Chinese Journal of Oceanology and Limnology
Vol. 35 No. 1, P. 203-214, 2017
http://dx.doi.org/10.1007/s00343-016-5250-7

Genome-wide mining, characterization, and development of microsatellite markers in *Marsupenaeus japonicus* by genome survey sequencing*

LU Xia (卢霞)[1], LUAN Sheng (栾生)[1], KONG Jie (孔杰)[1, **], HU Longyang (胡龙洋)[1], MAO Yong (毛勇)[2], ZHONG Shengping (钟声平)[2]

[1] *Key Laboratory for Sustainable Utilization of Marine Fisheries Resources*, *Ministry of Agriculture*, *Yellow Sea Fisheries Research Institute*, *Chinese Academy of Fishery Sciences*, *Qingdao 266071*, *China*

[2] *College of Ocean and Earth Sciences*, *Xiamen University*, *Xiamen 361005*, *China*

Received Oct. 8, 2015; accepted in principle Nov. 4, 2015; accepted for publication Nov. 10, 2015
© Chinese Society for Oceanology and Limnology, Science Press, and Springer-Verlag Berlin Heidelberg 2017

Abstract The kuruma prawn, *Marsupenaeus japonicus*, is one of the most cultivated and consumed species of shrimp. However, very few molecular genetic/genomic resources are publically available for it. Thus, the characterization and distribution of simple sequence repeats (SSRs) remains ambiguous and the use of SSR markers in genomic studies and marker-assisted selection is limited. The goal of this study is to characterize and develop genome-wide SSR markers in *M. japonicus* by genome survey sequencing for application in comparative genomics and breeding. A total of 326 945 perfect SSRs were identified, among which dinucleotide repeats were the most frequent class (44.08%), followed by mononucleotides (29.67%), trinucleotides (18.96%), tetranucleotides (5.66%), hexanucleotides (1.07%), and pentanucleotides (0.56%). In total, 151 541 SSR loci primers were successfully designed. A subset of 30 SSR primer pairs were synthesized and tested in 42 individuals from a wild population, of which 27 loci (90.0%) were successfully amplified with specific products and 24 (80.0%) were polymorphic. For the amplified polymorphic loci, the alleles ranged from 5 to 17 (with an average of 9.63), and the average PIC value was 0.796. A total of 58 256 SSR-containing sequences had significant Gene Ontology annotation; these are good functional molecular marker candidates for association studies and comparative genomic analysis. The newly identified SSRs significantly contribute to the *M. japonicus* genomic resources and will facilitate a number of genetic and genomic studies, including high density linkage mapping, genome-wide association analysis, marker-aided selection, comparative genomics analysis, population genetics, and evolution.

Keyword: *Marsupenaeus japonicus*; genome-wide SSR markers; genome survey sequencing; functional annotation

1 INTRODUCTION

Microsatellites, also known as simple sequence repeats (SSRs), are randomly repeated DNA sequences composed of 1–6 bp long units (Tautz and Renz, 1984). They are ubiquitous in prokaryote and eukaryote genomes, present in both protein encoding and noncoding regions (Tóth et al., 2000; Subramanian et al., 2003; Hoffman and Nichols, 2011). SSRs can generate and maintain extensive length polymorphism because they have a high enough mutation rate of between 10^{-3} and 10^{-4} mutations per gamete per generation (Weber and Wong, 1993; Bhargava and Fuentes, 2010), which makes them important for genome evolution (Tautz et al., 1986; Kashi et al., 1997). Additionally, because of their locus-specificity, co-dominant inheritance, reproducibility, and abundance SSR markers have become powerful molecular tools for a wide range of applications in genetic analysis and breeding, such as population genetics, parentage

* Supported by the National High Technology Research and Development Program of China (863 Program) (No. 2012AA10A409)
** Corresponding author: kongjie@ysfri.ac.cn
LU Xia and LUAN Sheng contributed equally to this work.

analysis, genetic mapping, association mapping, comparative mapping, and quantitative trait loci (QTL) analysis (Bruford and Wayne, 1993; Jarne and Lagodav, 1996; Cruz et al., 2005; Varshney et al., 2005; Bohra et al., 2011). Although the advent of SNP markers has attracted more attention and funding for their development and applications in various animal and plant genera, many studies have indicated that SSR markers are still being extensively developed as an irreplaceable molecular tool for various purposes (Barchi et al., 2011; Ansari et al., 2013; Barzegar et al., 2013; Doulati-Baneh et al., 2013; Guo et al., 2013; Liu et al., 2013; Smee et al., 2013; Yuan et al., 2013; Zitouna et al., 2013; Vukosavljev et al., 2013; Lee et al., 2014; Parobek et al., 2014).

However, conventional experimental approaches to develop SSRs for non-model species are laborious, time consuming, and expensive (Zane et al., 2002). Fortunately, recent developments in DNA sequencing technology and the falling costs of next-generation sequencing (NGS) have provided novel methods for SSR identification with high efficiency and lower costs (Castoe et al., 2010; Jennings et al., 2011; Triwitayakorn et al., 2011; Wang et al., 2011, 2013; Nybom et al., 2014; Tadano et al., 2014; Králová-Hromadová et al., 2015). Genome survey sequencing (GSS) based on the NGS platform, proven particularly useful in identifying genome-wide SSRs in non-model species with the advantages of low costs and high throughput (Miller et al., 2007; Barchi et al., 2011; Rowe et al., 2011; Zhou et al., 2013; Xu et al., 2014). Previous studies have demonstrated that SSR markers distributed in non-coding regions have practical implications (Hancock, 1995; Hosseini et al., 2008). Although genic-SSRs may have an insufficient degree of polymorphism for genetic analysis because they are often more conserved, SSR markers developed in functional genes can also be useful for association analysis and marker-aided selection (Cheng et al., 2007; Shikano et al., 2010; Guichoux et al., 2011). The GSS approach could solve the above issue, because it is not only productive and efficient for identifying SSRs in non-coding regions, but it also efficiently predicts putative gene functions and identifies potential exon-intron boundaries because it possesses high similarity and long sequence lengths (Strong and Nelson, 2000).

The kuruma prawn, *Marsupenaeus japonicus*, is one of the most cultivated and consumed shrimps with a native range encompassing an Indo-West Pacific distribution from Japan, through Southeast Asia to East Africa and the Red Sea (Holthuis, 1980; Hewitt and Duncan, 2001). However, very few molecular genetic/genomic resources are publically available for *M. japonicus*. The characterization and distribution of SSRs remains ambiguous for this species, and the use of SSR markers in genomic studies and marker-assisted selection is limited. Consequently, the aim of the present study was to characterize and develop genome-wide SSR markers in *M. Japonicus* by GSS. The newly identified SSRs would be useful for extending our current knowledge of *M. japonicus* genome organization and for genome mapping, genome-wide association studies, marker-aided selection, comparative genomic analysis, and population genetics.

2 MATERIAL AND METHOD

2.1 Genome survey sequencing

An *M. japonicus* individual, the maternal parent of an F_1 mapping population, was chosen for genome survey sequencing. Genomic DNA was extracted from its muscles following the methods described by Wang et al. (2006). Two paired-end DNA libraries were constructed with insert sizes of 180 and 500 bp, and then sequenced using the Illumina HiSeq2000 platform following the manufacturer's protocol, which has been described by Etter et al. (2011).

