



A transcriptome analysis of mitten crab testes (*Eriocheir sinensis*)

Wei Zhang^{1*}, Haolei Wan^{2,3*}, Hui Jiang^{1*}, Yunlong Zhao¹, Xiaowei Zhang², Songnian Hu² and Qun Wang¹

¹School of Life Science, East China Normal University, Shanghai, China.

²Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

³Graduated University of the Chinese Academy of Sciences, Beijing, China.

Abstract

The identification of expressed genes involved in sexual precocity of the mitten crab (*Eriocheir sinensis*) is critical for a better understanding of its reproductive development. To this end, we constructed a cDNA library from the rapid developmental stage of testis of *E. sinensis* and sequenced 3,388 randomly picked clones. After processing, 2,990 high-quality expressed sequence tags (ESTs) were clustered into 2,415 unigenes including 307 contigs and 2,108 singlets, which were then compared to the NCBI non-redundant (nr) protein and nucleotide (nt) database for annotation with Blastx and Blastn, respectively. After further analysis, 922 unigenes were obtained with concrete annotations and 30 unigenes were found to have functions possibly related to the process of reproduction in male crabs – six transcripts relevant to spermatogenesis (especially Cyclin K and RecA homolog DMC1), two transcripts involved in nuclear protein transformation, two heat-shock protein genes, eleven transcription factor genes (a series of zinc-finger proteins), and nine cytoskeleton protein-related genes. Our results, besides providing valuable information related to crustacean reproduction, can also serve as a base for future studies of reproductive and developmental biology.

Key words: reproduction, testis, EST, *Eriocheir sinensis*.

Received: April 16, 2010; Accepted: June 15, 2010.

The mitten crab (*Eriocheir sinensis*) (Henri Milne Edwards, 1854) is one of the most important aquaculture species in China, its culture under facility conditions having started in the early 1980's (Li *et al.*, 2007). The annual output has increased during the past decade in China, from 200,000 tons in the year 2000 to 420,000 tons in 2004 (Chen *et al.*, 2007). With the development of intensive culture, various problems have appeared in cultured populations, amongst others, sexual precocity. More and more individual crabs mature when small-sized. After sexual maturation, energy and nutrients are mainly diverted into gonad maturation or reproduction, with little left over for somatic growth, with the consequential devaluation of the commercial product and economic losses.

Previous studies (Chen *et al.*, 2003; Zhao and Lu, 2003; Li *et al.*, 2005) have revealed two main external factors as leading to sexual precocity in crabs, the environment and food. Nevertheless, few scientific investigations have been dedicated to elucidating the internal factors inducing sexual precocity. Furthermore, the regulative mechanisms of gonad maturation in *E. sinensis* at the molecular level are

unknown. The initial steps towards an understanding of molecular mechanism of gonad maturation in *E. sinensis* should be the identification of the respective reproduction-related gonadal transcripts (Klinbunga *et al.*, 2006; Preechaphol *et al.*, 2007). The animal testis is functionally important, both in reproduction and the secretion of hormones for growth and developmental regulation. The understanding of the molecular mechanism underlying testis development in the mitten crab is crucial to control testis maturation. Consequently, it is essential to discover reproduction-related transcripts in a testis cDNA library of *E. sinensis*, in order to reveal the corresponding molecular mechanisms of testis maturation and spermatogenesis.

Expressed sequence tag (EST) analysis is a powerful approach for discovering new transcripts and analyzing gene expression profiles in specific tissues or cells (Gieser and Swaroop, 1992; Tassanakajon *et al.*, 2006; Gai *et al.*, 2009; Hou *et al.*, 2010). Such an approach could be of aid in understanding the biological functions of testis tissues at the transcriptome level. The main objectives of the present study were (1) to discover transcripts potentially related to reproduction and development in *E. sinensis* by constructing a testis cDNA library and through EST analysis, and (2) to sequence the library in order to provide useful, pertinent genomic information.

Send correspondence to Qun Wang, School of Life Science, East China Normal University, 3663, North Zhong Shan Road, 200062 Shanghai, China. E-mail: qun_300@hotmail.com.

*These authors contributed equally to this work.

Healthy male mitten crabs (100-120 g) in the early stages of reproduction – the period of rapid testis development – were obtained from a commercial crab farm near Shanghai, China in August. The crabs were placed in an ice-bath for 1-2 min until they were lightly anesthetized. Through dissection, all the testes were collected and immediately frozen in liquid nitrogen and stored at 80 °C for future use.

