

Article

## Large-Scale Isolation of Microsatellites from Chinese Mitten Crab *Eriocheir sinensis* via a Solexa Genomic Survey

Liang-Wei Xiong<sup>1,2</sup>, Qun Wang<sup>3</sup> and Gao-Feng Qiu<sup>1,\*</sup>

<sup>1</sup> Key laboratory of Freshwater Aquatic Genetic Resources Certificated by Ministry of Agriculture, College of Life Science, Shanghai Ocean University, 999 Hucheng Huan Road, Shanghai 201306, China; E-Mail: xlwei2002@163.com

<sup>2</sup> Jiangsu Animal Husbandry & Veterinary College, Jiangsu 225300, China

<sup>3</sup> College of Life Science, East China Normal University, Shanghai 200062, China; E-Mail: qwang@bio.ecnu.edu.cn

\* Author to whom correspondence should be addressed; E-Mail: gfqiu@shou.edu.cn; Tel./Fax: +86-21-6190-0436.

Received: 26 September 2012; in revised form: 14 November 2012 / Accepted: 21 November 2012 / Published: 3 December 2012

---

**Abstract:** Microsatellites are simple sequence repeats with a high degree of polymorphism in the genome; they are used as DNA markers in many molecular genetic studies. Using traditional methods such as the magnetic beads enrichment method, only a few microsatellite markers have been isolated from the Chinese mitten crab *Eriocheir sinensis*, as the crab genome sequence information is unavailable. Here, we have identified a large number of microsatellites from the Chinese mitten crab by taking advantage of Solexa genomic surveying. A total of 141,737 SSR (simple sequence repeats) motifs were identified via analysis of 883 Mb of the crab genomic DNA information, including mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeat motifs. The number of di-nucleotide repeat motifs was 82,979, making this the most abundant type of repeat motif (58.54%); the second most abundant were the tri-nucleotide repeats (42,657, 30.11%). Among di-nucleotide repeats, the most frequent repeats were AC motifs, accounting for 67.55% of the total number. AGG motifs were the most frequent (59.32%) of the tri-nucleotide motifs. A total of 15,125 microsatellite loci had a flanking sequence suitable for setting the primer of a polymerase chain reaction (PCR). To verify the identified SSRs, a subset of 100 primer pairs was randomly selected for PCR. Eighty two primer sets (82%) produced strong PCR products matching expected sizes, and 78% were polymorphic. In an analysis of 30 wild individuals from the Yangtze River with 20 primer sets, the number of alleles per locus

ranged from 2–14 and the mean allelic richness was 7.4. No linkage disequilibrium was found between any pair of loci, indicating that the markers were independent. The Hardy-Weinberg equilibrium test showed significant deviation in four of the 20 microsatellite loci after sequential Bonferroni corrections. This method is cost- and time-effective in comparison to traditional approaches for the isolation of microsatellites.

**Keywords:** microsatellite marker; *Eriocheir sinensis*; solexa sequencing

---

## 1. Introduction

Microsatellites or simple sequence repeats (SSRs), which are tandemly repeated units of one to six nucleotides, have been abundant in all prokaryotic and eukaryotic genomes analysed to date [1,2]. They are evenly distributed throughout genomes and are usually characterized by a high degree of length polymorphism, which makes them one of the most popular genetic markers for a wide range of applications including genetic mapping, marker-assisted selection breeding (MAS), genetic diversity studies, population structure analysis, gene flow and germplasm conservation studies [3–6]. However, a major drawback of the application of microsatellite markers is that they need to be isolated *de novo* from most organisms being examined for the first time.

Traditionally, the isolation of SSR markers has relied on the screening of genomic libraries using repetitive probes and sequencing of positive clones to develop SSR primers [7]. However, most of these steps are difficult, time-consuming, and relatively inefficient. Next-generation sequencing (NGS) technologies that speed up the process to generate a large number of sequences have been used recently to isolate SSR markers in studies of non-model animals [8,9], plants [10–12] and fishes [13–15].

The Chinese mitten crab, *Eriocheir sinensis*, is a euryhaline brachyuran with a native range extending from the eastern Pacific coast of China to the Korean Peninsula [16]. In China, the basic production technology of mitten crab populations has had a long history, and today, it is one of the most economically important indigenous organisms in freshwater aquaculture with an annual aquaculture production of 570,000 mt in 2009, valued at 4.0 billion USD, according to the State of World Fisheries and Aquaculture from FAO [17]. Unfortunately, like many other cultured decapod species the mitten crab has not been completely domesticated and many broodstock crabs are collected from wild populations. To develop a strain with good performance, conventional selective breeding programs of the crab have been conducted for several years. Because the selection was performed based mainly on phenotypic assessment and because the breeding cycle is long, the breeding programs of the mitten crab have been inefficient. As verified in many cultured species, molecular markers introducing selection (*i.e.*, MAS) are required to accelerate the course of the crab breeding [18]. Usually, MAS depends on a high-resolution genetic linkage map for various purposes including characterization of quantitative trait loci (QTL) [19]. Development of a large number of sequence-based genetic markers, such as microsatellites, is an essential step for MAS and linkage map construction. However, karyotypic analysis has shown that the diploid chromosome number in the mitten crab is large ( $2n = 146$ ) [20]. It is estimated that hundreds of SSRs are required for construction of a high-density linkage map. To date, approximately 47 microsatellite markers have been developed [21–24], which is inadequate for

construction of a linkage map. In this study, we used the Solexa sequencing technology for the whole genomic DNA survey of the Chinese mitten crab, in order to isolate polymorphic microsatellites on a large-scale for its linkage map construction.