Because Illumina HiSeq2000 produces some low-quality reads, a quality trim was performed on the first raw reads to make the further analysis more accurate, as follows: (1) removal of adapter contamination; (2) removal of bases that were not A, G, C, and T from the beginning of 5′ end; (3) trimming the ends of the low-quality reads (sequencing quality value <20); (4) removal of reads with more than 10% "N" bases; (5) abandon reads <100 bp following adapter removal and quality trimming. The clean reads were assembled into contigs in SOAPdenove (v2.04) with a K-mer of 17 by applying the *de Bruijn* graph structure (Jiao et al., 2014). The paired-end information was then used to join the unique contigs into scaffolds. After that, paired-end extracted reads that had one read uniquely aligned on a contig and another read located in a gap region were used to fill the intra-scaffold gaps (Jiao et al., 2014). The details for the SOAPdenove assembly are described in Li et al. (2010). Repetitive element occurrence was detected with CENSOR, which screens query sequences against a reference collection of repeats (http://www.girinst.org/censor) (Kohany et al., 2006; Bohra et al., 2014), adopting default parameters and using Viridiplantae as the target database.

2.2 Microsatellite mining and primer design

To identify SSR markers, the assembled scaffolds were screened for SSRs in SciRoKo (Kofler et al., 2007). To assess the distribution and characterization of microsatellite abundance, the mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs were screened from the scaffolds. In the search for an SSR standard, we defined SSRs as mononucleotides ≥10 bases, dinucleotides ≥12 bases, trinucleotides ≥15 bases, tetranucleotides ≥20 bases, pentanucleotides ≥25 bases, and hexanucleotides ≥30 bases (Cardle et al., 2000); the maximum number of bases interrupting two SSRs in a compound microsatellite was 100 bases. Additionally, because DNA is double stranded and an SSR start site could be arbitrarily chosen, reverse-complement repeat motifs and translated or shift motifs were also considered and grouped together in the analysis (e.g., AG/CT representing AG and GA, TC and CT) (Jurka and Pethiyagod, 1995). The software Websat (http://wsmartins.net/websat/) was used to design SSR primers (Martins et al., 2009). Primer pairs were designed to meet the following restrictions: the target amplification size range was 100–500 bp, the primer annealing temperature was restricted to 55–60°C and the primer length was 19–27 bp. We termed such loci with PCR primers as potential microsatellite markers (Potentially Amplifiable Loci or PAL).

2.3 Functional annotation of genome-wide SSR markers

For homology annotation of all of the SSR-containing sequences, non-redundant sequences were subjected to the public database GO (Gene Ontology, http://www.geneontology.org/). The sequence annotation was based on sequence similarity according to gene ontology (Barchi et al., 2011) with the annotation parameters at an E-value cut-off of 10^{-10}.

2.4 Laboratory verification of the SSR markers in a natural population

We randomly selected 30 primer pairs from the microsatellite loci with relatively longer repeats for laboratory verification, containing six di-, tri-, tetra-, penta-, and hexanucleotide loci. Forty-two *M. japonicus* individuals collected from a natural population on the coast of Fujian Province (China) were used for verification of the SSR primers. The total genomic DNA was also extracted from muscle tissue following the methods described by Wang et al. (2006). The PCR reactions were conducted in a 10 μL PCR mix, containing 1×PCR buffer, 1.5 mmol/L gCl_2, 0.2 mmol/L dNTPs, 0.2 U Taq DNA polymerase (Promega, USA), approximately 20 ng of genomic DNA, and 0.25 μmol/L of each primer pair. PCR amplification was performed in a MJ-PCR-200 thermal cycler (Bio-Rad, USA) with the following procedure: initial denaturation at 94°C for 5 min, 35 cycles of denaturation at 94°C for 30 s, annealing temperature for 30 s (Table 1), and extension at 72°C for 40 s, followed by a final step at 72°C for 10 min. The PCR products were evaluated by 20 g/L agarose gel electrophoresis and then loaded into an Applied Biosystems 3130 sequencer and scored using GeneMapper version 3.7 software (Applied Biosystems) with GeneScan-500 LIZ Size Standard as an internal size standard. For each polymorphic SSR locus amplified, polymorphic information content (PIC) was calculated with the following formula:

$$\mathrm{PIC} = 1 - \sum p_i^2 ,$$

where, p_i s the proportion of the ith allele (Biswas et al., 2014).

3 RESULT

3.1 Genome survey sequencing

Genome survey sequencing was performed for a single *M. japonicus* individual by sequencing two paired-end DNA libraries with insert sizes of 180 and 500 bp based on the Illumina HiSeq2000 platform. All of the raw data are accessible in the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the accession number SRX1162702. According to the Lander–Waterman algorithm, the *M. japonicus* genome size was calculated by the following equation:

$$G=k_{\mathrm{num}}/k_{\mathrm{depth}}=b_{\mathrm{num}}/b_{\mathrm{depth}},$$

where, k_{num} is the number of K-mer, k_{depth} is the expected depth of K-mer, b_{num} is the number of bases, and b_{depth} is the expected depth of bases. According to the equation, the size *M. japonicus* genome was estimated as ~2.4 Gb. In total, 79.75 Gb of sequencing data were produced, equivalent to ~33.23 × genome coverage (Table 2). The assembled data with a K-mer of 29 was selected for the final results according to the comprehensive evaluation of technique indexes (such as the length of all the scaffolds, number of scaffolds, and scaffold N50). The assembly statistics of the genome survey sequencing data is summarized in Table 3. A total of 3 244 404 scaffolds were assembled

Table 1 Characteristics of 30 microsatellite primers for lab verification in *M. japonicus*