Total RNA was isolated with a Unizol Reagent (Biostar, Shanghai, China) and mRNA purified by using Oligotex mRNA Kits (Qiagen). First-strand and double-strand cDNA were synthesized separately by way of Superscript reverse transcriptase (Invitrogen) and DNA polymerase I (Promega). First-strand synthesis was carried out separately with oligo-dT (with XhoI-linker sequence) and random primers (50 µM) in equal amounts. In the case of oligo-dT-primed cDNA, the second strand was synthesized and linked to an *EcoRI*-linker. Products of second-strand synthesis were separated on a 1% agarose gel, and cDNAs longer than 500 bp were isolated and extracted. Size fractions of cDNA were then ligated into a pBluescript SK+ vector (Stratagene) using T4 DNA ligase (Promega). Plasmids were transformed into *Escherichia coli* (DH10B) cells (Invitrogen), and grown overnight on solid LB medium containing IPTG (200 mg/mL) and X-Gal (20 mg/mL). Colony selection was based on blue/white (LacZ) staining. Finally, over 3,000 individual clones were randomly picked from the library and sequenced from the 5' end by using a T3 universal sequencing primer. The dideoxy-dye-terminator method was used for capillary sequencing on a 3730 XL system (Applied Biosystems).

Phred analysis (Ewing *et al.*, 1998) was applied for determining DNA bases. Low quality sequences were omitted from further analysis. Cross-match was used to remove vectors and *E. coli* DNA sequences from insert sequences, and all sequences shorter than 100 bp were removed by way of scripts programmed in Perl language. High-quality ESTs were assembled into contigs using Phrap software. All the unigenes were compared against the GenBank (NCBI) non-redundant protein (nr) and nucleotide (nr/nt) databases through Blastx and Blastn, respectively (E-value < 1.010⁻⁵). By analyzing all unigenes for their functional characteristics using Gene Ontology (GO) (Harris *et al.*, 2004), we annotated the unigenes by searching (Blastx) the updated Universal Protein Resource (Uniprot) database. The sum of unigenes within the subcategories of every major category may exceed 100% because some transcripts were classified into more than one subcategory in each of the three major categories. Furthermore, all unigenes with particular annotations were matched to the Kyoto Encyclopedia of Genes and Genomes (KEGG) online, in order to forecast their functions and biochemical pathways.

3,388 clones were randomly picked from the library and sequenced from the 5' end. 2,990 high-quality ESTs

were longer than 100 bp, with an average length of 598 bp. Of these (GenBank Accession no. GE339624-GE342613), 882 (29.5%) were clustered into 307 contigs, while 2,108 (70.5%) remained singlets (Table 1). Of the 2,415 unigenes (307 contigs and 2,108 singlets), 922 were annotated by comparison with the NCBI nr and nt database. The remaining 1,493 were not definitively identified.

By GO comparison of gene expression profiles of mitten crab testis tissue, and based on three functional categories (Figure 1), 455 unigenes of the total 2,415 (38.8%) were categorized. Within the category “Biological process”, the subcategories “cellular metabolic process”, “primary metabolic process” and “macromolecule metabolic process” contained the highest number of unigenes, followed by several metabolisms related to testis cell-cycle. Our analyses indicated that mitten crab testis cells are rapidly growing and undergo active metabolism, consistent with the energy requirements of spermatogenesis. The in-depth analysis of GO results revealed that most cDNA library unigenes were distributed among those subcategories encompassing growth and development and originating from the functional categories of “Molecular function” and “Biological process”; as well as the subcategories “binding”, “catalytic activity”, “biological regulation”, and “developmental process”, all of which representing cell-division and cell-death.

According to KEGG results, 314 unigenes were assigned to specific pathways, with 35.2% functioning in basic-metabolism processes, such as specific carbohydrate, energy, amino acid and nucleotide metabolisms. The remainder was involved in genetic information processing (36.3%), environmental information processing (8.9%), and cellular processes (15.3%), respectively.

In accordance with the purpose of the project, we were particularly interested in discovering transcripts in the reproduction process. According to our annotations, of all the 2,415 unigenes, 30 were found to have significant functions during the processes of spermatogenesis and sexual maturity. Based on previously published reports, these unigenes, comprising 101 ESTs, are identified as being functionally involved in these processes in crabs. According to

Table 1 - Overview of EST analysis of the testis cDNA library from *Eriocheir sinensis*.

