Vol. 11: 193–199, 2010
doi: 10.3354/ab00308

AQUATIC BIOLOGY
Aquat Biol

Published online December 21

Characterization of single nucleotide polymorphisms from expressed sequence tags of Chinese mitten crab *Eriocheir sinensis*

Hong Jun Li[1], Chong Bo He[1,*], Qing Yang[2], Zhong Guo Shan[1], Ke Fei Tan[1], Xiang Gang Gao[1]

[1]The Key Laboratory of Marine Fishery Molecular Biology of Liaoning Province, Liaoning Ocean and Fishery Science Institute, Dalian 116023, PR China

[2]National Marine Environmental Monitoring Center, Dalian 116023, PR China

ABSTRACT: The Chinese mitten crab *Eriocheir sinensis* is an important aquaculture species in Asia and causes considerable economic and ecological damage as an invasive species in North America and Europe. Until recently, molecular markers available for genetic analysis in this species were limited to a few microsatellite sequences used in population studies. Here we describe the discovery of 3191 single nucleotide polymorphisms (SNPs) from the alignment of expressed sequence tags (ESTs) in *E. sinensis*. The observed frequency of SNPs was estimated at 0.78 per 100 bp of contig sequences. C/T substitutions were frequent and accounted for 32.2% of all SNPs. A subset of these SNPs (n = 38) was selected for validation by allele-specific PCR with melting temperature (T_m)-shift primers based on their observed frequency in sequence data; 12 (31.6%) of these were polymorphic in a panel of 40 wild-caught crabs. Eight (21.1%) did not amplify any product, and 18 (47.4%) failed due to amplification failure of 1 allele-specific primer. A table of optimal codons was deduced from the analysis of the *E. sinensis* EST dataset, and the results implied that the gene expression levels and GC-content might play important roles in codon usage bias. Sixteen codons ending with a G or C base were defined as 'optimal codons,' which may provide useful information for predicting gene expression in crabs. These are the first SNPs developed for the Chinese mitten crab and will provide a useful complement to currently available genetic markers.

KEY WORDS: *Eriocheir sinensis* · Single nucleotide polymorphism · SNP · Codon bias usage · Allele-specific PCR · T_m-shift

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most abundant sequence variations encountered in a genome (Picoult-Newberg et al. 1999, Berger et al. 2001). Although moderately distributed in the human genome with 1 SNP every 500 to 1000 bp (Sachidanandam et al. 2001), they reach high densities in other organisms, such as in some insects (e.g. 1 SNP every 125 bp in the *Aedes* mosquito genome, Morlais & Severson 2003) or some mollusks (e.g. 1 SNP every 40 bp in the oyster genome, Curole & Hedgecock 2005). Within the coding regions, a SNP is either nonsynonymous and results in an amino acid sequence change, or it is synonymous and does not alter the amino acid sequence. Synonymous codons (those encoding the same amino acid) are not equally used. This phenomenon is called codon usage bias, which exists in a wide range of biological systems from prokaryote to eukaryote (Angellotti et al. 2007).

Because of their high frequency, codominant inheritance, amenability to high-throughout genotyping analysis, and lower mutation rate compared to microsatellites (Wang et al. 2008), SNPs have become mark-

*Corresponding author. Email: hechongbo@hotmail.com

ers of choice for population genetics and genomic mapping studies (Rafalski 2002, Li et al. 2010a, Zhang & Guo 2010). Generally, locus discovery and genotyping are involved in the development of SNP markers. The rapidly developing expressed sequence tag (EST) databases make it possible to identify functionally-associated SNPs (Li et al. 2009), which are especially useful in correlative analysis and marker-assisted breeding. More recently, several highly multiplexed and ultra-high-throughout SNP genotyping systems have become available (Kwok 2001, Matsuzaki et al. 2004). However, most of these systems remain expensive and require significant up-front instrument investment. For projects focusing on a limited number of SNPs in candidate gene studies or population genetics, melting temperature (T_m)-shift is a more simple and economical SNP genotyping method involving a single allele-specific PCR reaction followed by melting curve analysis (Wang et al. 2005).