Locus (scaffold)	Primer sequences		Repeat unit	T_a (°C)	Allele	PIC
	Forward primers (5′→3′)	Reverse primers (5′→3′)				
573703	CGAGGGAGCAACTTAGCTGT	GATCCCTAAACACCTCGGCC	$(GA)_{55}$	57	10	0.817
563220	CCTCCCTCCCACACTCTTTC	CCTCCCTCCCACACTCTTTC	$(AG)_{49}$	57	8	0.781
38161	GGAGGAGGGAATATGGCACG	GCAACTACGCTTCGCTTACG	$(TG)_{30}$	56	7	0.753
918	TTCGACACTGCACGAACACT	TCAATTTTGCCCGACAGCTT	$(TG)_{29}$	57	8	0.781
2386	GGGACCAAGCTCAGATCGAG	CTGTGCTTGGTTACTCCCGT	$(CA)_{29}$	-	-	-
3048	CTCGATCTCCATTCCGTGCA	TCACGCTTCACTCACTGTCC	$(AC)_{29}$	58	10	0.803
2145	TGTGTGCACATGTAGGCCTT	CACACACATATGCATGCCCA	$(TGA)_{20}$	57	9	0.791
22166	CCCTTTCCCCACTCCTTTCC	CTCCATATGCATGCCCCCTC	$(AAG)_{20}$	58	13	0.839
61496	CCTTCTTCCTCCTTCCTCGC	AGCGTGGCTCTCGTATTGTT	$(TAC)_{20}$	57	7	0.742
197436	ACCAATGACACACGACGTCA	AGACAAGAGAAATAAACCGACCGA	$(TCT)_{20}$	58	10	0.810
218161	TTGCGAAAACAATGGGGCAG	CAGGACGGGTTCACTTCACA	$(GAT)_{20}$	57	9	0.782
218927	GCGGAAGGGATTTACCACCA	GCCTCGCCGTTGACTTTTTA	$(AGA)_{20}$	58	8	0.773
290335	GAGGGTAATGAGCCAACCCC	TCTGCCTGCCTGTCAATTGT	$(GATA)_{21}$	57	11	0.827
342070	ACATAAACAAAACGCACACAAAGA	TCACTGACACGGTGACCTTA	$(ATAC)_{19}$	57	5	0.692
360252	GTTCCTCTTCTCTTCGCCGT	CCCTGACGCTCTGAAATTCG	$(TGTC)_{19}$	58	8	0.765
190847	TGCTGTCTTGCTGTCTGTGT	CGGGATTATAAAGCACGGCG	$(CTGT)_{16}$	-	-	-
154340	TGAATATGACCACGCACAGGT	CTTTCCGTTCGACAGACACG	$(GAGT)_{15}$	58	9	0.801
189183	AGGAGGGAGGAGGATAGGGA	ATGCACGTTCAAGCACACAT	$(GACA)_{15}$	-	-	-
16631	CTACCCACCCACAGTCACAC	AGGAGGCGTCAACAAGAACA	$(AAGCA)_{12}$	58	10	0.823
430289	AGGGTAGGCGTGGTTTGAAT	AGGTTGCCAGCTCTAACCAC	$(GTTAG)_{12}$	-	-	-
447545	CTCGCCAGTGTCTGTCCTTT	AAAGACGAACGGCGGGAAAT	$(CTCTT)_{12}$	58	5	0.702
474174	ACATCCGAGCTTGGTAACGT	CCCATTGCTGTCTTGTATTGCC	$(TGTTA)_{12}$	58	6	0.723
572627	GACTCGACCGAAACCGAACA	CGGAGCCACGTGTTGTTAGT	$(CCGTC)_{12}$	58	17	0.912
801323	TCCTTCGAGCACTTTGCCAT	ATCTTGGTGGCAGTGAACGT	$(AGTTC)_{12}$	58	14	0.865
219076	CGTTTCCCTCACCCTCTTCC	AATGGAGAGGGAAGGGGTGA	$(CTCACT)_{11}$	-	-	-
268119	GATGAGCTTGTCCCCTGCTT	AGGTAGGTGATGCTGCAACC	$(CCCTCG)_{11}$	58	14	0.871
412755	GATACTTCAGGACCTCGCCG	ATACAGAGAGGGGCGTGACT	$(TGGTAC)_{11}$	58	9	0.797
549336	TCTCGCCACCTCTACCGTTA	CGTGTGTGTGCAAGAGTGTG	$(ACACGC)_{11}$	-	-	-
938819	ATTTCCGCTGGGTTGTTCCT	CTCATCGCCATCACCATTGC	$(GATGGC)_{11}$	58	16	0.900
3309	TCAAGTGATTGCCATAACCTCCT	CCCCCACGGCATTAATCACT	$(GTTGTA)_{10}$	58	8	0.758
Mean					9.63	0.796

Table 2 Summary of *M. japonicus* genome survey sequencing data

Paired-end libraries	Read length (bp)	Insert size (bp)	Total data (G)	Sequencing depth (×)
L1	101	180	71.28	29.70
L2	101	500	8.47	3.53
Total				33.23

from the genome survey sequencing data, among which there were 127 538 ones larger than 1 000 bp with the largest length of 15 860 bp.

3.2 SSR distribution in the *M. japonicus* genome

The general information on the SSR-containing sequences is summarized in Table 4. The total number of scaffold sequences with SSRs was 289 962, among which 126 210 contained more than one SSR. The lengths of SSR-containing sequences ranged from 201 bp to 15 960 bp, with an average of 468.1 bp. A total of 396 061 SSRs were detected, including 69 116 present in compound formation. Among the 326 945 perfect SSRs, dinucleotide repeats were dominant (44.08%), followed by mononucleotide repeats

(29.67%), trinucleotide repeats (18.96%), and tetranucleotides (5.66%). Interestingly, we found that the hexanucleotide repeats (1.07%) were two-fold higher than pentanucleotide repeats (0.56%) (Fig.1).

A more detailed investigation of individual repeat types was performed and is presented in Fig.2. The results showed that the AT-rich repeats were more abundant than GC-rich repeats. The most frequent mono-, di-, and trinucleotide repeat motifs were A (88.05%), AG (54.12%), and AAT (29.14%), respectively (Fig.2a); however, the motifs G (11.95%), CG (0.21%), and CCG (0.49%) were very rare. AT-rich motifs were also predominant in tetra-, penta-, and hexanucleotide repeats (Fig.2b), e.g., AAAC and AAAG were the most common tetranucleotides and

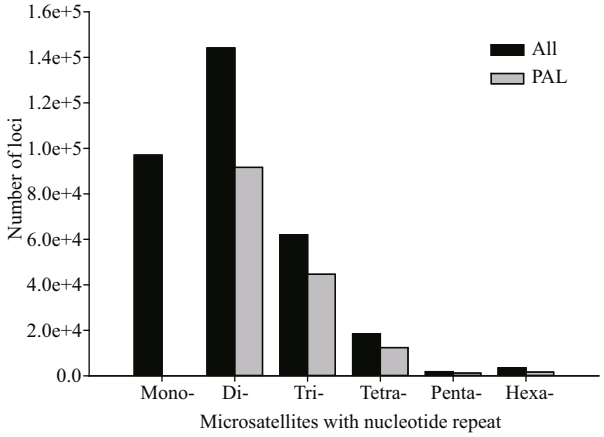

Fig.1 Frequency distribution of different SSR repeat units and the PAL subset identified in *M. japonicus* scaffolds

PAL: potentially amplifiable loci, were SSR loci with successfully designed primers.

Table 3 *M. japonicus* genome assembly statistics

K-mer=29	
No. of all scaffolds	**3 244 404**
Bases in all scaffolds	1 366 164 721 bp
No. of large scaffolds (>1 000 bp)	127 538
Largest length	15 960 bp
Scaffold N50	1 337 bp
Scaffold N90	1 047 bp
G+C content	38.213%
No. of all contigs	**3 343 150**
Bases in all contigs	1 362 806 802
No. of large contigs (>1 000 bp)	113 071
Largest length	15 430 bp
Contig N50	1 327 bp
Contig N90	1 046 bp

Table 4 SSR characteristics in the *M. japonicus* genome assembly

SSR mining	Total
Number of SSR-containing sequences	289 962
The longest length of SSR-containing sequences (bp)	15 960
The shortest length of SSR-containing sequences (bp)	201
The average length of SSR-containing sequences (bp)	468.1
Number of scaffolds with more than one SSR	126 210
Number of identified SSRs	396 061
Number of perfect SSR Loci	326 945
Number of SSRs present in compound formation	69 116