Description	Number
Number of total sequences	3,388
Number of success sequences	2,990
Average length of success sequences (bp)	578
Number of Unigenes	2,415
Number of contigs	307
Number of singlets	2,108
Number of matched genes (%)	922 (38.2%)
Number of unmatched genes (%)	1493 (61.8%)

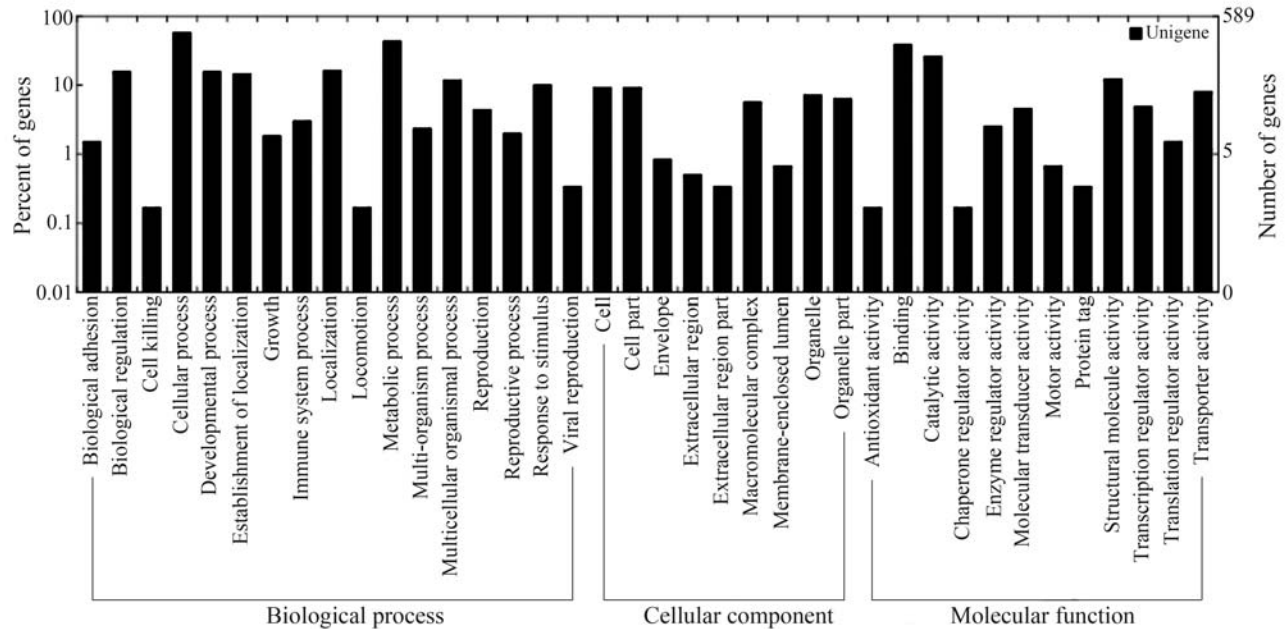


Figure 1 - Functional annotation of unigenes based on known proteins in the Uniprot database. Gene Ontology (GO) terms at the 2nd level were plotted here, and in this ontology, “Biological process”, “Cellular component”, and “Molecular function” are categorized independently.

their predicted functions, they were classified into five broad groups (Table 2).

Several transcripts related to spermatogenesis were identified in the library. There was a seminal plasma glycoprotein 120 in the testis cDNA library, the first time such a seminal plasma glycoprotein has been found in mitten crabs. Seminal plasma is a very complex fluid rich in many types of macromolecules involved in fertilization. Seminal plasma glycoprotein 120 may play a key role in mitten crab sperm capacitation. From the testis cDNA library, seven ESTs were annotated as related to ubiquitin. Ubiquitin is a small, highly-conserved regulatory protein, that is ubiquitously expressed in eukaryotes, its most prominent function being to label proteins for proteasomal degradation (Cook and Petrucelli, 2009). There were three heat-shock protein genes in the library from *E. sinensis*, including two heat-shock protein 70 and one heat-shock cognate protein 70. Hsp-70 has the characteristic of regulating initial meiosis. In mice the *Hsc70* gene is activated at the round spermatid stage, and regulated at both the transcriptional and translational levels during spermatogenesis (Matsumoto *et al.*, 1993). A total of nine transcripts were associated with Cytoskeleton Protein, which belong to a cytoskeleton gene family. These are reported to be involved in the spermatogenesis process in mice (Alsheimer and Benavente, 1996). In the library, two unigenes encoding β -actin and α -tubulin were identified. It has been reported that transcripts of the β -actin gene are more abundant in round spermatids than in pachytene spermatocytes, whereas those of the α -tubulin gene are widely distributed in testis germ cells, where their abundance influences the occurrence of spermatozoa abnormality (Hecht *et al.*, 1984). In the testis cDNA library,

