Pakistan J. Zool., vol. 55(2), pp 795-802, 2023 DOI: https://dx.doi.org/10.17582/journal.pjz/20190829030819

Genome Survey and Large-scale Isolation of Microsatellites through Illumina Sequencing from the Red Swamp Crayfish *Procambarus clarkii*

Xiao Zhang[1], Guoqing Shen[1], Yang Wang[1], Pengdan Huang[1,2], Yan Shui[2], Zenghong Xu[2] and Huaishun Shen[1, 2]*

[1]*Wuxi Fisheries College, Nanjing Agricultural University, Nanjing, 210095, PR China*
[2]*Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization, Ministry of Agriculture, Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, No. 9 Shanshui East Road, Wuxi 214081, PR China*

ABSTRACT

Microsatellites, also known as simple sequence repeats (SSRs), are a type of DNA marker that are highly polymorphic in the genome. However, the microsatellite markers available for the red swamp crayfish *Procambarus clarkii* are limited. In this study, a sizeable number of microsatellites were identified using the data of the *P. clarkii* genome survey generated from Illumina Hi-Seq PE sequencing. A total length of 1,625 Mb genomic sequences of *P. clarkii* were assembled based on 124 Gb raw sequencing data. Based on a Kmer analysis, the *P. clarkii* genome size was projected as approximately 3.6 Gb. A total of 336,275 SSRs (simple sequence repeats) identified motifs were identified. There were 167,999 the most common type di-nucleotide repeat motifs (49.94%). The second most common repeats were the tri-nucleotide ones (131,405, 39.09%). A total of 312,454 microsatellite loci had flanking sequences. A subset of 56 primer pairs was randomly selected for the PCR experiments to validate the SSRs that had been identified. Forty-six primer pairs (80%) had PCR products that matched the sizes expected, and the products from 40 primer pairs (71%) exhibited polymorphism. Thirty-nine *P. clarkii* individuals were analyzed for their primer sets, and 22 were identified. The average allelic number in the 22 polymorphic loci was 3.6. The 22 microsatellite loci all had good polymorphisms. Among them, the polymorphism information content of the 12 loci is greater than 0.5, which is highly polymorphic.

Article Information
Received 29 August 2019
Revised 30 December 2019
Accepted 30 April 2020
Available online 01 June 2022 (early access)
Published 30 December 2022

Authors' Contribution
XZ performed the experimental work and produced the tables and figures. GS, YW, PH, YS and ZX helped in improvement of the manuscript. HS supervised the work, performed statistical analysis and wrote the paper.

Key words
Microsatellite marker, *Procambarus clarkia*, Genome survey, Solexa sequencing, Polymorphic loci

INTRODUCTION

Microsatellites, also known as simple sequence repeats (SSR), are distributed in the eukaryotic genome and are the primary component of the eukaryotic genome repeats, consisting of 2 to 6 nucleotide tandem repeats (Queller *et al*., 1993; Freimer and Slatkin, 1996; Tamura and Satodate, 1996). SSRs are abundant in the genome and are codominantly expressed and widely distributed. Based on their polymorphism, SSRs are widely used in genetic structure analysis, germplasm identification, paternity testing, genetic linkage map construction, QTL mapping, and marker-assisted selection breeding (MAS) (Weissenbach, 1993; Yue *et al*., 2008, 2010; Liao *et al*., 2009; Gulcher, 2012; Andriantahina *et al*., 2013; Liming *et al*., 2013; Miah *et al*., 2013). To achieve these research purposes, it is necessary to obtain a sufficient number of SSRs. The traditional technique for isolating SSRs is based on using SSR-related probes to screen genomic libraries, which is inefficient, difficult, and time consuming. In recent years, next-generation sequencing (NGS) technologies, which have found wide use in exploring genomic information in both non-model and model organisms, have provided a more efficient method to isolate SSRs in many organisms, especially non-model organisms (Xiong *et al*., 2012; Kang *et al*., 2014; Loughnan *et al*., 2015; Heras *et al*., 2016; Ariede *et al*., 2018).