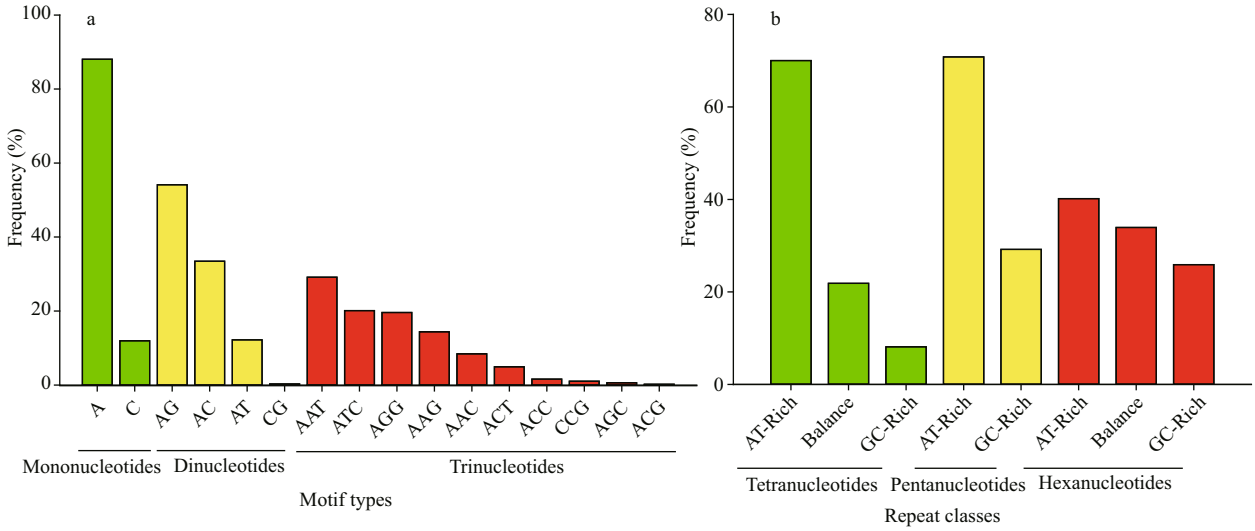

Fig.2 Frequency distribution of (a) two, four, and ten mono-, di-, and trinucleotides motifs, respectively and (b) AT-rich, AT-GC-balance and GC-rich tetra-, penta-, and hexanucleotides motifs in *M. japonicus*

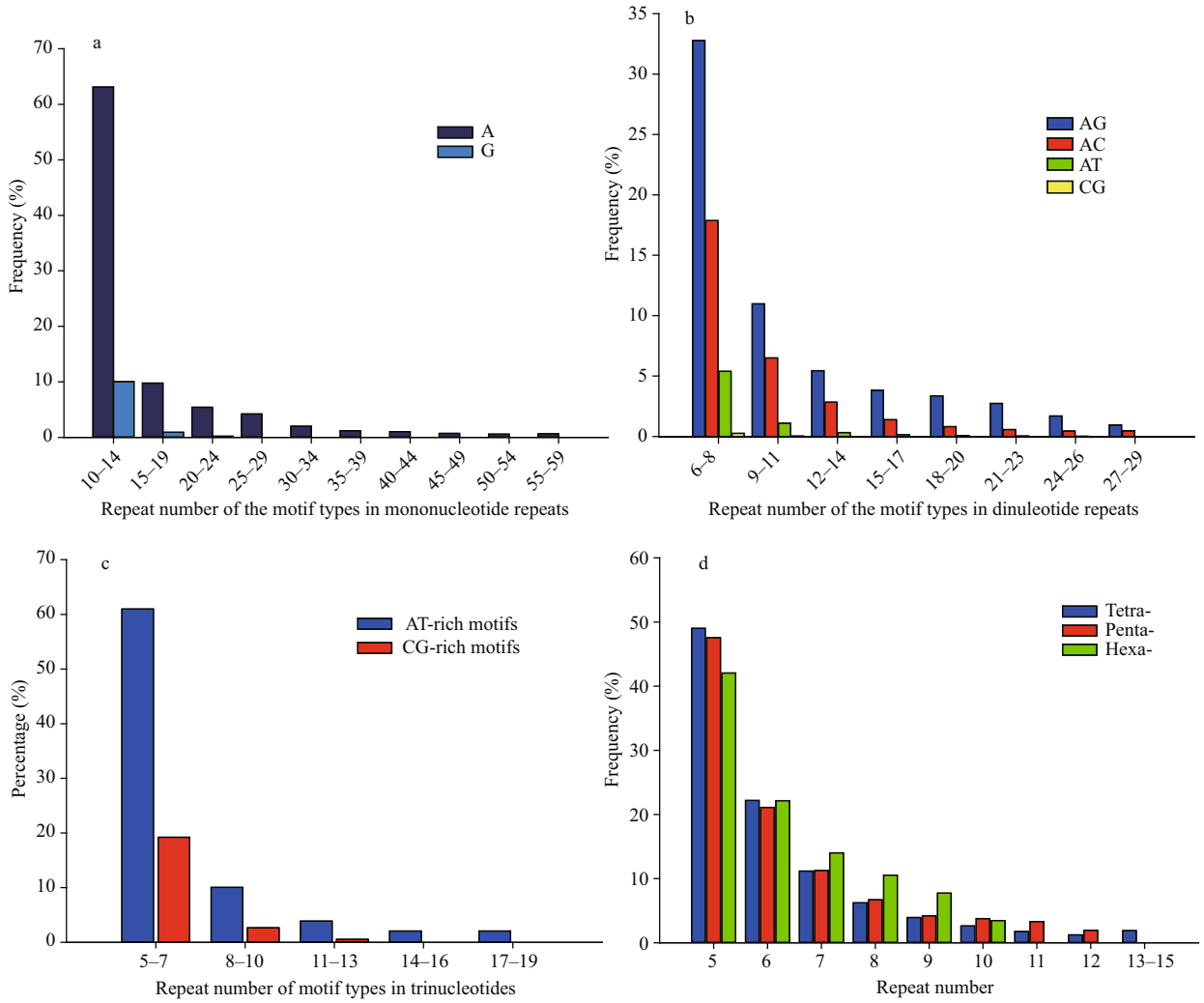

Fig.3 Frequency distribution of *M. japonicus* SSRs by number of repeats for (a) two mononucleotide motifs, (b) four dinucleotide motifs, (c) AT- and GC-rich trinucleotide motifs, and (d) tetra-, penta-, and hexanucleotides

Table 5 Repeat number distribution for all trinucleotide repeat motifs in *M. japonicus*

	Range of repeat number				
	5–7	8–10	11–13	14–16	17–19
Motifs	Percent (%)	Percent (%)	Percent (%)	Percent (%)	Percent (%)
AAT/ATT	21.031	3.947	1.922	0.905	0.291
AGG/CCT	16.624	2.497	0.519	0.007	0.000
AAG/CTT	13.880	2.750	0.940	0.570	0.390
ACT/ATG	10.198	1.440	0.474	0.302	0.132
AGT/ATC	9.873	1.431	0.485	0.262	0.139
AAC/GTT	6.007	0.486	0.063	0.019	0.013
ACC/GGT	1.607	0.095	0.017	-	-
CCG/CGG	0.427	0.050	0.008	-	-
ACG/CTG	0.277	0.021	-	-	-
AGC/CGT	0.276	-	-	-	-

AAGAG and AAAAG were the dominant hexanucleotides. However, the hexanucleotide repeats had a relatively lower degree of preference for A/T than the motifs of other sizes (Fig.2b).

3.3 SSR distribution in the *M. japonicus* genome

The number of repeat motifs within the same and among different size units were distributed differently. The details of the repeat number distributions of different size units (mono-, di-, tri-, tetra-, penta-, and hexanucleotides) are listed in Fig.3 and Table 5. The results revealed that shorter motifs and AT-rich types had relatively larger repeat numbers. The number A/T in mononucleotide repeats ranged from 10 to 59, but that of C/G ranged from 10 to 26, and the most common repeat number (10–12) accounted for 62.81% of the total mononucleotide repeats (Fig.3a). The number AG/CT and AC/GT repeats ranged from

6 to 29, but that of CG/CG ranged from 6 to 7, and the most common repeat number (6–9) accounted for 64.69% of the total dinucleotide repeats (Fig.3b). The number of AAT/ATT repeats ranged from 5 to 16, but CCG/CGG was mainly distributed in repeats of five (Table 4). The most common repeat number (5–7) accounted for 80.11%, 82.36%, 79.92%, and 78.23% of the total tri-, tetra-, penta-, and hexanucleotide repeats, respectively (Fig.3c, d).