## 2. Results

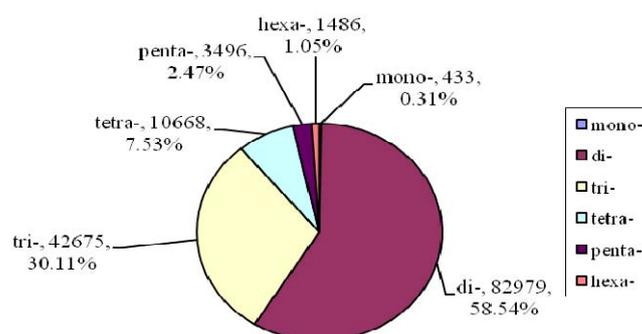
### 2.1. Genome Survey and Assembly

Solexa genomic surveying produced a total of 76.27 Gb of raw genomic data. We assembled the short reads using SOAP *de novo*, a genome assembler developed specifically for use with next-generation short-read sequences [25]. After excluding the data from poor libraries (reads with more than 10% of Q < 20 bases) and filtering low-quality sequences (reads with ambiguous bases “N”), 56.20 Gb reads remained as high-quality reads for *de novo* assembly. Finally, 883 Mb of sequence data were obtained from 1,096,936 scaffolds with a length range from 0.1 kb to 10 kb.

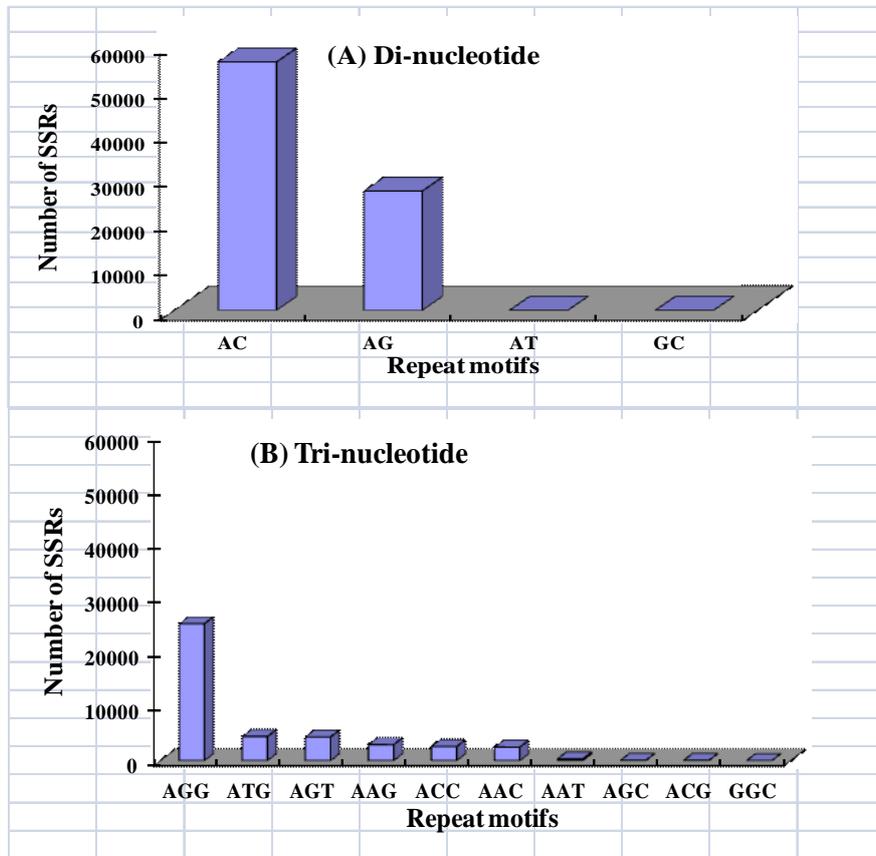
### 2.2. Microsatellite Loci Discovery and Primer Pair Design

The resultant 883 Mb of DNA sequence was analyzed to evaluate different types of perfect mono-, di-, tri-, tetra-, penta- and hexa-nucleotides. A total of 141,737 distinct microsatellite loci were identified. The SSR distribution density was approximately 161 loci per Mb. The most abundant type of repeat motif was a di-nucleotide (58.54%), followed by tri-nucleotide (30.11%), tetra-nucleotide (7.53%), penta-nucleotide (2.47%), hexa-nucleotide (1.05%), and mono-nucleotide (0.31%) repeat units (Figure 1). There were large differences in the relative abundance of special repeat motifs. As shown in Figure 2A, among the di-nucleotide sequences, the motif AC had the highest frequency, representing 67.55% of the sampled sequences, followed by AG (32.44%). Motifs AT and GC (<0.01% each) were comparatively rare. The most frequent tri-nucleotide was AGG (59.32%), whereas AAT (0.74%), AGC (0.11%), ACG (0.09%), and GCC (0.01%) were comparatively scarce (Figure 2B). The frequency distributions from mono- to hexa-nucleotide repeats were calculated and are shown in Figure 3. The bulk of repeat sequences were centralized in the domain composed of low copy number, and fewer sequences were seen with increasing copy number. Among the di-nucleotide repeat sequences, repeats with 10–11 copies were the most common (33.28%); among tri-nucleotide repeat sequences, repeats with 8–9 copies were the most common (22.28%). The size of each repeat sequence was determined by the copy number of its repeat unit (Table 1).