The classification of *Procambarus clarkii* includes the arthropods, crustaceans, Decapoda, and the family Lobaridae, and they are commonly known as crayfish. The red swamp crayfish (*P. clarkia*) is native to northern Mexico and the southern USA, and it has spread to Europe, Africa, Asia, and other parts of the Americas, and it has

* Corresponding author: shenhuaishun@ffrc.cn
0030-9923/2023/0002-795 $ 9.00/0

become an invasive pest (Gherardi, 2006). It was brought to China in 1929 (Yue *et al*., 2008) For more than half a century, its natural population has developed rapidly and has been widely distributed in most provinces and cities in China, especially in the lower and middle parts of the Yangtze River, where it has become a dominant population. Over the years, it has become an important species in the aquaculture industry due to its high protein, low fat, rich nutrition and delicious flavor (Wang *et al*., 2005; Shen *et al*., 2014).

So far, studies on crayfish using microsatellite markers have primarily focused on the evaluation of germplasm resources and germplasm quality of the original crayfish population in different river basins (i.e., the genetic diversity analysis of the population of the original crayfish in different regions) and the identification of relatives of the original crayfish population. Little is known about the genetic linkage map construction, QTL character localization, and MAS of *P. clarkii*. Traditional methods have developed very few SSRs (<50) in *P. clarkia* (Belfiore and May, 2000; Harris and Crandall, 2000; Zhu and Yue, 2008; Li *et al*., 2012; Shi *et al*., 2018), which limits these studies. This study utilized Solexa sequencing technology to perform a genome survey of *P. clarkii* and isolate large-scale polymorphic microsatellites.

MATERIALS AND METHODS

Ethics statement

This study had the approval of the Animal Care and Use Committee of the Freshwater Fisheries Center at the Chinese Academy of Fishery Sciences.

Animal collection

A crayfish farm in Xuyi, Jiangsu Province, China, served as the source of the *P. clarkia* samples. The muscle tissue of the crayfish served as the source of the genomic DNA for sequencing. The total genomic DNA was extracted using the standard procedure of proteinase K and phenol-chloroform. A Nanodrop spectrophotometer (Madison, WI, USA) was used to quantify the genomic DNA, and the DNA was sized using a 0.8% agarose gel. In the analysis of population genetics, the total DNA was separately extracted from the muscle tissues of 39 *P. clarkii* individuals using the same method.

DNA sequencing

Library construction involved the use of approximately 5 ug of qualified genomic DNA. The libraries were prepared following the manufacturer's instructions (Illumina, San Diego, CA, USA). Briefly, qualified DNA samples were subjected to random interruption using a Covaris ultrasonic crushing instrument to grow 350 bp DNA fragments, and the whole library was prepared using the steps of end repair, adding an A tail and a sequencing joint, purification and PCR amplification. The constructed library was sequenced using Illumina Hi-Seq at Beijing Novogene Ltd, China. The genome data have been deposited in the NCBI Sequence Read Archive (SRA) (accession number: PRJNA554704).

De novo assembly

The raw reads that were generated by the Illumina Hi-Seq sequencer were first trimmed by removing the adapter sequences. The pair end reads were removed when the low quality bases (quality scores less than 5) contained in the single-ended sequencing reads exceeded 20% or the content of the "N" base in the single-end sequencing reads was greater than 10% (Erlich *et al*., 2008; Cock *et al*., 2010). The subsequent assembly was performed using the high-quality reads that resulted from this procedure. Before the genome assembly, we used a k-mer based analysis to estimate the genome size and heterozygosity ratio. The read sequences were aligned using the SOAP de novo software (http://soap.genomics.org.cn/ soapdenovo.html) with default settings, and we adopted the De Bruijin graph data structures to construct the contigs (Pevzner *et al*., 2001; Li *et al*., 2009; Luo *et al*., 2012). Using the connection relation between the reads and inserted fragment size information, the contigs were assembled into the scaffolds.