3.4 Functional annotation of SSRs

Among all of the SSR-containing sequences, 58 256 exhibited a significant GO annotation with three categories. The biological process category covered most of the functional genes (51.12%), followed by the molecular function (24.70%) and cellular component (24.18%) categories. In the biological process category, the dominant subcategories were metabolic process (19.44%), single-organism process (17.40%), and cellular process (17.22%) (Fig.4a). Among the molecular function terms, a main proportion of clusters assigned to binding (42.15%) and catalytic activity (37.16%) (Fig.4b). In the cellular component category, the major subcategories were membrane (22.27%), cell (16.02%), and cell part (15.83%) (Fig.4c).

3.5 Development and laboratory verification of genome-wide SSR markers

In the present study, primers were designed for the di- to hexanucleotide repeats to develop genome-wide SSR markers in *M. japonicus*. With the exceptions of compound repeats and mononucleotide loci, primers were successfully designed for 63.60%, 72.05%, 67.03%, 64.46%, and 46.74% of the di-, tri-, tetra-, penta-, and hex-nucleotide loci, respectively, proving themselves to be promising candidates for either PCR amplification or PAL (Fig.1).

A subset of 30 SSR primer pairs were synthesized and tested in 42 individuals from a natural population to assess the PAL availability rate. Among the primer pairs tested, 90.0% could be amplified with prominent PCR products of the expected size and 80.0% were polymorphic (Table 1). The hexanucleotide (47 alleles for four loci with an average of 11.75) and pentanucleotide (52 alleles for five loci with an average of 10.40) exhibited the highest polymorphism, followed by trinucleotide (56 alleles for six loci with an average of 9.33) and dinucleotide loci (43 alleles for five loci with an average of 8.60). However, tetranucleotide loci exhibited the lowest polymorphism (33 alleles for four loci with an average of 8.25). A total of 231 alleles were recorded from the 24 SSR loci, and the number of alleles ranged from 5 to 17 (with an average of 9.63). The PIC values for the polymorphic SSR loci ranged from 0.692 to 0.912 with an average of 0.796.

4 DISCUSSION

Despite their abundance and wide geographical distribution, little effort has been devoted to decoding the *M. Japonicus* genome, which has limited the use of SSR markers in genomic studies and marker-assisted selection. In the present study, we analyzed the genome-wide SSR distribution and frequency of 1 to 6 bp sequences in the *M. japonicus* genome via GSS. Our study represents the first step towards decoding the *M. japonicus* genome, and also enables the development of SSR markers for genetic analysis. An extensive set of 326 945 perfect SSRs were detected without previously knowing anything about the genome sequence. As expected, the majority of the SSR markers had no Gene Ontology assignment because the SSRs were from *M. japonicus*, but a total of 58 256 of the SSR-containing scaffold sequences still had significant GO hits. The newly identified SSRs would be useful for extending our current knowledge of *M. japonicus* genome organization, and for genetic mapping, genome-wide association studies, marker-aided selection, comparative genomic analyses, and population genetics.

In the kuruma prawn, dinucleotide repeats were dominant, which was consistent with previous studies (Tóth et al., 2000; Somridhivej et al., 2008; Wang et al., 2011; Xu et al., 2011; Iranawati et al., 2012; Ji et al., 2012; Thanh et al., 2014). However, some studies have reported that the most abundant repeat motifs were trinucleotides (Stàgel et al., 2008). The higher frequency of trinucleotides in other studies might have been because they were identified from EST sequences, trinucleotide repeats prevail in the protein-coding exons of all taxa (Tóth et al., 2000). This difference might also have arisen from using different species, search parameters, and algorithms (Cavagnaro et al., 2010; Biswas et al., 2014). Interestingly, the *M. japonicus* genome had particularly high counts of hexanucleotide compared to pentanucleotide repeats, which was consistent with other studies (Tóth et al., 2000; Xu et al., 2011; Castoe et al., 2012; Vukosavljev et al., 2012). The higher counts of hexanucleotides might be because they were selected and changed during the process of

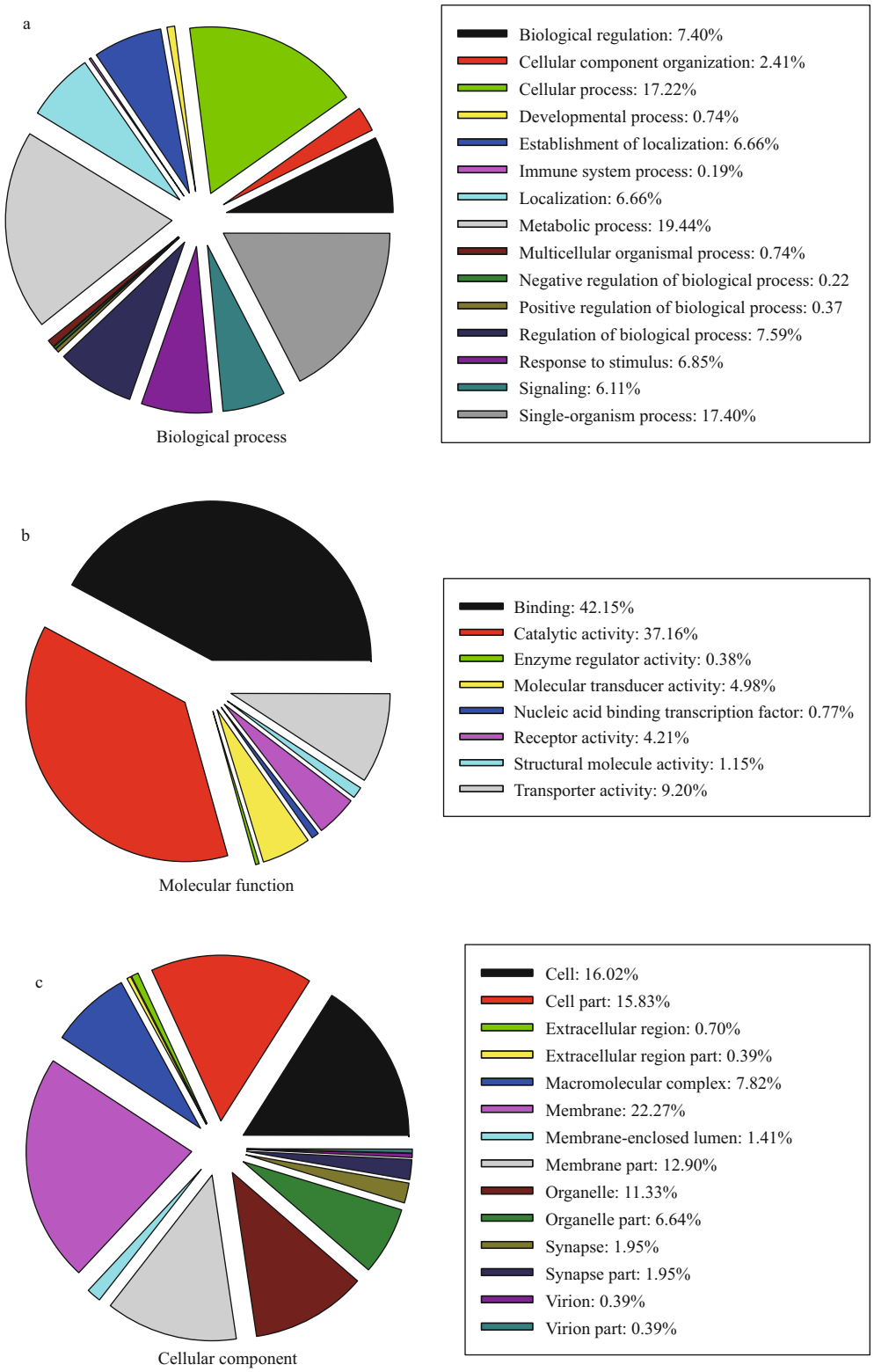

Fig.4 Gene Ontology annotation of genome-wide SSR-containing sequences developed in *M. japonicas*

a. biological process; b. molecular function; c. cellular component.

evolution, because variation in hexanucleotides leads to increased morphological and functional proteins (Wahba et al., 1963; Touriol et al., 2003; Turanov et al., 2009).

As shown in Fig.2, there is a remarkable variation in the frequency of individual repeat patterns in *M.*

japonicus. The base composition of SSR patterns in *M. japonicus* is strongly biased toward A/T in all unit sizes, which may be the result of a high A+T content (61.79%) in the *M. japonicus* genome. Additionally, Tóth et al. (2000) also reported that poly (A/T) tracts were more abundant than poly (C/G) sequences in some species (e.g., primates, rodents, other mammals, non-mammalian vertebrates, arthropods, *Caenorhabditis elegans*, plants, yeast, and other fungi). The most abundant dinucleotide motif in *M. japonicus* was AG (61.74%), which was consistent with previous studies (Nagy et al., 2007; Zeng et al., 2010). However, this contradicted observations in other reports, where AT was the predominant motif (Shirasawa et al., 2010; Barchi et al., 2011; Wang et al., 2011). AAT was the dominant trinucleotide motif in *M. japonicus*, which was consistent with the research of Shirasawa et al. (2010) and Tóth et al. (2000); however, it was in contrast to other studies, where AAC was the predominant motif (Barchi et al., 2011; Wang et al., 2011). The same situation was detected for tetra-, penta-, and hex-nucleotide repeats. Large variation in motif abundance among species has also been reported in other studies (Cruz et al., 2005; Tanguy et al., 2008). One of the reasons for the differences in dominant motifs might be that they were from different resources (Genome vs EST), because characteristic differences between species can be observed in intergenic regions and introns (Tóth et al., 2000).

The common trends in this study were, that AT-rich motifs were dominant and the CG-rich motifs, such as CG (2.12%), CCG (1.02%), and CCCG (0.070%), were very rare in the *M. japonicus* genome, which has also been reported in other species (Tóth et al., 2000; Kumpatla and Mukhopadhyay, 2005; Barchi et al., 2011; Wang et al., 2011). Additionally, the AT-rich motifs had relatively longer repeat lengths than CG-rich motifs (Fig.3). One of the main causes for the rarity and short length of the CG-rich motifs might have been that the three hydrogen bonds between C/G enable more stability and reduce the chance of DNA polymerase slippage (Tóth et al., 2000). Another reason might be that CG-rich motifs are common in exons with less variance, whereas other regions have fewer of them (Tóth et al., 2000; Subramanian et al., 2003). Moreover, CG-rich repeats might form higher structures such as hairpin loops, which might stabilize the slipped structure (Schlötterer and Tautz, 1992; Sinden, 1999). Importantly, the variation in SSR lengths may be a good indication of polymorphism.

Because mononucleotide repeats are not suitable for marker development (Cavagnaro et al., 2010), only di- to hexanucleotide repeats were considered for primer design; primers were successfully designed for 151 541 SSR loci. We obtained a high PCR amplification efficiency and a high level of polymorphic loci in this study (90% of primer pairs yielded specific products, 80% of which were polymorphic), which likely resulted from the long repeats (Table 1). Previous studies have indicated that the SSR loci with long repeats are more prone to mutations both in plants and animals (Cavagnaro et al., 2010; Biswas et al., 2014). Our results also allowed us to understand the possible relationships between the degree of polymorphism and particular features of microsatellites, from which we found that hex- and pentanucleotides are more polymorphic than tetranucleotide repeats. These results might have been biased by the selection of the SSR loci analyzed, because we selected loci with long repeats for verification. In a future study, we will validate more SSR markers for linkage mapping and association analysis. The many polymorphic loci now available could be used for evolutionary and population genetic research, as well as high-resolution chromosome linkage mapping studies of *M. japonicus*.

5 CONCLUSION

Overall, the present study has contributed a detailed characterization and development of genome-wide SSR markers in *M. japonicus* by GSS. The kuruma prawn genome was relatively rich in microsatellites, and dinucleotide repeats and AT-rich SSRs were prevalent. Primers were successfully designed for almost half of the SSRs and a surprisingly high percentage (80%) of the polymorphic markers were verified in the laboratory. A total of 58 256 SSR-containing sequences had significant GO annotation, these are good functional molecular marker candidates for association studies and comparative genomic analysis. The newly identified genome-wide SSR markers enhance *M. japonicus* genomic resources and will facilitate many genetic and genomic studies.

6 ACKNOWLEDGEMENT

We thank Majorbio Pharm Technology Co., Ltd. (Shanghai, China) for assisting with the RAD-Seq technique and Novogene Bioinformatics Technology Co., Ltd. (Beijing, China) for data analysis.

References

Ansari M J, Al-Ghamdi A, Kumar R, Usmani S, Al-Attal Y, Nuru A, Mohamed A A, Singh K, Dhaliwal H S. 2013. Characterization and gene mapping of a chlorophyll-deficient mutant *clm1* of *Triticum monococcum L*. *Biologia Plantarum*, **57**(3): 442-448.

Barchi L, Lanteri S, Portis E, Acquadro A, Valè G, Toppino L, Rotino G L. 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, **12**: 304.

Barzegar R, Peyvast G, Ahadi A M, Rabiei B, Ebadi A A, Babagolzadeh A. 2013. Biochemical systematic, population structure and genetic variability studies among Iranian Cucurbita (*Cucurbita pepo* L.) accessions, using genomic SSRs and implications for their breeding potential. *Biochemical Systematics and Ecology*, **50**: 187-198.

Bhargava A, Fuentes F F. 2010. Mutational dynamics of microsatellites. *Molecular Biotechnology*, **44**(3): 250-266, http://dx.doi.org/10.1007/s12033-009-9230-4.

Biswas M K, Xu Q, Mayer C, Deng X X, Niedz R P. 2014. Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*). *PLoS One*, **9**(8): e104182, http://dx.doi.org/10.1371/journal.pone. 0104182.

Bohra A, Dubey A, Saxena R K, Penmetsa R V, Poornima K N, Kumar N, Farmer A D, Srivani G, Upadhyaya H D, Gothalwal R, Ramesh S, Singh D, Saxena K, Kishor P B K, Singh N K, Town C D, May G D, Cook D R, Varshney R K. 2011. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). *BMC Plant Biology*, **11**: 56.