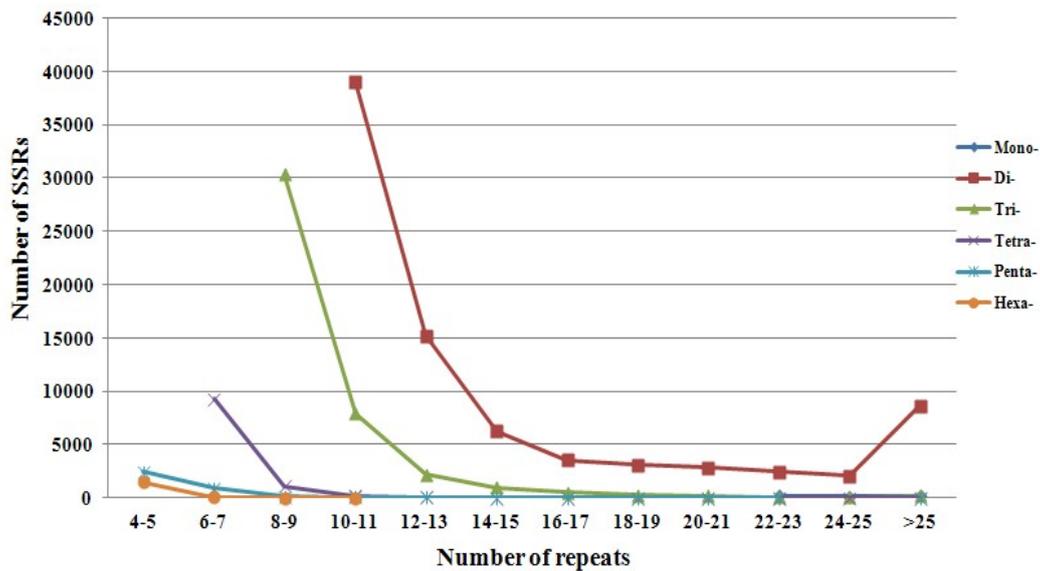
**Figure 1.** The percentages of mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeats in motif sequences from the Chinese mitten crab.



**Figure 2.** Observed counts of identified microsatellite loci for different repeat sequence motifs of (A) di- nucleotide and (B) tri-nucleotide repeats from the Chinese mitten crab.



**Figure 3.** Observed number of microsatellites with mono-, di-, tri-, tetra-, penta- and hexa-nucleotide motifs in 883 Mb sequence.



**Table 1.** Frequency of simple sequence repeats (SSRs) in *Eriochloa sinensis*.

Motif length	Repeat numbers											Total number	%	
	4-5	6-7	8-9	10-11	12-13	14-15	16-17	18-19	20-21	22-23	24-25			>25
Mono-										184	100	149	433	0.31
Di-				38990	15208	6239	3541	3052	2830	2441	2070	8608	82979	58.54
Tri-			30358	7936	2189	969	497	264	151	73	46	192	42675	30.11
Tetra-		9278	1093	208	61	8	2	6	1	7	3	1	10668	7.53
Penta-	2403	909	125	33	16	3	2	2	1	1		1	3496	2.47
Hexa-	1460	23	2	1									1486	1.05
Total	3863	10210	31578	47168	17474	7219	4042	3324	2983	2706	2219	8951	141737	100
%	2.73	7.20	22.28	33.28	12.33	5.09	2.85	2.35	2.10	1.91	1.57	6.32		

To estimate the number of loci that represented promising candidates for PCR amplification-based scoring of microsatellite length variation, we screened the loci to determine which of them contained suitable flanking PCR primer sites; we referred to such loci as “potentially amplifiable loci” or PAL. We identified 15,125 PAL and designed their corresponding primer pairs, which represented 10.67% of microsatellite loci identified.

### 2.3. SSR Validation and Population Genetic Analysis

A subset of 100 PAL was selected for validation. Primers were designed for these loci and tested using the genomic DNA of a panel of five individuals. Eighty-two primer sets (82%) successfully yielded amplicons matching the expected sizes although they contained some nonspecific bands, and eighteen primer pairs did not give any amplification product. Of the eighty-two primer sets, four sets generated monomorphic products in all the tested individuals. A panel of 20 SSRs was used for further polymorphism testing in 30 individuals from a wild population. The primer sequences, repeat motifs, annealing temperatures, number of alleles, PCR ranges and the heterozygosity for the 20 new microsatellite loci are summarized in Table 2. The amplification results showed that all the loci were polymorphic. The number of alleles per locus varied from 2 to 14 with an average of 7.4. No linkage disequilibrium was found between any pair of loci ( $p > 0.05$  indicating that the markers were independent). The Hardy-Weinberg equilibrium (HWE) test, indicating the deviation from the expected heterozygosity, showed significant deviation in four (*Eri3*, *Eri8*, *Eri11* and *Eri14*) of the 20 loci in the wild population after sequential Bonferroni corrections. Null alleles were presumed in five (*Eri3*, *Eri6*, *Eri8*, *Eri14* and *Eri16*) of the 20 loci. Expected heterozygosities ( $H_e$ ) ranged from 0.510 to 0.971 (mean:  $0.800 \pm 0.147$ ) and observed heterozygosities ( $H_o$ ) from 0.326 to 0.958 (mean:  $0.689 \pm 0.170$ ).

**Table 2.** Characteristics of 20 microsatellite loci in *Eriocher sinensis*, tested with 30 samples.

Locus	Primer sequence(5'-3')	Repeat motif	$T_a^a$	Number of alleles	Allele size range (bp) <sup>b</sup>	$H_o$	$H_E$	$P^c$	GeneBank Accession No.
<i>Eri1</i>	F:GATAGACCGTAAATGAGACGGCTG R:GGACGGAGAAAACCTAGAGACCACA	(GGA) <sub>9</sub>	63	8	151–174 (157)	0.643	0.759	0.130	KC143114
<i>Eri2</i>	F:GGATTTACTTAAGTTGGGGCTCGT R:CGACGCAGTTTTGTCTAGAGACCT	(GAG) <sub>8</sub>	63	10	119–161 (146)	0.750	0.918	0.057	KC143100
<i>Eri3<sup>d</sup></i>	F:CAGCGAAAAACAGGAAGCATTAG R:GGAAAGGGAAAGTGAAGGATGAAT	(AC) <sub>11</sub>	63	12	140–212 (166)	0.793	0.950	0.000	KC143117
<i>Eri4</i>	F:TTCTTTGAGCGACATGCAAAAGT R:AGACAGACAGACAAAAACGCTCCT	(TG) <sub>29</sub>	62	14	132–192 (150)	0.950	0.971	0.100	KC143110
<i>Eri5</i>	F:TAGGGGTTTTAGGTGTGGTGATA R:ATTTATGTGGAGGGAATGGGAGAT	(TGA) <sub>9</sub>	62	8	125–167 (143)	0.658	0.855	0.361	KC143115
<i>Eri6<sup>d</sup></i>	F:CAACCACTACAACATCAAAACCACC R:GACTTTACGACCACGAAATGGAG	(CA) <sub>26</sub>	62	9	91–181 (109)	0.792	0.897	0.151	KC143108
<i>Eri7</i>	F:TAACCTAAACAGCAACAGCAGCAA R:AAAGGGTTAGAAAGGAAGGAGGGT	(AGT) <sub>8</sub>	63	4	128–146 (134)	0.567	0.693	0.251	KC143118
<i>Eri8<sup>d</sup></i>	F:TGTTGAGTGTGATGTTGTGATGC R:TAATAGCGCCAAACTTTGTTGAT	(GT) <sub>12</sub>	65	11	160–226 (196)	0.958	0.935	0.001	KC143116
<i>Eri9</i>	F:TGCATATGTTGTTTTACTGACGTGT R:CATCATCACCATCATCATCACAAA	(AC) <sub>18</sub>	63	12	140–188 (154)	0.884	0.945	0.641	KC143109
<i>Eri10</i>	F:TACCTTTTTCAGGGTGAGTGAAGG R:AAGGACAGGAGGGAAAATGAGAGT	(GT) <sub>23</sub>	63	7	138–208 (160)	0.865	0.917	0.317	KC143102
<i>Eri11</i>	F:ATGTTTATTTTACAAACGCGAAGC R:TGTCTTCTGTCTCTGTCTGTGTG	(ACAT) <sub>6</sub>	63	4	157–177 (157)	0.583	0.732	0.002	KC143103
<i>Eri12</i>	F:ACCCATCTCAAGTCCAGACTCATC R:AGAGGATGCAAGGGAAATAAGGAG	(CCT) <sub>8</sub>	63	7	152–170 (161)	0.700	0.784	0.113	KC143113
<i>Eri13</i>	F:AAGGAAGGCAGTTAGGAGGGTATG R:TTATTATTGTGGCGACGAAGGGT	(AC) <sub>11</sub>	65	8	189–259 (221)	0.668	0.950	0.147	KC143106
<i>Eri14<sup>d</sup></i>	F:TGTTGTTGTGTCATGCTGTCTTTT R:TCAGAAACACCGCACTCGATATAA	(TCTCA) <sub>5</sub>	62	2	142–147 (142)	0.420	0.510	0.000	KC143119
<i>Eri15</i>	F:TCACCCCTTACTGAGCATAACACA R:CCTTATCCTGCGACTCGTAATGTT	(CA) <sub>12</sub>	63	5	93–109 (105)	0.736	0.773	0.250	KC143105
<i>Eri16<sup>d</sup></i>	F:TCCTCCATATGCTCTTTGTAGGTG R:AAGGCCAGGAGTATGGTGAAC	(AC) <sub>11</sub>	63	8	150–192 (156)	0.699	0.786	0.115	KC143111
<i>Eri17</i>	F:CAGCATGTCCAGTCTCTTCTGTGT R:GCTGAGAGAATATGTATGATGACATGG	(GT) <sub>15</sub>	65	6	141–159 (145)	0.772	0.864	0.054	KC143107
<i>Eri18</i>	F:TGGCATTGATTGATGTGAGTAGTG R:CTAACCTTCTCGACACCTTTGCAT	(GGT) <sub>10</sub>	62	5	85–106 (94)	0.326	0.551	0.269	KC143104
<i>Eri19</i>	F:CAGACCCCTCCCGATGATACTACTAC R:CTATCCACTCAGCTACCGCCTCT	(ATCT) <sub>8</sub>	63	4	133–155 (143)	0.563	0.668	0.531	KC143112
<i>Eri20</i>	F:GAGATGGAGGTAGATGATCGAGGA R:CAAGGCACTCAATCTCAACCTTTT	(GAAG) <sub>7</sub>	63	4	120–144 (132)	0.447	0.532	0.516	KC143101