Genome size estimation

A K-mer-based analytical method was used to estimate the heterozygosity and genome size (Marcais and Kingsford, 2011). In brief, we iteratively selected a sequence of K bases from a continuous sequence. If the length of each sequence is L, then the K-mer length is K, then L-K+1 K-mer can be obtained, and we use K=17 for analysis. The genome size = the K-mer count/the peak of the depth distribution.

Identification of microsatellite loci

Simple repeats in the DNA sequences were detected using the SSR search software. The detection software is composed of three modules. The first module is used for the detection of all the simple repeat sequences of the DNA sequences, and the second module is used to filter the results from the first module to eliminate simple repeat sequences that are too close, while the third module is to use primer 3 to design primers (Rozen and Skaletsky, 2000). The design conditions of the SSR primers are as follows: (1) the optimal length of the primer is 24; (2) the minimum length of the primer is 20; (3) the longest primer length is

28; (4) the optimum annealing temperature of the primer is 63 °C; (5) the minimum annealing temperature of the primer is 60 °C; (6) the maximum annealing temperature of the primer is 65 °C; (7) the maximum temperature difference of annealing between a pair of primers is 1 °C.

SSR marker validation and population genetic analysis

To test the amplified quality of the SSRs identified, a 56-primer pair subset was selected and amplified using genomic DNAs of a panel of 4 individuals. Using the specific amplification products produced from the primers, the SSR polymorphisms were further evaluated and were amplified consistently across the individuals using an additional 39 individual samples.

The PCR reactions were performed in a 25 μL reaction system that contained 3 μL DNA (~30 ng), 1 μL reverse primer and 1 μL forward primer (10 μM each), 12.5 μL 2× Taq PCR MasterMix (Takara, Dalian, China), and supplemented with distilled water. The PCR reaction procedure was as follows: 3 min at 94 °C followed by 30 cycles of 94 °C for 30 s, annealing at 60 °C for 30 s and at 72 °C for 30 s, with a final extension of 72 °C for 10 min. The alleles were separated on 12% denaturing polyacrylamide gels, which were stained using 0.1% silver nitrate (Bassam *et al*., 1991).

The size of the bands in the polyacrylamide gels was calculated using Quantity One software (http://www.bio-rad.com) to determine the genotype of the samples. The number of alleles (Aa), observation heterozygosity (Ho), expectation heterozygosity (He), and the effective allele number (Na) was counted using PopGene32 Software (version 1.32) and used to evaluate the polymorphisms of the microsatellite loci.

RESULTS

Genome survey and assembly

After strict filtering of the sequencing data generated by the Illumina Hi-Seq PE sequencing, high quality clean data was obtained. The sequencing quality met the requirements (Q20= 93.78%, Q30= 86.75%); the sequencing error rate was normal (<0.05%), and the nucleotide database comparison results showed that there was a small amount of manson's schistosome contamination. The collected output data totaled 120.59 Gb (Supplementary Table 1).

The genome features can be estimated from the sequences obtained by sequencing prior to genome assembly. We used a k-mer-based analytical method to estimate the genome size and heterozygosity. The survey analysis indicated that the genome size obtained by (kmer-number/depth) is approximately 3627.30 Mbp, and the modified genome size is 3579.38 Mbp (Fig. 1A, Supplementary Table 2). The heterozygosity rate was 0.88% and the repetition rate was 77.65% (Supplementary Table 2). Estimates indicated that the average GC content of the *P. clarkii* genome was 45.21% (Fig. 1B).

The initial assembly was performed using kmer 41; the total number of contigs was 3,711,514; the length of contig N50 was 879 bp, and the total length of of the contigs was 1,535,295,243 bp (Supplementary Table 3). The total number of scaffolds was 2,940,268; scaffold N50 was 1426 bp, and its total length was 1,625,211,665 bp (Supplementary Table 4).