Bohra A, Jha U C, Kavi Kishor P B, Pandey S, Singh N P. 2014. Genomics and molecular breeding in lesser explored pulse crops: current trends and future opportunities. *Biotechnology Advances*, **32**(8): 1 410-1 428.

Bruford M W, Wayne R K. 1993. Microsatellites and their application to population genetic studies. *Current Opinion in Genetics & Development*, **3**(6): 939-943.

Cardle L, Ransay L, Milbourne D, Macaulay M, Marshall D, Waugh R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, **156**(2): 847-854.

Castoe T A, Polle A W, Gu W J, de Koning A P J, Daza L M, Smith E N, Pollock D D. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. *Molecular Ecology Resource*, **10**(2): 341-347.

Castoe T A, Poole A W, de Koning A P J, Jones K L, Tomback D F, Oyler-McCance S J, Fike J A, Lance S L, Streicher J W, Smith E N, Pollock D D, Hansson B. 2012. Rapid microsatellite identification from illumina paired-end genomic sequencing in two birds and a snake. *PLoS One*, **7**(2): e30953, http://dx.doi.org/10.1371/journal.pone.0030953.

Cavagnaro P F, Senalik D A, Yang L M, Simon P W, Harkins T T, Kodira C D, Huang S W, Weng Y Q. 2010. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics*, **11**: 569.

Cheng L, Liao X, Yu X, Tong J. 2007. Development of EST-SSRs by an efficient FIASCO-based strategy: a case study in rare minnow (*Gobiocyrpis Rarus*). *Animal Biotechnology*, **18**(3): 143-152.

Cruz F, Pérez M, Prese P. 2005. Distribution and abundance of microsatellites in the genome of bivalves. *Gene*, **346**: 241-247.

Doulati-Baneh H, Mohammadi S A, Labra M. 2013. Genetic structure and diversity analysis in *Vitis vinifera* L. cultivars from Iran using SSR markers. *Scientia Horticulturae*, **160**: 29-36.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit R J. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resource*, **11**(4): 591-611.

Guo E M, Cui Z X, Wu D H, Hui M, Liu Y, Wang H X. 2013. Genetic structure and diversity of Portunus trituberculatus in Chinese population revealed by microsatellite markers. *Biochemical Systematics and Ecology*, **50**: 313-321.

Hancock J M. 1995. The contribution of slippage-like processes to genome evolution. *Journal of Molecular Evolution*, **41**(6): 1 038-1 047.

Hewitt D R, Duncan P F. 2001. Effect of high water temperature on the survival, moulting and food consumption of *Penaeus* (*Marsupenaeus*) *japonicus* (Bate, 1888). *Aquaculture Research*, **32**(4): 305-313.

Hoffman J I, Nichols H J. 2011. A novel approach for mining polymorphic microsatellite markers *in silico*. *PLoS One*, **6**: e23283, http://dx.doi.org/10.1371/journal.pone. 0023283.

Holthuis L B. 1980. FAO Species Catalogue. Vol. 1. Shrimps and Prawns of the World. An Annotated Catalogue of Species of Interest to Fisheries. FAO Fisheries Synopsis, No. 125, 1. Food and Agricultural Organization of the United Nations, Rome. 271p.

Hosseini A, Ranade S H, Ghosh I, Khandekar P. 2008. Simple sequence repeats in different genome sequences of Shigella and comparison with high GC and AT-rich genomes. *DNA Sequence*, **19**(3): 167-176, http://dx.doi.org/10.1080/10425170701461730.

Iranawati F, Jung H, Chand V, Hurwood D A, Mather P B. 2012. Analysis of genome survey sequences and SSR marker development for Siamese mud carp, *Henicorhynchus siamensis*, using 454 pyrosequencing. *International Journal of Molecular Sciences*, **13**(12): 10 807-10 827.

Jarne P, Lagodav P J L. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, **11**(10): 424-429.

Jennings T N, Knaus B J, Mullins T D, Haiq S M, Cronn R C. 2011. Multiplexed microsatellite recovery using massively parallel sequencing. *Molecular Ecology Resource*, **11**(6): 1 060-1 067.

Ji P F, Liu G M, Xu J, Wang X M, Li J T, Zhao Z X, Zhang X F, Zhang Y, Xu P, Sun X W, Liu Z J. 2012. Characterization

of common carp transcriptome: sequencing, *de novo* assembly, annotation and comparative genomics. *PLoS One*, **7**(4): e35152.

Jiao W Q, Fu X T, Dou J Z, Li H D, Su H L, Mao J X, Yu Q, Zhang L L, Hu X L, Huang X T, Wang Y F, Wang S, Bao Z M. 2014. High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA Research*, **21**(1): 85-101.

Jurka J, Pethiyagod C. 1995. Simple repetitive DNA sequences from primates: compilation and analysis. *Journal of Molecular Evolution*, **40**(2): 120-126.

Kashi Y, King D, Soller M. 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics*, **13**(2): 74-78.

Kofler R, Schlotterer C, Lelley T. 2007. SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, **23**(13): 1 683-1 685.

Kohany O, Gentles A J, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in repbase: repbase submitter and censor. *BMC Bioinformatics*, **7**: 474.

Králová-Hromadová I, Minárik G, Bazsalovicsová E, Mikulíček P, Oravcová A, Pálková L, Hanzelová V. 2015. Development of microsatellite markers in *Caryophyllaeus laticeps* (Cestoda: Caryophyllidea), monozoic fish tapeworm, using next-generation sequencing approach. *Parasitology Research*, **114**(2): 721-726.

Kumpatla S P, Mukhopadhyay S. 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome*, **48**(6): 985-998.

Lee G A, Sung J S, Lee S Y, Chung J W, Yi J Y, Kim Y G, Lee M C. 2014. Genetic assessment of safflower (*Carthamus tinctorius* L.) collection with microsatellite markers acquired via pyrosequencing method. *Molecular Ecology Resource*, **14**(1): 69-78, http://dx.doi.org/10.1111/1755-0998.12146.

Li R Q, Fan W, Tian G et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature*, **463**(7279): 311-317.

Liu H F, Li S Q, Hu P, Zhang Y Y, Zhang J B. 2013. Isolation and characterization of EST-based microsatellite markers for *Scatophagus argus* based on transcriptome analysis. *Conservation Genetics Research*, **5**(2): 483-485, http://dx.doi.org/10.1007/s12686-012-9833-0.

Martins W S, Lucas D C S, Neves K F S, Bertioli D J. 2009. WebSat-A web software for microsatellite marker development. *Bioinformation*, **3**(6): 282-283.

Miller M R, Dunhamv J P, Amores A, Cresko W A, Johnson E A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**(2): 240-248.

Nagy I, Stágel A, Sasvári Z, Röder M, Ganal M. 2007. Development, characterization, and transferability to other Solanaceae of microsatellite markers in pepper (*Capsicum annuum* L.). *Genome*, **50**(7): 668-688.

Nybom H, Weising K, Rotter B. 2014. DNA fingerprinting in botany: past, present, future. *Investigative Genetic*, **5**: 1, http://dx.doi.org /10.1186/2041-2223-5-1.

Parobek C M, Jiang L Y, Patel J C, Alvarez-Martínez M J, Miro J M, Worodria W, Andama A, Fong S, Huang L, Meshnick S R, Taylor S M, Juliano J J. 2014. Multilocus microsatellite genotyping array for investigation of genetic epidemiology of *Pneumocystis jirovecii*. *Journal of Clinical Microbiology*, **52**(5): 1 391-1 399.