eleven unigenes encoding zinc finger proteins were identified. The identification of *Zfp37* genes suggests that ZF genes are generally involved in the complex pathways from round spermatids to spermatozoa in mitten crab. Furthermore, it is also the first time *Zfp37* genes have been found to participate in the processes of spermatogenesis and sex maturity in crustaceans (Sepp and Choo, 2005; Cho *et al.*, 2008), although their specific function in sperm development needs to be further examined.

The unigene annotated to the Histone H1 gene, comprising 39 ESTs, is one of the most abundantly expressed in the testis. The representative characteristic of the spermatogenesis process is the gradual replacement from histone in somatocytes to testis-specific protein in spermatogonia (Meistrich *et al.*, 1985; Thomas *et al.*, 1989). Therefore, the observation of a highly expressed Histone H1 gene is taken to mean that our crabs were in an early phase of sexual maturity. Furthermore, four ESTs constructed one Arginine kinase unigene in the testis cDNA library. Arginine kinase is a member of what appears to be a highly conserved family of phosphotransferases that reversibly catalyze the transfer of phosphate from phosphoarginine to ADP (Wu *et al.*, 2008). Cyclin K is one kind of cyclin which participates in controlling the cell cycle by binding to Ser/Thr cyclin-dependent kinases, thereby regulating their activities (Yu *et al.*, 2008). The coordination between cell proliferation and differentiation is important in normal development. These processes usually have an antagonistic relationship, in that differentiation is blocked in proliferative cells, and terminally differentiated cells do not divide. In some instances, cyclins, cyclin-dependent kinases and their inhibitors play important roles in this antagonistic reg-

Table 2 - Reproduction-related genes in the testis cDNA library from *Eriocheir sinensis*.

Annotation	Matched species	Number of ESTs	E-value
Spermatogenesis related genes			
Seminal plasma glycoprotein 120	<i>Danio rerio</i>	1	2.00E-07
Arginine kinase	<i>Eriocheir sinensis</i>	4	7.00E-77
Cyclin K	<i>Apis mellifera</i>	1	5.00E-45
Ubiquitin	<i>Portunus pelagicus</i>	7	2.00E-67
Vasa-like protein	<i>Litopenaeus vannamei</i>	1	7.00E-25
RecA homolog DMC1	<i>Pacifastacus leniusculus</i>	3	6.00E-89
Nuclear protein transformation related-genes			
Histone H1	<i>Mytilus galloprovincialis</i>	39	6.00E-16
Histone H2A-IV	<i>Trichomonas vaginalis</i>	1	1.00E-17
Heat shock protein genes			
Heat shock protein 70	<i>Pachygrapsus marmoratus</i>	2	3.00E-96
Heat-shock cognate protein 70	<i>Bombyx mori</i>	1	2.00E-74
Transcription factor related genes			
Zinc finger protein	<i>Strongylocentrotus purpuratus</i>	1	1.00E-39
Zinc finger protein 160	<i>Xenopus tropicalis</i>	1	1.00E-21
Zinc finger protein 219	<i>Macaca mulatta</i>	1	1.00E-13
Zinc finger protein 33A	<i>Macaca mulatta</i>	1	4.00E-08
Zinc finger protein 33B	<i>Macaca mulatta</i>	1	3.00E-19
Zinc finger protein 36-like 3	<i>Macaca mulatta</i>	1	3.00E-21
Zinc finger protein 37a	<i>Equus caballus</i>	1	6.00E-42
Zinc finger protein 51	<i>Rattus norvegicus</i>	1	6.00E-13
Zinc finger protein 561	<i>Macaca mulatta</i>	1	1.00E-18
Zinc finger protein 658	<i>Canis familiaris</i>	1	1.00E-27
Zinc finger RNA binding protein	<i>Apis mellifera</i>	1	3.00E-36
Cytoskeleton protein genes			
Actin D	<i>Litopenaeus vannamei</i>	3	1.00E-126
Beta actin	<i>Pacifastacus leniusculus</i>	3	3.00E-71
Cytoplasmic actin A3a1	<i>Homalodisca coagulata</i>	1	4.00E-71
Alpha tubulin a1	<i>Mesenchytraeus solifugus</i>	1	4.00E-75
Alpha-Tubulin at 84B	<i>Xenopus laevis</i>	1	3.00E-97
Tubulin beta-1 chain	<i>Homarus americanus</i>	3	0
Tubulin beta-2 chain	<i>Homarus americanus</i>	3	1E-146
Tubulin alpha-3 chain	<i>Homarus americanus</i>	14	1.00E-139
Beta tubulin	<i>Phytophthora capsici</i>	1	1.00E-21