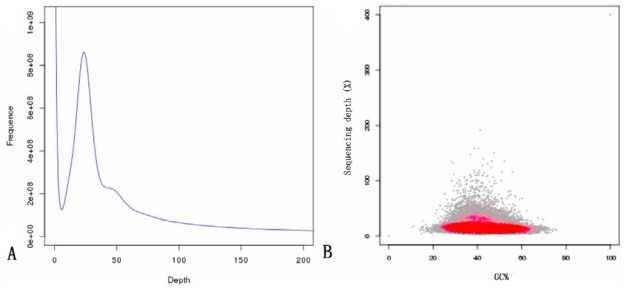

Fig. 1. Results of a genomic survey using Illumina sequencing. (A), A K-mer (K=17) analysis to estimate the genome size of *Procambarus clarkii*. The x-axis is the depth (X); the Y-axis is the proportion that represents the frequency at that depth divided by the total frequency of all depths. (B), GC content and average sequencing depth. The X-axis represents the GC content, and the Y-axis represents the average depth. The red region represents the relatively dense part of the points in the scatter plot.

Microsatellite loci discovery and primer pair design

By filtering out the SSR sequences at both ends of the sequence (less than 100 bp from both ends of the sequence), a total of 336,184 SSR sequences were detected. Among them are 312,455 SSRs designed with primers and a length distribution of 12~136 bp. The statistical results of the SSR sequence length distribution designed are shown in Table I. The length of the SSR units is concentrated in 2~6 bp, and the length distribution of the SSR units is shown in Table I.

The di-nucleotide repeat motifs totaled 167,908, and they are the most common type of repeat motif (49.94%). The second most common was the tri-nucleotide repeats (131,405, 39.09%). This was followed by tetra-nucleotide (29,536, 8.79%), penta-nucleotide (5,870, 1.75%), and hexa-nucleotide (1,465, 0.44%) repeat units (Table I). The relative abundance of the special repeat motifs is highly variable. Figure 2A shows the composition of the primary motifs in the di-nucleotide repeats: AC/GT (48, 420, 28.82%), CA/TG (40, 128, 23.89%), GA/TC (28, 813, 17.15%) and AG/CT (24, 503, 14.59) %), reaching

Table I. Frequency of simple sequence repeats (SSRs) in *Procambrus clarkii.*

Motif length	Repeat number											Total number	%
	4	5	6	7	8	9	10-13	14-17	18-19	22-25	>25		
Di-			54610	29750	18315	11961	20545	10089	3810	3494	15425	167908	49.94
Tri-	79642	28104	11598	5133	2580	1495	2038	389	219	111	96	131405	39.09
Tetra-	18602	5811	2268	1188	638	327	347	182	117	43	13	29536	8.78
Penta-	3814	1133	516	218	70	36	62	15	5	1	0	5870	1.75
Hexa-	1117	244	59	21	11	7	5	1	0	0	0	1465	0.44
Total	103175	35292	69051	36310	21614	13826	22997	10676	4151	3649	15534	336184	100
%	30.68	10.49	20.53	10.8	6.43	4.11	6.84	3.17	1.23	1.08	4.62		

15.57%, 13.27% and 13.81% of the total, respectively. The motifs GC/GC (526, 0.31%) and CG/CG (389, 0.23%) were relatively small. The AAT/ATT motif (14,242) was the most common (10.84%) tri-nucleotide motif, followed by CAC/GTG (10,980, 8.36%) and TAA/TTA (10,749, 8.18%),while CGC/GCG (458, 0.35%), CCG/CGG (327, 0.25%) and CGA/TCG (267, 0.20%) were relatively rare (Fig. 2B). Figure 3 shows the calculations of the frequency distributions from the di-nucleotide to the hexa-nucleotide repeats. Most of the repeat sequences were concentrated in a domain that consisted of low copy numbers, and fewer sequences were observed as the copy number increases. The most abundant di-nucleotide repeat sequence was a repeat with 6 copies (32.52%). The most abundant tri-nucleotide repeat sequence was a repeat with 4 copies (60.61%). Table I provides the number of each repeat sequence under the copy number of its repeat unit.

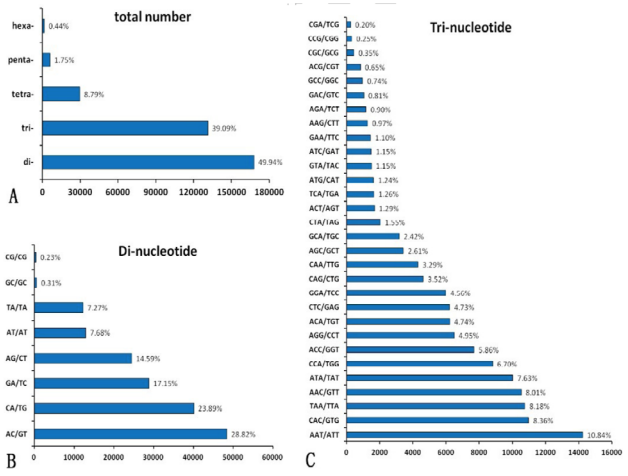

Fig. 2. Distribution of SSR motifs in *Procambarus clarkii*. The Y-axis represents motif types and the X-axis represents the count of motifs in whole genome of *Procambarus clarkii*.

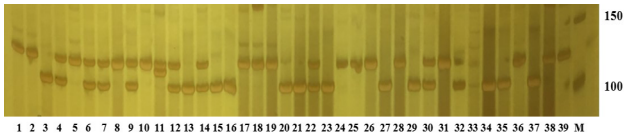

Fig. 3. The amplification band of scaffold1946501 loci for 39 individuals of *Procambrus clarkii*. The M is Marker; 150 is 150 bp, and 100 is 100 bp.

Of all the 336,184 SSR sequences, 312,455 SSRS could be designed for their corresponding primer pairs that reached 92.91% of the total identified microsatellite loci. These microsatellites represent potentially amplifiable microsatellite loci (PAL).

SSR validation and population genetic analysis

To conduct validation, a subset of 56 PAL was randomly chosen. These loci provided the source of primers for tests with the genomic DNA from a panel of 10 individuals. Amplicons were successfully produced from 46 primer sets (80%), and they matched the sizes expected. Nine primer pairs did not produce any amplification products. Six sets of the 46 primer sets generated monomorphic products in all the individuals tested.

In an analysis of 39 *P. clarkii* individuals with 22 primer sets, 78 alleles were detected, with the number produced at each site ranging between 2 and 6. The average number of alleles was 3.6. The mean observed heterozygosity (HO) ranges between 0.2 and 1 and averages 0.559. The average expected heterozygosity (HE) ranges between 0.337 and 0.821 and averages 0.615. The content (PIC) ranges between 0.269 and 0.746 and averages 0.515. The primer pairs for 22 microsatellite loci sequences are available in Supplementary Table 5, and the amplification result using one of primer pairs is shown in Figure 3.