The Chinese mitten crab *Eriocheir sinensis* is native to East Asia and predominantly lives in fresh water but migrates seawards to breed (Clark et al. 1998). It is considered to be one of the most important economic crustaceans for aquaculture in many countries (Ying et al. 2006, Gao et al. 2010). Research on this species has focused primarily on development of hatchery seed production techniques and grow-out modes, and a series of key techniques involving spawning, larval rearing, juvenile nursing, and grow-out systems have been established (Wang et al. 2006, Li et al. 2007). Despite the importance of crab culture, crab breeding is still in its infancy, especially in the application of molecular breeding technologies. Genetic markers are important for aquaculture genetics and breeding. They are essential for stock identification, pedigree analysis, and genomic mapping (Liu & Cordes 2004). Here we describe the detection and characteristics of 12 *E. sinensis* EST-associated SNPs by allele-specific PCR with T_m-shift primers. A codon usage table was deduced from the analysis of the EST database, using EST counts as a rough assessment of gene expression.

MATERIALS AND METHODS

Data mining for SNPs. In total, 16 960 Chinese mitten crab EST sequences were downloaded from GenBank (www.ncbi.nlm.nih.gov/). The EST database was aligned and assembled using CAP3 (Huang & Madan 1999) with default parameters. For each contig generated from the CAP3 assembly, BLASTX was conducted against the non-redundant protein (NR) database to assist functional identification of any related ESTs in different contigs. The program QualitySNP (Tang et al. 2006) was used to detect putative SNPs from the contig sequences. The program utilized the CAP3 output files as input to detect SNPs based on the base redundancy in the sequence alignments. The contigs that contained 4 or more sequences were identified for searching for candidate SNPs upon visual inspection. A single-base mutation that occurred in 2 or more ESTs and that was surrounded by good flanking sequences was identified as a candidate SNP for validation analysis.

Analysis of codon usage bias. In order to investigate codon usage bias, a table of optimal codons was deduced from the analysis of an EST dataset using EST counts as a rough assessment of gene expression (Duret & Mouchiroud 1999). Open reading frames (ORFs) were identified using EMBOSS GUI v1.12: getorf (http://bips.u-strasbg.fr/EMBOSS/) with minimum nucleotide size of 300 bp. In order to normalize codon usage within datasets of differing amino acid compositions, relative synonymous codon usage (RSCU) and GC-content of each ORF were calculated by online software Codonw (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html). RSCU values are calculated by dividing the observed codon usage by that expected when all codons for the same amino acid are used equally. RSCU values close to 1.0 indicate a lack of bias for the corresponding codon. A synonymous codon was inferred as 'optimal' when its RSCU was significantly higher in genes expressed at a high level (defined as contigs) than those expressed at a low level (defined as singletons).

Primer design and SNP genotyping. The T_m-shift method (Wang et al. 2005) was used to genotype SNPs. For each SNP locus, the primer set included a common reverse primer (CR) and 2 forward allele-specific primers (AS1 and AS2), with the 3′ terminal base of each specific primer matching one of the SNP alleles (Fig. 1A). The common primer was typically placed no more than 20 bp downstream from the SNP for favoring allele discrimination. GC tails of different length, 14 bases for 1 primer and 6 for the second, were added to each of the 2 allele-specific primers to discriminate melting curve of amplification products. As a rule, the long tail was attached to an allele-specific primer with higher T_m base (G or C) at its 3′ end, and the short tail to the other allele-specific primer with lower T_m base (A or T).

For SNP polymorphism analysis, 40 wild individuals of *Eriocheir sinensis* were randomly obtained from Panjin City, Liaoning Province, China. Genomic DNA was extracted from leg muscle following the standard phenol-chloroform method. Allele-specific PCR was carried out in a final volume of 25 µl containing 10 ng DNA, 1× PCR SYBR® Premix Ex *Taq*™ buffer (Takara), and 0.2 µM each of the 3 primers. The PCR program was as follows: initial denaturation at 95°C for 30 s, followed by 40 cycles of 3-step amplification profile of 5 s at 95°C for denaturation, 30 s at 60°C for annealing,

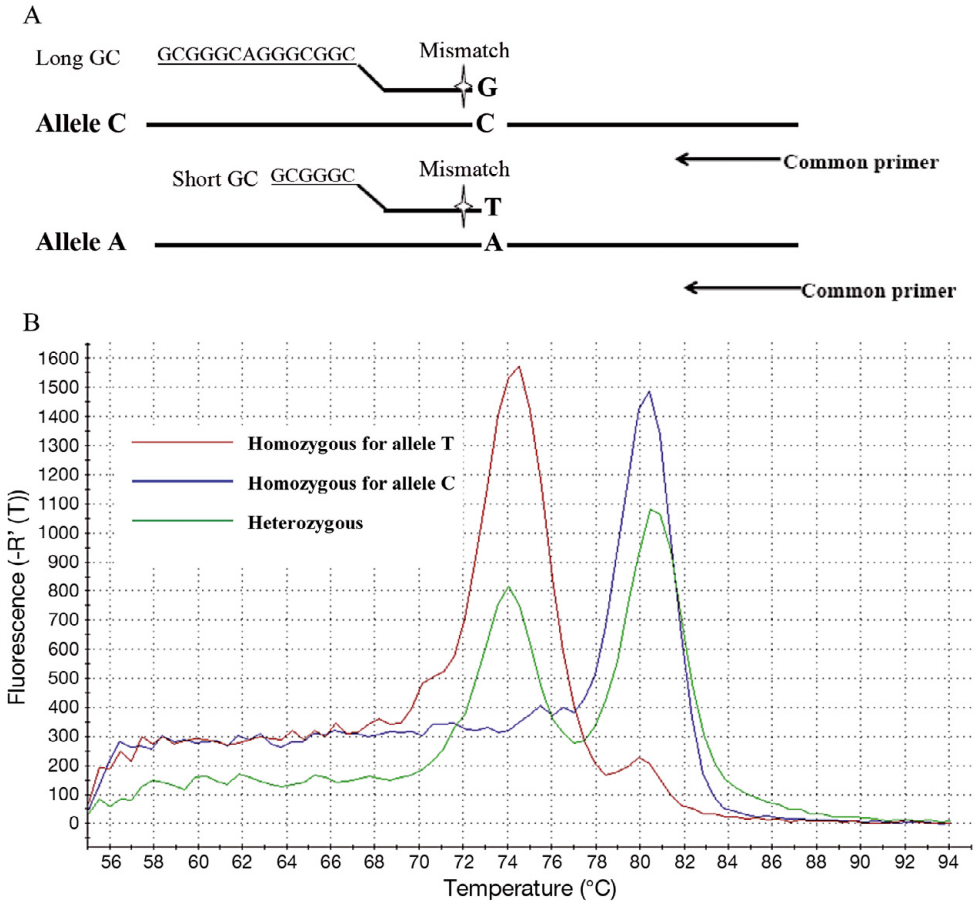

Fig. 1. *Eriocheir sinensis*. Melting temperature (T_m)-shift single nucleotide polymorphism (SNP) genotyping. (A) Schematic explanation of the primer design for the T_m-shift method. (B) Method validation and melting curve analysis. The melting curve of locus *Es1082* that were genotyped with the T_m-shift method are shown here

and 20 s at 72°C for extension. Melting curves were obtained using Stratagene Mx3000p real-time thermal cycler with the default 'dissociation step' to measure the fluorescence intensity of the PCR product in a linear denaturation ramp from 65 to 95°C. POPGENE 32 (Yeh et al. 2000) was used to calculate observed and expected heterozygosities (H_o and H_e) and to test departures from Hardy-Weinberg equilibrium (HWE).

RESULTS

Characteristics of putative SNPs

In total, 16 960 *Eriocheir sinensis* ESTs available from GenBank were subjected to cluster analysis to identify putative SNPs. The contig assembly resulted in 1330 contigs with an average size of 9.3 sequences contig^{-1} with an average length of 776 bp. Among 561 contigs with 4 or more ESTs, a total of 3191 SNPs were identified from 353 contigs, with an average of 0.78 SNPs per 100 bp (Table 1). A total of 1093 transversions (A/C, A/T, C/G, G/T) and 1485 transitions (A/G, C/T; Fig. 2) were detected, with C/T being the most common (829, 32.2%) and T/G the least common (212, 8.2%) variation observed.

SNP genotyping by allele-specific PCR with T_m-shift primer

Contigs that contained 4 or more sequences (n = 561) were identified for searching candidate SNPs for validation analysis by the software QualitySNP (Tang et al. 2006). According to the frequency of occurrence and conservation of flanking sequences, 38 candidate SNP loci were chosen for validation by allele-specific PCR with T_m-shift primers. When tested among 40 individuals of a wild population, 12 (31.6%) primer pairs were successful and showed bi-allelic polymorphisms. Eight (21.1%) did not amplify any product, and 18 (47.4%) failed due to amplification failure of one allele-specific primer. All polymorphic loci had 2 alleles, and the minor allele frequency ranged from 0.150 to 0.500. H_o

Table 1. *Eriocheir sinensis.* Summary of single nucleotide polymorphisms (SNPs) from the expressed sequence tag (EST) database

Parameter	Value
No. sequences for assembly	16960
No. putative transcripts	5954
No. singletons	4624
No. contigs	1330
2–3 ESTs	769
4–5 ESTs	193
6–10 ESTs	162
11–20 ESTs	103
>20 ESTs	103
Average contig size	9.3
Average contig length (bp)	776
Total consensus length (bp)	1031807
Number of contigs with SNPs	353
Number of total putative SNPs	3191
Transitions	1485
C↔T	829
A↔G	656
Transversions	1093
A↔T	329
A↔C	326
T↔G	212
C↔G	226
Other types[a]	172
Insertion/deletion	441
SNP frequency (bp)	128.4
Indel frequency (bp)	750.7

[a]Other types represent tri-, tetra-, or penta-allelic polymorphisms

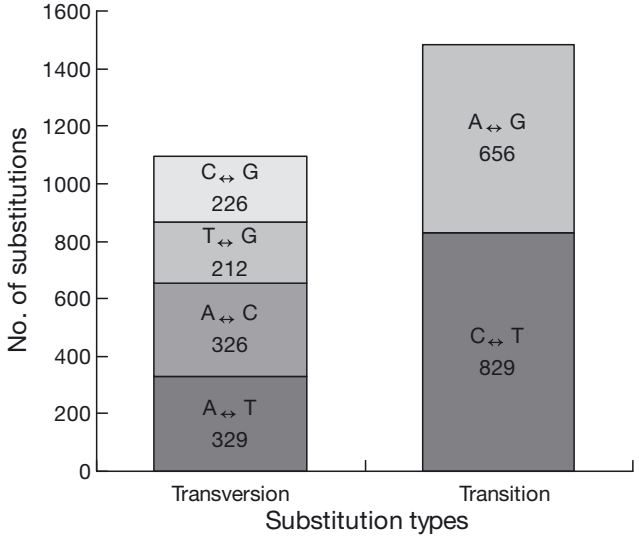

Fig. 2. *Eriocheir sinensis.* Distribution of substitution types of putative single nucleotide polymorphisms detected in the expressed sequence tags

and H_e ranged from 0.056 to 0.833 and 0.155 to 0.507, respectively (Table 2). In Fig. 1B, we show an example of T_m-shift genotyping assay for the locus *Es1082* based on allele-specific PCR.

Codon usage bias

To investigate the synonymous codon usage variation among genes of Chinese mitten crab, we used 8028 sequences containing ORFs longer than 100 codons (300 bp) to explore variation in RSCU. The cluster of the 8028 sequences produced 1203 contigs and 1643 singletons, which represented high-expression and low-expression genes, respectively. Table 3 reports the average RSCU values of all codons from *Eriocheir sinensis* genes; the optimal codons (those with positive ΔRSCU values) are highlighted. The average GC-content of codons was 56.6% for all genes, and of the 16 optimal codons that we inferred from ΔRSCU values, all were G or C ending codons, which suggested that genes with higher expression level were GC-rich and were preferred over the codons with G or C at the synonymous position.

DISCUSSION

EST analysis is not only the most efficient approach for gene discovery, but it is also an effective approach for the development of DNA markers (Li et al. 2009, 2010a,b). In the present study, we detected a total of 3191 putative SNPs in the *Eriocheir sinensis* EST database, and the average SNP frequency was estimated at 128.4 bp of contig sequences. This frequency is much higher than in the human genome with 1 SNP per kb (Frazer et al. 2007) and in Atlantic salmon *Salmo salar* with 1 per 641 bp (Hayes et al. 2007), but considerably lower than in Pacific oyster *Crassostrea gigas* with 1 per 60 bp in coding regions and 1 per 40 bp in non-coding regions (Sauvage et al. 2007). Considering the the strategy for identifying SNPs, where real SNPs may be ignored due to the number-limited ESTs, the SNP frequency might be higher in the genome of *E. sinensis*. Our research on this crab, together with a study on shrimp (Gorbach et al. 2009), perhaps implies a high SNP density in the genome of all crustaceans.

Many commercial SNP genotyping protocols have been developed using various assay methods based on hybridization, ligation, primer extension, enzymatic cleavage, or PCR-based amplification methods (reviewed by Syvänen 2001, Chen & Sullivan 2003). Although these techniques are efficient for validating SNPs, they differ in experimental design, associated costs, and time needed to carry out the validation pro-

Table 2. *Eriocheir sinensis*. Expressed sequence tag–single nucleotide polymorphism (EST–SNP) markers for the Chinese mitten crab. GC tails are underlined, and additional deliberate mismatches are boxed. AS1 and AS2: allele-specific primers, CR: common reverse primer, H_o: observed heterozygosity, H_e: expected heterozygosity

Locus (accession no.)	Putative function	SNP type	Primer sequence(5′-3′)	H_o	H_e	Minor allele and frequency
Es1082	Cytochrome c oxidase subunit III	C/T	AS1: GCGGGCAGGGCGGCAACCACTGGTTTAGTAAAATGAGTC AS2: GCGGGCAACCACTGGTTTAGTAAAATAATTT CR: AATTAAAAGGTTTATATTGTGTTGG	0.056	0.155	T 0.438
Es1291	Cryptocyanin 1	G/T	AS1: GCGGGCAGGGCGGCGCAAGGACTGGAACTGCTGCG AS2: GCGGGCGCAAGGACTGGAACTGCTTAT CR: TCCCTGAAGTAGGCGGCGTTG	0.333	0.282	G 0.150
Es1299	Toll-like receptor	C/T	AS1: GCGGGCAGGGCGGCACGCTGTCTCAGCCGCGCCTC AS2: GCGGGCACGCTGTCTCAGCCGCTACTT CR: GGAGAGGGAGTTCCCGTCCAAG	0.500	0.501	T 0.400
Es1310	C-type lectin	G/A	AS1: CGGGCAGGGCGGCCGGTTTTGTTGAAGTAGGAGGCGAG AS2: GCGGGCCGGTTTTGTTGAAGTAGGAGGATTAA CR: GTGCCATAGCTGGCGAGGAAAG	0.194	0.300	A 0.163
Es685	Crustin	C/T	AS1: GCGGGCAGGGCGGCATAAATTATGAATATGTTGTGCAC AS2: GCGGGCATAAATTATGAATATGTTGTATAT CR: TTCTTGCGGCCTCTGGTTTC	0.250	0.263	T 0.500
Es96	Hydroxypyruvate isomerase	G/A	AS1: GCGGGCAGGGCGGCGGCTTTAAGGCTGTGGAAGCG AS1: GCGGGCGGCTTTAAGGCTGTGGAAATA CR: AGAATTTCGGCAGACTCCCTG	0.778	0.507	G 0.450
Es58	Serine threonine protein kinase	G/A	AS1:GCGGGCAGGGCGGCGCTCAGGTACTATCGAAACTGACGG AS2: GCGGGCGCTCAGGTACTATCGAAACTGAATA CR: GACCCCGTAGAAATGAGGTACGAAG	0.833	0.501	A 0.400
Es350	Pacifastin light chain	G/A	AS1: GCGGGCAGGGCGGCCCACAGGGTGTGAAGGATGGCG AS2: GCGGGCCCACAGGGTGTGAAGGATAACA CR: AGGTTGGGTTGCCCTCACACG	0.500	0.482	A 0.225
Es422	Arginine kinase	G/A	AS1: GCGGGCAGGGCGGCTGCGTTGCGGCCGCTCCACGGAG AS2: GCGGGCTGCGTTGCGGCCGCTCCATGTAA CR: TGTACTGGGCCTCGGTGAGGCAGG	0.389	0.380	A 0.338
Es394	Serine protease	G/A	AS1: GCGGGCAGGGCGGCAATCCCGTCATCACCGCCACGCCG AS2: GCGGGCAATCCCGTCATCACCGCCACTCTA CR: CTAATGCCGCAAACGGGCTCGTAG	0.083	0.475	G 0.325
Es110	C-type lectin	C/A	AS1: GCGGGCAGGGCGGCAAGCAGGTCCAGGCTGGCGCC AS2: GCGGGCAAGCAGGTCCAGGCTGGTGTA CR: CACCGGAGCCCTTCATCTGCT	0.056	0.468	A 0.200
Es283	Hydrogen-transporting atp	G/A	AS1: CGGGCAGGGCGGCACCAAGTCTAGATATATCCCCAGGG AS2: GCGGGCACCAAGTCTAGATATATCCCTAGGA CR: AGTGATGATTCTGTCTACACATTAC	0.250	0.335	A 0.188

Table 3. Codon usage (relative synonymous codon usage, RSCU values) deduced from the *Eriocheir sinensis* expressed sequence tag (EST) dataset. High and low refer to relative synonymous codon usage (RSCU) values of highly and lowly expressed genes, respectively. Optimal codons, defined by a higher RSCU value in highly expressed genes, are in **bold** and indicated by **ΔRSCU ≥0.2 and *ΔRSCU ≥0.08

Amino acid	Codon	RSCU	
		High	Low
Ala	GCA	0.730	0.806
	GCC*	1.680	1.591
	GCG	0.731	0.717
	GCU	0.859	0.886
Cys	**UGC****	1.159	1.052
	UGU	0.841	0.948
Asp	GAC	1.305	1.255
	GAU	0.695	0.745
Glu	GAA	0.634	0.708
	GAG	1.366	1.292
Phe	**UUC****	1.423	1.293
	UUU	0.577	0.706
Gly	GGA	0.773	0.833
	GGC**	1.567	1.437
	GGG	0.899	0.949
	GGU	0.760	0.781
His	CAC	1.372	1.326
	CAU	0.628	0.674
Ile	AUA	0.422	0.548
	AUC**	1.809	1.531
	AUU	0.769	0.921
Lys	AAA	0.476	0.684
	AAG**	1.524	1.316
Leu	CUA	0.299	0.363
	CUC	1.601	1.544
	CUG*	2.062	1.927
	CUU	0.953	0.900
	UUA	0.238	0.363
	UUG	0.848	0.902
Asn	**AAC***	1.380	1.250
	AAU	0.619	0.750
Pro	CCA	0.888	1.007
	CCC*	1.437	1.282
	CCG	0.759	0.714
	CCU	0.916	0.996
Gln	CAA	0.498	0.552
	CAG	1.502	1.448
Arg	AGA	0.774	0.929
	AGG	1.511	1.525
	CGA	0.539	0.581
	CGC*	1.462	1.275
	CGG	0.920	0.952
	CGU	0.793	0.738
Ser	**AGC***	1.233	1.126
	AGU	0.657	0.730
	UCA	0.826	0.971
	UCC*	1.748	1.591
	UCG	0.690	0.639
	UCU	0.845	0.942
Thr	ACA	0.827	1.003
	ACC*	1.560	1.397
	ACG	0.831	0.809
	ACU	0.782	0.790
Val	GUA	0.443	0.469
	GUC*	1.090	1.047
	GUG*	1.811	1.747
	GUU	0.656	0.737
Tyr	**UAC***	1.452	1.338
	UAU	0.547	0.661

cess. In this study, we found T_m-shift PCR with melting curve analysis to be an efficient, cost effective, and reliable way for SNP validation. Of the 38 primer pairs tested in this study, 12 primer pairs could amplify genomic DNA of Chinese mitten crab and showed bi-allelic polymorphisms. We observed that some genotyping assays displayed poor discrimination in T_m between the 2 alleles' melting curves. Wang et al. (2005) demonstrated that this problem can be overcome for transition SNPs (AT to GC changes) by attaching the long tail (14 bp) to the allele-specific primer with the higher T_m base (G or C) at the 3′ end, and the short tail to the other allele-specific primer with the lower T_m base (A or T). We noticed that in some primer pairs, uneven height of melting peaks could make genotype scoring difficult when one primer amplified substantially more efficiently than the other. To resolve this issue, we added more efficient primer at half of the original concentration (0.1 µM). This is not always strictly necessary, because genotypes can be identified even under the original conditions.

Non-random usage of synonymous codons is a widespread phenomenon, observed in genomes from many species in all domains of life. Such codon usage biases may result from mutational biases, from natural selection acting on silent changes in DNA, or both (Duret 2002). Generally speaking, highly expressed genes, producing abundant proteins, use a subset of optimal codons which are recognized by the most abundant tRNAs. In this study, we defined 16 codons as the 'optimal codons' of Chinese mitten crab (Table 3), which will provide significant information for gene expression prediction. All 'optimal codons' inferred from ΔRSCU values were -G or -C ending codons, despite the average GC-content of 56.6% in this study. Although further investigations are necessary to understand the causes affecting codon usage bias, our results imply that the gene expression levels and GC-content might play important roles in codon usage bias.

LITERATURE CITED

Angellotti MC, Bhuiyan SB, Chen G, Wan XF (2007) CodonO: codon usage bias analysis within and across genomes. Nucleic Acids Res 35:W132–W136

Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G, Dickson BJ (2001) Genetic mapping with SNP markers in *Drosophila*. Nat Genet 29:475–481

Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenomics J 3:77–96

Clark PF, Rainbow PS, Robbins RS, Smith B, Yeomans WE, Thomas M, Dobson G (1998) The alien Chinese mitten crab, *Eriocheir sinensis* (Crustacea: Decapoda: Brachyura), in the Thames catchment. J Mar Biol Assoc UK 78: 1215–1221

Curole JP, Hedgecock D (2005) High frequency of SNPs in the

Pacific Oyster genome. 13th Int Plant Anim Genomes Conf, San Diego, CA. Abstract available at http://intl-pag.org/13/abstracts/PAG13_W026.html

Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649

Duret L, Mouchiroud D (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc Natl Acad Sci USA 96:4482–4487

Frazer KA, Ballinger DG, Cox DR, Hinds DA and others (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861

Gao XG, Li HJ, Li YF, Sui LJ, Zhu B, Liang Y, Liu WD, He CB (2010) Sixteen polymorphic simple sequence repeat markers from expressed sequence tags of the Chinese mitten crab *Eriocheir sinensis.* Int J Mol Sci 11:3035–3038

Gorbach DM, Hu ZL, Du ZQ, Rothschild MF (2009) SNP discovery in *Litopenaeus vannamei* with a new computational pipeline. Anim Genet 40:106–109

Hayes B, Laerdahl JK, Lien S, Moen T and others (2007) An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. Aquaculture 265:82–90

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet 2: 235–258

Li XD, Dong SL, Lei YZ, Li YH (2007) The effect of stocking density of Chinese mitten crab *Eriocheir sinensis* on rice and crab seed yields in rice-crab culture systems. Aquaculture 273:487–493

Li RH, Li Q, Kong LF (2009) Characterization of expressed sequence tag-derived single-nucleotide polymorphisms in the bay scallop *Argopecten irradians irradians.* Fish Sci 75:1389–1400

Li HJ, Zhu D, Gao XG, Li YF, Wang J, He CB (2010a) Mining single nucleotide polymorphisms from EST data of hard clam *Meretrix meretrix.* Conserv Genet Res 2:69–72

Li HJ, Liu WD, Gao XG, Zhu D, Wang J, Li YF, He CB (2010b) Identification of host-defense genes and development of microsatellite markers from ESTs of hard clam *Meretrix meretrix*. Mol Biol Rep doi: 10.1007/s11033-010-0165-4

Liu ZJ, Cordes JF (2004) DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238: 1–37

Matsuzaki H, Dong S, Loi H, Di X and others (2004) Genotyping over 100 000 SNPs on a pair of oligonucleotide arrays. Nat Methods 1:109–111

Morlais I, Severson DW (2003) Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti.* Insect Mol Biol 12:631–639

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. Genome Res 9:167–174

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM and others (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Sauvage C, Bierne N, Lapègue S, Boudry P (2007) Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas.* Gene 406:13–22

Syvänen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2: 930–942

Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JA (2006) QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. BMC Bioinformatics 7:438

Wang HZ, Wang HJ, Liang XM, Cui YD (2006) Stocking models of Chinese mitten crab (*Eriocheir japonica sinensis*) in Yangtze lakes. Aquaculture 255:456–465

Wang J, Chuang K, Ahluwalia M, Patel S, Umblas N, Mirel D, Higuchi R, Germer S (2005) High-throughput SNP genotyping by single-tube PCR with T_m-shift primers. Biotechniques 39:885–893

Wang S, Sha ZX, Sonstegard TS, Liu H and others (2008) Quality assessment parameters for EST-derived SNPs from catfish. BMC Genomics 9:450

Yeh FC, Yang R, Boyle TJ, Ye Z, Xiyan JM (2000) POPGENE 32, Microsoft Windows-based freeware for population genetic analysis. Molecular Biology and Biotechnology Centre. University of Alberta, Edmonton. Available at www.ualberta.ca/~fyeh/popgene.html

Ying XP, Yang WX, Zhang YP (2006) Comparative studies on fatty acid composition of the ovaries and hepatopancreas at different physiological stages of the Chinese mitten crab. Aquaculture 256:617–623

Zhang LS, Guo XM (2010) Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. Aquaculture 302:124–129

Editorial responsibility: Francesco Patti, Ischia, Italy

Submitted: June 21, 2010; Accepted: November 5, 2010
Proofs received from author(s): December 13, 2010