ulation (Chang *et al.*, 2007; Karagiannis and Balasubramanian, 2007). A cyclin K transcript, possibly involved in spermatogenesis, was identified in the library, as well as a Vasa-like protein which encodes an RNA helicase protein member of the DEAD-box family and plays a key role in spermatozoa formation, from bacteria to mammals (Ye *et al.*, 2007). Most DEAD-box proteins are essential for cell viability (Irion and Leptin, 1999). Vasa proteins are involved in gamete production as it is essential to the assembly of germ cell cytoplasm in many invertebrate and vertebrate species (Sellars *et al.*, 2007). A unique RecA

homolog DMC1 gene, a specifically expressed gene in meiosis and involved in spermatogenesis, was also discovered in the testis cDNA library. The protein encoded by this gene is essential for meiotic homologous recombination, and plays an important role in generating diverse genetic information (Sato *et al.*, 1995; Matsuda *et al.*, 1996). A unique RecA homolog DMC1 may be involved in mitten crab spermatogenesis, by playing catalytic and structural roles in interhomolog recombination during meiosis. In our project, considerable efforts were made to annotate genes, especially those related to reproduction, thus effectively

increasing the repertoire of crustacean genes en route of genomic studies.

Acknowledgments

This research was supported by grants from the National Nature Science Foundation of China (No.30671607), the Key Basic Research Project of Science and Technology Commission of the Shanghai Municipal Government (06DJ14003), the Scientific Research Foundation for Returned Overseas Chinese Scholars, the State, the Short-term Training Abroad Fund of ECNU, the Award for Doctoral Candidate Academic Newcomer of ECNU, and the PhD Program Scholarship Fund of ECNU (No.2010042).

References

- Alzheimer M and Benavente R (1996) Change of karyoskeleton during mammalian spermatogenesis: Expression pattern of nuclear lamin C2 and its regulation. *Exp Cell Res* 228:181-188.
- Chang EJ, Begum R, Chait BT and Gaasterland T (2007) Prediction of cyclin-dependent kinase phosphorylation substrates. *PLoS One* 2:e656.
- Chen D-W, Zhang M and Shrestha S (2007) Compositional characteristics and nutritional quality of Chinese mitten crab (*Eriocheir sinensis*). *Food Chem* 103:1343-1349.
- Chen Z, Cheng Y and Wang W (2003) Changes of hepatopancreas index, lipid content and fatty acid composition in *Eriocheir sinensis* during precocity. *J Fish China* 27:57-61.
- Cho SY, Chung M, Park M, Park S and Lee YS (2008) ZIFIBI: Prediction of DNA binding sites for zinc finger proteins. *Biochem Biophys Res Comm* 369:845-848.
- Cook C and Petrucelli L (2009) A critical evaluation of the ubiquitin-proteasome system in Parkinson's disease. *Biochim Biophys Acta* 1792:664-675.
- Ewing B, Hillier L, Wendl MC and Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175-185.
- Gai Y, Wang L, Zhao J, Qiu L, Song L, Li L, Mu C, Wang W, Wang M, Zhang Y, *et al.* (2009) The construction of a cDNA library enriched for immune genes and the analysis of 7535 ESTs from Chinese mitten crab *Eriocheir sinensis*. *Fish Shellfish Immunol* 27:684-694.
- Gieser L and Swaroop A (1992) Expressed sequence tags and chromosomal localization of cDNA clones from a subtracted retinal pigment epithelium library. *Genomics* 13:873-876.
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258-D261.
- Hecht NB, Kleene KC, Distel RJ and Silver LM (1984) The differential expression of the actins and tubulins during spermatogenesis in the mouse. *Exp Cell Res* 153:275-280.
- Hou X, Mao Q, Zhang W, Jia L and Wang Q (2010) Differentially expressed genes during accessory sex gland seasonal development in *Eriocheir sinensis*. *J Crust Biol* 30:93-100.
- Irion U and Leptin M (1999) Developmental and cell biological functions of the *Drosophila* DEAD-box protein abstract. *Curr Biol* 9:1373-1381.
- Karagiannis J and Balasubramanian MK (2007) A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p subunit. *PLoS One* 2:e433.
- Klinbunga S, Preechaphol R, Thumrunthanakit S, Leelatanawit R, Aoki T, Jarayabhand P and Menasveta P (2006) Genetic diversity of the giant tiger shrimp (*Penaeus monodon*) in Thailand revealed by PCR-SSCP of polymorphic EST-derived markers. *Biochem Genet* 44:222-236.
- Li X, Lei Y, Gao X, Ma C and Dong S (2007) Calcium carbonate supersaturation and precipitation in Chinese mitten crab (*Eriocheir japonica sinensis*) larval ponds in China: Mass mortality, crystal form analysis, and safety saturation index. *Aquaculture* 272:361-369.
- Li Y, Kang X, Zhao X, Wen X and LQ H (2005) Intrinsic factors of precocity in mitten-handed crab *Eriocheir sinensis*. *Fish Sci* 24:34-36.
- Matsuda Y, Habu T, Hori T and Morita T (1996) Chromosome mapping of the mouse homologue of DMC1, the yeast meiosis-specific homologous recombination gene. *Chromosome Res* 4:249-250.
- Matsumoto M, Kurata S, Fujimoto H and Hoshi M (1993) Haplotype specific activations of protamine 1 and hsc70t genes in mouse spermatogenesis. *Biochim Biophys Acta* 1174:274-278.
- Meistrich ML, Bucci LR, Trostle-Weige PK and Brock WA (1985) Histone variants in rat spermatogonia and primary spermatocytes. *Dev Biol* 112:230-240.
- Preechaphol R, Leelatanawit R, Sittikankeaw K, Klinbunga S, Khamnamtong B, Puanglarp N and Menasveta P (2007) Expressed sequence tag analysis for identification and characterization of sex-related genes in the giant tiger shrimp *Penaeus monodon*. *J Biochem Mol Biol* 40:501-510.
- Sato S, Seki N, Hotta Y and Tabata S (1995) Expression profiles of a human gene identified as a structural homologue of meiosis-specific recA-like genes. *DNA Res* 2:183-186.
- Sellars MJ, Lyons RE, Grewe PM, Vuocolo T, Leeton L, Coman GJ, Degnan BM and Preston NP (2007) A PL10 vasa-like gene in the kuruma shrimp, *Marsupenaeus japonicus*, expressed during development and in adult gonad. *Mar Biotech* 9:377-387.
- Sepp A and Choo Y (2005) Cell-free selection of zinc finger DNA-binding proteins using *in vitro* compartmentalization. *J Mol Biol* 354:212-219.
- Tassanakajon A, Klinbunga S, Paunglarp N, Rimphanitchayakit V, Udomkit A, Jitrapakdee S, Sritunyalucksana K, Phongdara A, Pongsomboon S, Supungul P, *et al.* (2006) *Penaeus monodon* gene discovery project: The generation of an EST collection and establishment of a database. *Gene* 384:104-112.
- Thomas KH, Wilkie TM, Tomashefsky P, Bellve AR and Simon MI (1989) Differential gene expression during mouse spermatogenesis. *Biol Reprod* 41:729-739.
- Wu Q-Y, Li F and Wang X-Y (2008) Evidence that amino-acid residues are responsible for substrate synergism of locust arginine kinase. *Insect Biochem Mol Biol* 38:59-65.
- Ye D, Lv D, Song P, Peng M, Chen Y, Guo M, Yang Q and Hu Y (2007) Cloning and characterization of a rice field eel vasa-like gene cDNA and its expression in gonads during natural sex transformation. *Biochem Genet* 45:211-224.

Yu W, Ramakrishnan R, Wang Y, Chiang K, Sung TL and Rice AP (2008) Cyclin T1-dependent genes in activated CD4 T and macrophage cell lines appear enriched in HIV-1 co-factors. *PLoS One* 3:e3146.

Zhao W and Lu J (2003) The relationship between hormone biosynthesis of mandibular organ and precociousness in *Eriocheir sinensis*. *J Fish China* 27:289-294.

Internet Resources

Phrap software, <http://www.phrap.org/>.

Associate Editor: Klaus Hartfelder

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.