DISCUSSION

Genome size estimation by genome survey sequencing

As a type of non-model species, there is little

knowledge of the *P.clarkii* genomic landscape. Illumina Solexa sequencing technology supplies a quick and easy platform to explore genome sequences. The K-mer method has been used successfully in genome surveys to estimate the genome size of non-model species, which lacked prior knowledge of its size (Yu *et al*., 2015; Shi *et al*., 2018; Zhao *et al*., 2019). It was estimated that the genome scale reached 3.4 Gb based on our data. The proportion of the heterozygote rate was approximately 0.88%, and the *P. clarkii* genomic sequences had a repeat ration of 77.65%. in conclusion, the *P. clarkii* genome is large and complex with a high sequence repeat ratio and a high heterozygote rate. The genome size of the marbled crayfish *Procambarus virginalis* was determined to be approximately 3.5 Gb using genome sequencing and was estimated to be 3.3 Gb based on K-mer frequencies (Zhang *et al*., 2019). A 3.4 Gb estimated genome size of *P. clarkii* is similar to that of the closely related *P. virginalis*, which leads us to believe that our estimate of the size of the *P. clarkii* genome is within reasonable limits. The *P. clarkii* genome size is estimated to be approximately 8.50 Gb by flow cytometry and 1.86 Gb with genome survey sequencing in a former study (Shi *et al*., 2018). There are also some other differences between our data and those of Shi, including the fact that the the repeat ratio of the *P. clarkii* genomic sequences was high (~91.05%), and the proportion of the heterozygote rate was approximately 0.69%, in the study by Shi. The difference between our study and that of Shi may be due to the fact that the genome survey sequencing used to estimate the genome size of large and complex genomes is inefficient. The distribution of the K-mer depth is affected by the presence of genomic heterozygotes and repeats. However, since genome survey sequencing can provide an initial global view of a genome, we believe that the results of both our study and that of Shi could provide a degree of reference for a future *P. clarkii* whole genome project.

Highly efficient method to explore SSRs

The traditional methods that are used to explore the SSR markers have a low efficiency and are time consuming. Therefore, the published number of SSRs in *P. clarkii* is limited (Belfiore and May, 2000; Zhu and Yue, 2008). In this study, we used the Illumina genomic survey to explore an extensive set of the 336,184 microsatellite loci in which 312,455 loci are PAL. The enormous number of SSR loci in *P. clarkii* will provide better selection that is more convenient for the research on genetic linkage mapping, genetic diversity analysis, and the MAS breeding of this organism.

The di-nucleotide repeats contributed to nearly half (49.96%) of the genome SSRs in *P. clarkii*, while the tri-nucleotide repeats comprised the second largest proportion (39.08%). The di-nucleotide repeats was reported to be the most common in many species, including some crustaceans, such as the Chinese mitten crab (*Eriocheir sinensis*) and the Chinese shrimp (Kong and Gao, 2005; Xiong *et al*., 2012). AC/GT (28.82%) and CA/TG (23.89%) was the most abundant motif, while GC/GC and CG/CG was the least frequent in the *P. clarkii* genome, which is typical for many other eukaryotic genomes (Ross *et al*., 2003). This is also the case in the freshwater Chinese mitten crab. Both belong to Crustacea *Decapoda*. However, there is a different scenario for the AT/AT and TA/TA motif, which totals 14.95% in *P. clarkii* and less than 0.01% in *E. sinensis* (Xiong *et al*., 2012). The lowest frequency of the GC or CG repeats are observed in most organisms, including fungi and plants (Toth *et al*., 2000). AAT/ATT (10.84%) was the most frequent tri-nucleotide repeat in our dataset, as it was in humans (Toth *et al*., 2000) and pistachio (Motalebipour *et al*., 2016). However, the other tri-nucleotide repeats, such as AGC, AGG, and AAG, were the most frequent motifs in *E. sinensis*, fruit fly and Chinese shrimp, respectively (Schug *et al*., 1998; Kong and Gao, 2005; Xiong *et al*., 2012). These results indicate that the relative frequency of the di-nucleotide motifs is more conserved than those of the tri-nucleotide motifs.

Dependable SSR markers based on genomic information

In a total of all 336,184 SSR sequences, 312,455 SSRS can be designed by their corresponding primer pairs, which reached a high ratio (92.91%). We randomly selected a subset of 56 PAL from a total of 312,455 to examine using PCR amplification. Forty-six primer pairs, which reached 80%, yielded amplicons that matched the expected sizes. A high validation rate indicates that the SSRs examined in this study are reliable and of high quality. A population with 39 *P. clarkii* individuals was analyzed with 22 primer sets. Adjusting for the null alleles, the mean number of alleles per locus, HE, HO, and PIC are 3.6, 0.615, 0.559, and 0.515, respectively, demonstrating that the *P. clarkii* individuals have a relatively high genetic diversity.

CONCLUSION

In conclusion, the *P. clarkii* genome size was estimated to reach 3.6 Gb based on an analysis of K-mer in this study. A total length of 1,625 Mb genomic sequences of *P. clarkii* were assembled based on 124 Gb raw sequencing data, and a total of 336,275 SSR motifs were identified, in which 312,455 SSRS can be designed for their corresponding primer pairs. We believe that a high number of newly developed SSRs will result in a useful resource for genetic research in *P. clarkii*.

ACKNOWLEDGEMENTS

This study was supported by the Central Public-Interest Scientific Institution Basal Research Fund, Freshwater Fisheries Research Center, CAFS (2019JBFZ09), the Natural Science Foundation of Jiangsu Province (BK20181138), and the Aquatic Three Update Project of Jiangsu Province (Y2016-35).

Supplementary material

There is supplementary material associated with this article. Access the material online at: https://dx.doi.org/10.17582/journal.pjz/20190829030819

Statement of conflict of interest

The authors have declared no conflict of interests.

REFERENCES

Andriantahina, F., Liu, X.L. and Huang, H., 2013. Genetic map construction and quantitative trait locus (qtl) detection of growth-related traits in *litopenaeus vannamei* for selective breeding applications. *PLoS One*, **8**: 11. https://doi.org/10.1371/annotation/80183e4b-b1b4-48cc-b307-865af97df0ce

Ariede, R.B., Freitas, M.V., Hata, M.E., Matrochirico, V.A., Utsunomia, R., Mendonca, F.F., Foresti, F., Porto-Foresti, F. and Hashimoto, D.T., 2018. Development of microsatellite markers using next-generation sequencing for the fish *colossoma macropomum*. *Mol. Biol. Rep*., **45**: 9-18. https://doi.org/10.1007/s11033-017-4134-z

Bassam, B.J., Caetanoanolles, G. and Gresshoff, P.M., 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem*., **196**: 80-83. https://doi.org/10.1016/0003-2697(91)90120-I

Belfiore, N.M. and May, B., 2000. Variable microsatellite loci in red swamp crayfish, *Procambarus clarkii*, and their characterization in other crayfish taxa. *Mol. Ecol*., **9**: 2231-2234. https://doi.org/10.1046/j.1365-294X.2000.105339.x

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L. and Rice, P.M., 2010. The sanger fastq file format for sequences with quality scores and the solexa/illumina fastq variants. *Nucl. Acids Res*., **38**: 1767-1771. https://doi.org/10.1093/nar/gkp1137

Erlich, Y., Mitra, P.P., delaBastide, M., McCombie, W.R. and Hannon, G.J., 2008. Alta-cyclic: A self-optimizing base caller for next-generation sequencing. *Nat. Methods*, **5**: 679. https://doi.org/10.1038/nmeth.1230

Freimer, N.B. and Slatkin, M., 1996. Microsatellites: Evolution and mutational processes. In: *Variation in the human genome* (eds. D. Chadwick and G. Cardew). John Wiley and Sons Ltd., Chichester. pp. 51-67. https://doi.org/10.1002/9780470514887.ch4

Gherardi, F., 2006. Crayfish invading europe: The case study of *procambarus clarkii*. *Mar. Freshw. Behav. Physiol*., **39**: 175-191. https://doi.org/10.1080/10236240600869702

Gulcher, J., 2012. Microsatellite markers for linkage and association studies. *Cold Spring Harbor Protocols*., **2012**: 425-432. https://doi.org/10.1101/pdb.top068510

Harris, D.J. and Crandall, K.A., 2000. Intragenomic variation within its 1 and its 2 of freshwater crayfishes (decapoda : Cambaridae): Implications for phylogenetic and microsatellite studies. *Mol. Biol. Evol*., **17**: 284-291. https://doi.org/10.1093/oxfordjournals.molbev.a026308

Heras, S., Planella, L., