP-glucose 6-dehydrogenase	1	158296008	*D. melanogaster*	4.00E-52
Chromatin structure and dynamics				
Histone H2	1	210122937	*Strongylocentrotus purpuratus*	1.00E-50
Global transcription activator	1	194388750	*H. sapiens*	4.00E-44
Coenzyme transport and metabolism				
NADH-cytochrome b5 reductase 3	1	210121645	*H. sapiens*	2.00E-36
Cytoskeleton				
Actin	44	157143002	*Placopecten magellanicus*	1.00E-113
Myosin heavy chain	1	189007784	*Aequipecten irradians*	4.00E-65
Tubulin	4	158287825	*D. melanogaster*	2.00E-78
Energy production and conversion				
ATP synthase	4	210127593	*Gallus gallus*	4.00E-59
Cytochrome c14	1	82658182	*M. musculus*	8.00E-65
Isocitrate dehydrogenase (NAD)	1	48097748	*M. musculus*	1.00E-71
Malate dehydrogenase	1	14583133	*H. sapiens*	1.00E-54
Omega-crystallin	1	210124320	*Octopus dofleini*	5.00E-54
Pyruvate dehydrogenase E1 component	3	170062187	*Caenorhabditis elegans*	2.00E-29
Vacuolar ATP synthase	2	209154312	*R. norvegicus*	2.00E-69
General function prediction only				
ATP-binding cassette sub-family	1	57529531	*Bos taurus*	2.00E-36
Eukaryotic translation initiation factor 3	1	55742565	*H. sapiens*	7.00E-41
Guanine nucleotide-binding protein	1	121014	*Loligo forbesi*	6.00E-55
Ribonucleoprotein A1	4	170586940	*D. melanogaster*	2.00E-78
Palmitoyltransferase ZDHHC3	1	210094007	*H. sapiens*	9.00E-64
Ras-like GTP-binding protein RHO	1	132545	*Aplysia californica*	5.00E-49
Serine/threonine-protein kinase N2	1	126305915	*H. sapiens*	3.00E-90
Serine-threonine kinase receptor-associated protein	1	126339924	*M. musculus*	5.00E-38
Small glutamine	1	215505493	*B. taurus*	9.00E-46
Splicing factor, arginine/serine-rich	1	215509877	*G. gallus*	5.00E-43
Target of rapamycin complex subunit LST8	2	149052019	*M. musculus*	6.00E-26
TBC1 domain family member 20	3	210116716	*H. sapiens*	3.00E-17
tRNA selenocysteine	1	210086345	*R. norvegicus*	1.00E-14
Zinc-binding alcohol dehydrogenase	1	210082462	*H. sapiens*	3.00E-50
Inorganic ion transport and metabolism				
Calcium-transporting ATPase	2	152003987	*Artemia sanfranciscana*	1.00E-109
Sodium/potassium-transporting ATPase	1	134141898	*Ovis aries*	4.00E-70
Intracellular trafficking, secretion, and vesicular transport				
ADP-ribosylation factor GTPaseactivating protein	1	215496398	*Pongo abelii*	3.00E-56
Receptor expression-enhancing protein	2	109078228	*B. taurus*	5.00E-54
Signal peptidase complex catalytic subunit SEC11A	1	210118327	*P. abelii*	7.00E-49
Lipid transport and metabolism				
Retinol dehydrogenase	1	118464304	*M. musculus*	2.00E-20

Table 4 (*contd*.)

Gene annotation	ESTs	GI number	Reference organism	*E* value
Nucleotide transport and metabolism				
Ribonucleoside-diphosphate reductase	1	212511324	*Spisula solidissima*	4.00E-32
Posttranslational modification, protein turnover, chaperones				
26S proteasome non-ATPase regulatory subunit 11	1	72057611	*M. musculus*	1.00E-38
Glutathione peroxidase 2	3	169403976	*P. pygmaeus*	6.00E-36
Methionine-R-sulfoxide reductase B3	1	194212345	*P. abelii*	1.00E-65
SUMO-conjugating enzyme	1	66519638	*X. tropicalis*	5.00E-21
Ubiquitin carboxyl-terminal hydrolase	1	47230636	*H. sapiens*	5.00E-36
Replication, recombination and repair				
Eukaryotic initiation factor 4A-II	1	215492155	*Macaca fascicularis*	7.00E-44
RNA processing and modification				
Poly(A) polymerase	1	118091978	*M. musculus*	2.00E-65
Secondary metabolites biosynthesis, transport and catabolism				
Protein usf	1	156390586	*Aquifex pyrophilus*	9.00E-07
Signal transduction mechanisms				
14-3-3 protein epsilon	5	193673848	*D. melanogaster*	1.00E-15
Myosin catalytic light chain LC-1	7	149242702	*Todarodes pacificus*	8.00E-49
Serine/threonine-protein phosphatase	2	157104222	*X. tropicalis*	4.00E-50
Transcription				
RuvB-like 2	1	115968520	*X. laevis*	9.00E-46
Y-box factor	4	1175568	*Aplysia californica*	2.00E-56
Translation, ribosomal structure and biogenesis				
60S acidic ribosomal protein	2	158120678	*G. gallus*	9.00E-56
60S ribosomal protein	1	22203722	*P. abelii*	4.00E-61
Elongation factor	1	13111524	*Caenorhabditis elegans*	8.00E-58
Eukaryotic translation initiation factor	2	210101423	*X. tropicalis*	1.00E-66
Phenylalanyl-tRNA synthetase	1	213511304	*P. abelii*	1.00E-33
Function unknown				
Flotillin-1	1	212515491	*D. melanogaster*	3.00E-42
Flotillin-2a	1	94039390	*Danio rerio*	5.00E-41
ORM1-like protein	2	147900385	*X. laevis*	2.00E-62

data. Additionally, the cluster and redundancy information should be useful for further subtraction of the most abundant transcripts included in the cDNA library, further making EST analysis in ink sac more effective.