<sup>a</sup> annealing temperature; <sup>b</sup> allele size (size of sequenced allele); <sup>c</sup>  $p$ -values of test for deviation from Hardy-Weinberg equilibrium; <sup>d</sup> microsatellite loci revealed the presence of null alleles with MICRO-CHEKER 2.2.3.

### 3. Discussion

Prior to our study, less than 50 SSR markers had been developed in the mitten crab using conventional methods [21–23,26]. By taking advantage of the Solexa genomic survey here, we discovered an extensive set of 141,737 microsatellite loci in which 15,125 loci are PAL. Compared to the weeks or even months that would have been spent obtaining only tens of microsatellite loci by conventional methods, this process only took one month. This method consisting of a genomic survey is composed of only four steps: (i) isolation of genomic DNA; (ii) DNA sequencing and assembly; (iii) microsatellite loci discovery and primer design, and (iv) microsatellite verification [8]. No cloning or library screening is required. This new method targets all microsatellite repeat types (e.g., mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide), while conventional methods require an SSR-enriched genomic library for screening a limited number of specific microsatellite motifs, and the choice of motif can have an effect on the variability detected [27]. Thus, the Solexa genomic survey is an effective method for large-scale isolation of microsatellite markers. Furthermore, millions of base pairs of genomic sequence are available in the survey, potentially providing a framework for further genomic analyses and a useful resource for gene research.

It is estimated that the size of crustacean genomes ranges from 55 Mb to 1800 Mb [28]. The cumulative length of the assembled sequences of the Chinese mitten crab genome reached 883 Mb and will provide important information about the mitten crab genomic organization of repeat sequences. The di-nucleotide repeats contributed to a major proportion of genome SSRs, while only a very small proportion was contributed by mono-, tri-, tetra-, penta- and hexa-nucleotide repeats (Figure 1). The number of di-nucleotide repeats was also the most prevalent in the Chinese shrimp [29], fruit fly [30], pufferfish [31], human [32], and plant [33]. Moreover, the di-nucleotide repeat class was the only class of repeats found in the genome of the prokaryote methanogenic archaeon (*Methanococcus jannaschii*) [34]. These data indicate that di-nucleotide repeats may have an important biological significance in the genesis and development of repeat sequences. Among the di-nucleotide repeats, AC was the most frequent motif in the crab genome as observed in many other eukaryotes [35], with the exception of plants in which AT was the most abundant motif [36]. Interestingly, the frequency of GC and AT di-nucleotide repeats was in less than 0.01% each of the crab genome (Figure 2A). This result is similar to that of most organisms including plants whose genome contains the rare GC repeat class [37]. One explanation of this GC suppression may be that the GC repeats present structural problems, and this point was exemplified with the association of a similar CCG repeat with the fragile X site on the human X chromosome [29]. Among the tri-nucleotide repeats, AGG (59.32%) was the most frequent in our dataset, as it was in the Japanese pufferfish [31]. AAT, AGC, AAG were the most frequent motifs in human, fruit fly and Chinese shrimp, respectively [29,30,32]. We hypothesized that the relative frequency of different types of tri-nucleotide motifs varied according to species. 1789 5,789

When a subset of 100 primer pairs from 15,125 PAL was tested, 78 of them were successfully amplified with polymorphisms from five individuals. When 20 microsatellite markers were selected for further polymorphism testing, all of them showed high diversity and variation among the 30 individuals tested. Adjusting for null alleles, the mean number of alleles per locus,  $H_E$  and  $H_O$  are 7.4, 0.800 and 0.689 respectively, demonstrating a relatively high genetic diversity within crab individuals. This is similar to reports from studies in other locations [21,23,24,38,39]; however, four microsatellite loci

exhibited drastic departures from *HWE* as shown by the fact that  $H_E$  was apparently higher than  $H_O$  (Table 2). In this study, heterozygote deficiency at the four loci seems to be strong evidence for the deviations of *HWE*. Although there are several possible explanations for a deficiency of heterozygotes, here the consistent pattern across loci suggests that the individuals examined most likely do not originate from a single panmictic population (individuals in this study were sampled from the Yangtze River over quite a large geographical distance).

## 4. Experimental Section

### 4.1. Sample

A single specimen of Chinese mitten crab from one aquaculture farm in China was used as the sole source of tissue/DNA for a genomic DNA survey. The total genomic DNA was extracted from the muscle tissue using the standard proteinase K, phenol-chloroform procedure [40]. The genomic DNA concentration was quantified using an ultraviolet (UV) spectrophotometer (Nanodrop, Madison, WI, USA), and the DNA quality was assessed on a 0.8% agarose gel. In the population genetic analysis, the total DNA was extracted by the same method from ethanol-fixed tissues of 30 wild individuals who had been stored at the Key laboratory of Freshwater Aquatic Genetic Resources Certified by the Ministry of Agriculture, China.

### 4.2. DNA Sequencing

Approximately 1 mg of genomic DNA (>23 kb, OD260/OD280  $\approx$  1.80) was subjected to Solexa sequencing analysis at Beijing Genomics Institute (BGI; Shenzhen, China) using whole-genome shotgun sequencing strategy and Illumina Genome Analyzer sequencing technology. Libraries with an insert size of 170 bp and 500 bp were prepared following the manufacturer's instructions (Illumina, San Diego, CA, USA). After library preparation and quality control of the DNA samples, four lane (two 170 bp, two 500 bp) template DNA fragments were hybridized to the surface of flow cells on an Illumina Genome Analyzer II sequencer (GA2), amplified to form clusters, and sequenced following the standard Illumina protocol.

### 4.3. De Novo Assembly

The read sequence was aligned using the SOAP *de novo* software [25] with the default setting, which adopts the De Bruijn graph data structure to construct contigs [41]. The reads were then realigned to the contig sequence, and the paired-end relationship between the reads was transferred to linkage between contigs. Scaffolds starting with short paired-ends were constructed and then the scaffold process was iterated step-by-step using longer insert-size paired-ends. To fill the intra-scaffold gaps, the paired-end information was used to retrieve read pairs that had one read well-aligned on the contigs and another read located in the gap region. We then did a local assembly for the collected reads.

#### 4.4. Identification of Microsatellite Loci

The assembled sequences were scanned for perfect mono-, di-, tri-, tetra-, penta- and hexa-nucleotide tandem repeats (*i.e.*, microsatellite loci) that met the following criteria: a minimum pattern length of 22 bp, at least 11 repeat units in case of mono-nucleotide and at least 11, 8, 6, 5, and 4 repeat units for di-, tri-, tetra-, penta- and hexa-nucleotide SSR, respectively, using the SSRFinder program [42]. Each simple sequence was counted on one strand only, and the microsatellite loci were then sorted by the monomer sequence of the repeat (e.g., AG or AAG repeats) and by the number of tandemly repeated units. Non-unique repeat motifs (reverse-complement repeat motifs (e.g., AC and GT) and translated or shifted motifs (e.g., AAT, ATA, TAA, TTA, TAT and ATT)) were grouped together, so that there were a total of 2 unique 1mer repeats, 4 unique 2mer repeats, 10 unique 3mer repeats, 33 unique 4mer, 102 unique 5mer and 350 unique 6mer repeats [43].

#### 4.5. Screening of Loci suitable for PCR and Primer Design

Newly identified microsatellite loci are typically useful only if primers in the non-repeated flanking regions around the microsatellite can be designed and used successfully for PCR amplification. We therefore screened the assembled sequence with microsatellite loci for flanking regions with high quality PCR priming sites; we referred to such loci as “potentially amplifiable loci” or PAL. The primer-pair design process was automated to submit large batches of sequences to a local installation of the program Primer 3 [42]. We used fairly stringent criteria for the primer design, including the following specifications: (i) GC content >40%; (ii) melting temperatures ( $T_m$ ) 60–65 °C with a maximum of 1 °C difference between paired primers; (iii) amplicon length range 80–300 bp, and (iv) primer size  $24 \pm 4$  bp. All the remaining parameters were left at the default settings. If all criteria were met, a single primer-pair was chosen based on the highest Primer 3 assigned score and targeting the longest microsatellite element within a sequence.

#### 4.6. SSR Marker Validation and Population Genetic Analysis

A subset of 100 primer pairs was synthesized and screened for amplification quality using the genomic DNA of a panel of five wild individuals. From the primers that showed scorable amplification, those that also produced specific amplification products and amplified consistently across individuals were further evaluated for marker polymorphism with additional 30 wild individual organisms sampled from the Yangtze River in China. Standard PCR was carried out in a 10  $\mu$ L reaction containing 1  $\mu$ L of DNA (~10 ng), 0.5  $\mu$ L of forward primer and 0.5  $\mu$ L of reverse primer (10  $\mu$ M each), 5  $\mu$ L of 2 $\times$  Taq PCR MasterMix (Shanghai Xufei Company, China), and 3  $\mu$ L of distilled water. The temperature cycling conditions were as follows: 95 °C for 4 min followed by 35 cycles of 94 °C for 30 s, 1 min at the annealing temperature listed in Table 2 and 72 °C for 1 min, with a final extension of 72 °C for 10 min. The separation of alleles was performed on 8% denaturing polyacrylamide gels with a 50 bp DNA marker (TaKaRa) to calculate the length of the SSR amplicons. Gels were stained with silver nitrate as previously described [44]. The allelic determination was made manually with the software package of Gel-Pro Analyzer 4.5 (<http://www.mediacy.com/index.aspx?page=GelPro>). The number of alleles per locus and heterozygosity were calculated using Arlequin version 3.0 [45]. Tests for linkage

disequilibrium between pairs of loci and deviations from HWE ( $p < 0.05$ ) were estimated using GENEPOP version 4.0 [46], and the adjusted  $p$ -values for both analyses were obtained using a sequential Bonferroni test for multiple comparisons. MICRO-CHECKER version 2.2.3 [47] was used to test the presence of null alleles.

## 5. Conclusions

The Solexa sequencing method was applied to the development of microsatellite markers for the Chinese mitten crab. More than 15,000 microsatellites were achieved using this method. To the best of our knowledge, this is the first time that such a large number of microsatellites have been isolated from this crab. At the same time, the results of this study clearly demonstrate that in addition to the 454 pyrosequencing technology, Solexa sequencing technology is suitable for the isolation of microsatellites for non-model animals in an efficient and cost-effective way.

## Acknowledgements

We are grateful to Weibing Guang from Shanghai Ocean University and Xinghua Liaoyuan Special Cultivation Co., Ltd for assistance with specimen collection. This research was supported by the National Science & Technology Pillar Program (2012BAD26B04), a grant from the Shanghai Municipal Government (D8003100208), the Marine Biology Program of Shanghai Leading Academic Discipline (J50701) and Shanghai University Knowledge Service Platform (ZF1206).

## References

1. Weber, J.L. Informativeness of human (dC-dA)<sub>n</sub>(dG-dT)<sub>n</sub> polymorphisms. *Genomics* **1990**, *7*, 524–530.
2. Field, D.; Wills, C. Long, polymorphic microsatellites in simple organisms. *Proc. R. Soc. Lond. Ser. B* **1996**, *263*, 209–215.
3. Goldstein, D.B.; Schlotterer, C. *Microsatellites: Evolution and Applications*; Oxford University Press: New York, NY, USA, 1999.
4. Gupta, P.; Varshney, R. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **2000**, *113*, 163–185.
5. Gu, Z.X.; Gou, T.J.; Xi, X.B. Applications of microsatellite markers in studies of genetics and breeding of fish. *Chin. J. Agric. Biotechnol.* **2006**, *3*, 83–87.
6. Chistiakov, D.A.; Helleman, B.; Volckaert, F.A.M. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* **2006**, *255*, 1–29.
7. Zane, L.; Bargelloni, L.; Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **2002**, *11*, 1–16.
8. Yu, J.-N.; Won, C.; Jun, J.; Lim, Y.; Kwak, M. Fast and Cost-Effective Mining of Microsatellite Markers Using NGS Technology: An Example of a Korean Water Deer *Hydropotes inermis argyropus*. *PLoS One* **2011**, *6*, e26933.

9. Castoe, T.A.; Poole, A.W.; Gu, W.; de Koing, A.P.J.; Daza, J.M.; Smith, E.N.; Pollock, D.D. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Mol. Ecol. Resour.* **2010**, *10*, 341–347.
10. Jennings, T.N.; Knaus, B.J.; Mullins, T.D.; Haig, S.M.; Cronn, R.C. Multiplexed microsatellite recovery using massively parallel sequencing. *Mol. Ecol. Resour.* **2011**, *11*, 1060–1067.
11. Blanca, J.; Canizares, J.; Roig, C.; Ziarso, P.; Nuez, F.; Pico, B. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* **2011**, *12*, 104.
12. Setsuko, S.; Uchiyama, K.; Sugai, K.; Yoshimaru, H. Rapid development of microsatellite markers for *Pandanus boninensis* (Pandanales) by pyrosequencing technology. *Am. J. Bot.* **2012**, *99*, e33–e37.
13. Saarinen, E.V.; Austin, J.D. When Technology Meets Conservation: Increased Microsatellite Marker Production Using 454 Genome Sequencing on the Endangered Okaloosa Darter (*Etheostoma okaloosae*). *J. Hered.* **2010**, *101*, 784–788.
14. Kang, J.H.; Park, J.Y.; Jo, H.S. Rapid Development of Microsatellite Markers with 454 Pyrosequencing in a Vulnerable Fish, the Mottled Skate, *Raja pulchra*. *Int. J. Mol. Sci.* **2012**, *13*, 7199–7211.
15. Wang, J.; Yu, X.; Zhao, K.; Zhang, Y.; Tong, J.; Peng, Z. Microsatellite development for an endangered bream *Megalobrama pellegrini* (Teleostei, Cyprinidae) using 454 sequencing. *Int. J. Mol. Sci.* **2012**, *13*, 3009–3021.
16. Fushun, G.X.Z. Resources and Culturing Situation of Chinese Mitten Crab (*Eriocheir sinensis*) and Species Character Conservation. *J. Lake Sci.* **2001**, *13*, 267–271.
17. Food and Agriculture Organization of the United Nations. *FAO Yearbooks of Fishery Statistics Summary Tables, Aquaculture Production 2009*; FAO: Rome, Italy; 2010, p. 44.
18. Ozaki, A.; Sakamoto, T.; Khoo, S.; Nakamura, K.; Coimbra, M.R.M.; Akutsu, T.; Okamoto, N. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genomics* **2001**, *265*, 23–31.
19. Dekkers, J.C.M.; Hospital, F. The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* **2002**, *3*, 22–32.
20. Du, N.S.; Lai, W.; Xue, L.Z. The chromosomes of the Chinese mitten-handed crab, *Eriocheir sinensis* (Crustacea, Decapoda). *Zool. Res.* **1986**, *7*, 293–296.
21. Hanfling, B.; Weetman, D. Characterization of microsatellite loci for the Chinese mitten crab, *Eriocheir sinensis*. *Mol. Ecol. Notes* **2003**, *3*, 15–17.
22. Zhu, Z.Y.; Shi, Y.H.; Le, G.W. Isolation and characterization of polymorphic microsatellites from Chinese mitten crab, *Eriocheir sinensis*. *Mol. Ecol. Notes* **2006**, *6*, 838–839.
23. Mao, R.X.; Zhao, Y.Y.; Liu, F.J.; Jia, Z.Y.; Hou, N.; Chang, Y.M.; Lu, C.Y.; Liang, L.Q.; Sun, X.W. Development and characterization of new microsatellite loci from Chinese mitten crab (*Eriocheir sinensis*). *Conserv. Genet.* **2009**, *10*, 1117–1119.
24. Chang, Y.M.; Liang, L.Q.; Li, S.W.; Ma, H.T.; He, J.G.; Sun, X.W. A set of new microsatellite loci isolated from Chinese mitten crab, *Eriocheir sinensis*. *Mol. Ecol. Notes* **2006**, *6*, 1237–1239.

25. Li, R.; Yu, C.; Li, Y.; Lam, T.-W.; Yiu, S.-M.; Kristiansen, K.; Wang, J. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics (Oxford)* **2009**, *25*, 1966–1967.
26. Cheng, Q.; Yuan, C.; Wang, J.; Xu, J.; Lee, T.-H.; Wang, C. Development of 20 microsatellite loci in the Japanese mitten crab *Eriocheir japonica* and cross-amplification in the Chinese mitten crab *Eriocheir sinensis*. *Conserv. Genet. Resour.* **2010**, *2*, 47–50.
27. Bachtrog, D.; Agis, M.; Imhof, M.; Schlotterer, C. Microsatellite variability differs between dinucleotide repeat motifs: Evidence from *Drosophila melanogaster*. *Mol. Biol. Evol.* **2000**, *17*, 1277–1285.
28. Lecher, P.; Defaye, D.; Noel, P. Chromosomes and nuclear DNA of Crustacea. *Invertebr. Reprod. Dev.* **1995**, *27*, 85–114.
29. Kong, J.; Gao, H.A. Analysis of tandem repeats in the genome of Chinese shrimp *Fenneropenaeus chinensis*. *Chin. Sci. Bull.* **2005**, *50*, 1462–1469.
30. Schug, M.D.; Wetterstrand, K.A.; Gaudette, M.S.; Lim, R.H.; Hutter, C.M.; Aquadro, C.F. The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Mol. Ecol.* **1998**, *7*, 57–70.
31. Edwards, Y.J.K.; Elgar, G.; Clark, M.S.; Bishop, M.J. The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish, *Fugu rubripes*: Perspectives in functional and comparative genomic analyses. *J. Mol. Biol.* **1998**, *278*, 843–854.
32. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; *et al.* Initial sequencing and analysis of the human genome. *Nature (Lond.)* **2001**, *409*, 860–921.
33. Wang, Z.; Weber, J.L.; Zhong, G.; Tanksley, S.D. Survey of plant short tandem DNA repeats. *Theor. Appl. Genet.* **1994**, *88*, 1–6.
34. Van Belkum, A.; Scherer, S.; van Alphen, L.; Verbrugh, H. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 275–293.
35. Ross, C.L.; Dyer, K.A.; Erez, T.; Miller, S.J.; Jaenike, J.; Markow, T.A. Rapid divergence of microsatellite abundance among species of *Drosophila*. *Mol. Biol. Evol.* **2003**, *20*, 1143–1157.
36. Lagercrantz, U.; Ellegren, H.; Andersson, L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* **1993**, *21*, 1111–1115.
37. Toth, G.; Gaspari, Z.; Jurka, J. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **2000**, *10*, 967–981.
38. Chang, Y.; Liang, L.; Ma, H.; He, J.; Sun, X. Microsatellite analysis of genetic diversity and population structure of Chinese mitten crab (*Eriocheir sinensis*). *J. Genet. Genomics* **2008**, *35*, 171–176.
39. Herborg, L.-M.; Weetman, D.; Van Oosterhout, C.; Hanfling, B. Genetic population structure and contemporary dispersal patterns of a recent European invader, the Chinese mitten crab, *Eriocheir sinensis*. *Mol. Ecol.* **2007**, *16*, 231–242.
40. Sambrook, J.; Russell, D.W. *Molecular Cloning: A Laboratory Manual*; CSHL Press: Cold Spring Harbor, NY, USA, 2001; Volume 2.
41. Pevzner, P.A.; Tang, H.X.; Waterman, M.S. An Eulerian path approach to DNA fragment assembly. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9748–9753.

42. Rozen, S.; Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **2000**, *132*, 365–386.
43. Jurka, J.; Pethiyagoda, C. Simple receptive DNA sequences from primates: Compilation and analysis. *J. Mol. Evol.* **1995**, *40*, 120–126.
44. Bassam, B.J.; Caetanoanollles, G.; Gresshoff, P.M. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **1991**, *196*, 80–83.
45. Excoffier, L.; Lischer, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567.
46. Rousset, F. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **2008**, *8*, 103–106.
47. Van Oosterhout, C.; Hutchinson, W.F.; Wills, D.P.M.; Shipley, P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **2004**, *4*, 535–538.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).