Rowe H C, Renaut S, Guggisberg A. 2011. RAD in the realm of next-generation sequencing technologies. *Molecular Ecology*, **20**(17): 3 499-3 502.

Schlötterer C, Tautz D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, **20**(2): 211-215.

Shikano T, Ramadevi J, Shimada Y, Merilä J. 2010. Utility of sequenced genomes for microsatellite marker development in non-model organisms: a case study of functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *BMC Genomics*, **11**: 334, http://dx.doi.org/10.1186/1471-2164-11-334.

Shirasawa K, Asamizu E, Fukuoka H, Ohyama A, Sato S, Nakamura Y, Tabata S, Sasamoto S, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Isobe S. 2010. An interspecific linkage map of SSR and intronic polymorphism markers in tomato. *Theoretical and Applied Genetics*, **121**(4): 731-739.

Sinden R R. 1999. Biological implications of the DNA structures associated with disease-causing triplet repeats. *American Journal of Human Genetic*, **64**(2): 346-353.

Smee M R, Pauchet Y, Wilkinson P, Wee B, Singer M C, ffrench-Constant R H, Hodgson D J, Mikheyev A S. 2013. Microsatellites for the marsh fritillary butterfly: *de novo* transcriptome sequencing, and a comparison with amplified fragment length polymorphism (AFLP) markers. *PLoS One*, **8**(1): e54721.

Somridhivej B, Wang S L, Sha Z X, Liu H, Quilang J, Xu P, Li P, Hu Z L, Liu Z J. 2008. Characterization, polymorphism assessment, and database construction for microsatellites from BAC end sequences of channel catfish (*Ictalurus punctatus*): a resource for integration of linkage and physical maps. *Aquaculture*, **275**(1-4): 76-80.

Stàgel A, Portis E, Toppino L, Rotino G L, Lanteri S. 2008. Gene-based microsatellite development for mapping and phylogeny studies in eggplant. *BMC Genomics*, **9**: 357.

Strong W B, Nelson R G. 2000. Preliminary profile of the *Cryptosporidium parvum* genome: An expressed sequence tag and genome survey sequence analysis. *Molecular and Biochemical Parasitology*, **107**(1): 1-32.

Subramanian S, Mishra R K, Singh L. 2003. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology*, **4**: R13, http://dx.doi.org/10.1186/gb-2003-4-2-r13.

Tadano R, Nunome M, Mizutani M, Kawahara-Miki R, Fujiwara A, Takahashi S, Kawashima T, Nirasawa K, Ono T, Kono T, Matsuda Y. 2014. Cost-effective development of highly polymorphic microsatellite in Japanese quail

facilitated by next-generation sequencing. *Animal Genetics*, **45**(6): 881-884.

Tanguy A, Bierne N, Saavedra C et al. 2008. Increasing genomic information in bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. *Gene*, **408**(1-2): 27-36.

Tautz D, Renz M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, **12**(10): 4 127-4 138.

Tautz D, Trick M, Dover G A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature*, **322**(6080): 652-656.

Thanh N M, Jung H, Lyons R E, Chand V, Tuan N V, Thu V T M, Mather P. 2014. A transcriptomic analysis of striped catfish (*Pangasianodon hypophthalmus*) in response to salinity adaptation: *De novo* assembly, gene annotation and marker discovery. *Comparative Biochemistry and Physiology Part D*: *Genomics and Proteomics*, **10**: 52-63.

Tóth G, Gáspári Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*, **10**(7): 967-981.

Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats A C, Vagner S. 2003. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the Cell*, **95**(4): 169-178, http://dx.doi.org/10.1016/S0248-4900(03)00033-9.

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, Yoocha T, Sangsrakru D, Chanprasert J, Ngamphiw C, Jomchai N, Therawattanasuk K, Tangphatsornruang S. 2011. Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. *DNA Research*, **18**(6): 471-482.

Turanov A A, Lobanov A V, Fomenko D E, Morrison H G, Sogin M L, Klobutcher L A, Hatfield D L, Gladyshev V N. 2009. Genetic code supports targeted insertion of two amino acids by one codon. *Science*, **323**(5911): 259-261, http://dx.doi.org/10.1126/science.1164748.

Varshney R K, Graner A, Sorrells M E. 2005. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, **23**(1): 48-55.

Vukosavljev M, Di Guardo M, van de Weg W E, Arens P, Smulders M J M. 2012. Quantification of Allele Dosage in tetraploid Roses. *Science MED* (*Bologna*), **3**: 277-282.

Vukosavljev M, Zhang J, Esselink G D, van't Westende W P C, Cox P, Visser R G F, Arens P, Smulders M J M. 2013. Genetic diversity and differentiation in roses: a garden rose perspective. *Science Horticulturae*, **162**: 320-332, http://dx.doi.org/10.1016/j.scienta.2013.08.015.

Wahba A J, Gardner R S, Basilio C, Miller R S, Speyer J F, Lengyel P. 1963. Synthetic polynucleotides and the amino acid code. VIII. *Proceedings of the National Academy of Sciences of the United States of America*, **49**: 116-122.

Wang H X, Huan P, Lu X, Liu B Z. 2011. Mining of EST-SSR markers in clam *Meretrix meretrix* larvae from 454 shotgun transcriptome. *Genes & Genetic System*, **86**(3): 197-205.

Wang J Y, Song X M, Li Y, Hou X L. 2013. *In-silico* detection of EST-SSR markers in three *Brassica species* and transferability in *B*. *rapa*. *The Journal of Horticultural Science & Biotechnology*, **88**(2): 135-140.

Wang W J, Kong J, Dong S R, Luan S, Wang Q Y. 2006. Genetic mapping of the Chinese shrimp *Fenneropenaeus chinensis* using AFLP markers. *Acta Zoologica Sinica*, **52**(3): 575-584. (in Chinese with English abstract)

Weber J L, Wong C. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**(8): 1 123-1 128.

Xu P X, Wu X H, Luo J, Wang B G, Liu Y H, Ehlers J D, Wang S, Lu Z F, Li G J. 2011. Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. *BMC Genomics*, **12**: 467.

Xu P X, Xu S Z, Wu X H, Tao Y, Wang B G, Wang S, Qin D H, Lu Z F, Li G J. 2014. Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *The Plant Journal*, **77**(3): 430-442, http://dx.doi.org/10.1111/tpj.12370.

Yuan S X, Ge L, Liu C, Ming J. 2013. The development of EST-SSR markers in *Lilium regale* and their cross-amplification in related species. *Euphytica*, **189**(3): 393-419.

Zane L, Bargelloni L, Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**(1): 1-16.

Zeng S H, Xiao G, Guo J, Fei Z J, Xu Y Q, Roe B A, Wang Y. 2010. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics*, **11**: 94.

Zhou W, Hu Y Y, Sui Z H, Fu F, Wang J G, Chang L P, Guo W H, Li B B, Sun H. 2013. Genome survey sequencing and genetic background characterization of *Gracilariopsis lemaneiformis* (Rhodophyta) based on next-generation sequencing. *PLoS One*, **8**(7): e69909, http://dx.doi.org/10.1371/journal.pone.0069909.

Zitouna N, Marghali S, Gharbi M, Chennaoui-Kourda H, Haddioui A, Trifi-Farah N. 2013. Mediterranean *Hedysarum* phylogeny by transferable microsatellites from Medicago. *Biochemical Systematics and Ecology*, **50**: 129-135.