In summary, this study described the construction of cDNA library and partial genes expression profile of *S. maindroni*. The results from sequencing and analysis of 979 unigenes are powerful means for identifying novel genes and for gene expression profiling in this secretory. Although generation of ESTs and sequence analysis is the first step to understand the gene expression, the cDNA library can provide a framework for understanding the ink production, regulation and metabolic functions.

Acknowledgements

The authors are grateful to all the laboratory members for technical advice and helpful discussion. This research was supported by National Natural Science Foundation of China (40776076), Innovative Research Team in University (PCSIRT) (IRT0734), Major Programme of Science and Technology Planning Project of Zhejiang Province (2009C03017-3, 2007C12061), Transformation of Agriculture funds (2009GB2C220415) and KC Wong Magna Fund in Ningbo University.

References

Bao Y. and Lin Z. 2010 Generation, annotation, and analysis of ESTs from hemocyte of the bloody clam, *Tegillarca granosa*. *Fish Shellfish Immunol.* **29**, 740–746.

Clifton S. W. and Mitreva M. 2009 Strategies for undertaking expressed sequence tag (EST) projects. *Methods Mol. Biol.* **533**, 13–32.

Crookes W. J., Ding L. L., Huang Q. L., Kimbell J. R., Horwitz J. and McFall-Ngai M. J. 2004 Reflectins: the unusual proteins of squid reflective tissues. *Science* **303**, 235–238.

D'Agostino N., Traini A., Frusciante L. and Chiusano M. L. 2007 Gene models from ESTs (GeneModelEST): an application on the Solanum lycopersicum genome. *BMC Bioinformatics* **8**, suppl 1, 9.

Di Cristo C., Fiore G., Scheinker V., Enikolopov G., d'Ischia M., Palumbo A. and Di Cosmo A. 2007 Nitric oxide synthase expression in the central nervous system of Sepia officinalis: an in situ hybridization study. *Eur. J. Neurosci.* **26**, 1599–1610.

Fiore G., Poli A., Di Cosmo A., d'Ischia M. and Palumbo A. 2004 Dopamine in the ink defence system of Sepia officinalis: biosynthesis, vesicular compartmentation in mature ink gland cells, nitric oxide (NO)/cGMP-induced depletion and fate in secreted ink. *Biochem. J.* **378**, 785–791.

Frentiu F. D., Adamski M., McGraw E. A., Blows M. W. and Chenoweth S. F. 2009 An expressed sequence tag (EST) library for Drosophila serrata, a model system for sexual selection and climatic adaptation studies. *BMC Genomics* **10**, 40.

Luo H., Sun C., Song J., Lan J., Li Y., Li X. and Chen S. 2010 Generation and analysis of expressed sequence tags from a cDNA library of the fruiting body of Ganoderma lucidum. *Chin. Med.* **5**, 9.

Malde K. and Jonassen I. 2008 Repeats and EST analysis for new organisms. *BMC Genomics* **9**, 23.

Meng X., Chang Y., Qiu X. and Wang X. 2010 Generation and analysis of expressed sequence tags from adductor muscle of Japanese scallop Mizuhopecten yessoensis. *Comp. Biochem. Physiol. Part* **D5**, 288–294.

Palumbo A. 2003 Melanogenesis in the ink gland of Sepia officinalis. *Pigment. Cell Res.* **16**, 517–522.

Palumbo A., d'Ischia M., Misuraca G., De Martino L. and Prota G. 1994 A new dopachrome-rearranging enzyme from the ejected ink of the cuttlefish *Sepia officinalis*. *Biochem. J.* **299**, 839–844.

Palumbo A., Poli A., Di Cosmo A. and d'Ischia M. 2000 N-Methyl-D-aspartate receptor stimulation activates tyrosinase and promotes melanin synthesis in the ink gland of the cuttlefish Sepia officinalis through the nitric Oxide/cGMP signal transduction pathway. A novel possible role for glutamate as physiologic activator of melanogenesis. *J. Biol. Chem.* **275**, 16885–16890.

Perez F., Ortiz J., Zhinaula M., Gonzabay C., Calderon J. and Volckaert F. A. 2005 Development of EST-SSR markers by data mining in three species of shrimp: Litopenaeus vannamei, Litopenaeus stylirostris, and Trachypenaeus birdy. *Mar. Biotechnol.* **7**, 554–569.

Priyanka B., Sekhar K., Sunita T., Reddy V. D. and Rao K. V. 2010 Characterization of expressed sequence tags (ESTs) of pigeonpea (*Cajanus cajan* L.) and functional validation of selected genes for abiotic stress tolerance in Arabidopsis thaliana. *Mol. Genet. Genomics* **283**, 273–287.

Rudd S. 2003 Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci.* **8**, 321–329.

Sathiyamoorthy S., In J. G., Gayathri S., Kim Y. J. and Yang D. C. 2010 Generation and gene ontology based analysis of expressed sequence tags (EST) from a Panax ginseng C. A. Meyer roots. *Mol. Biol. Rep.* **37**, 3465–3472.

Received 6 June 2012; accepted 1 August 2012
Published on the Web: 28 November 2012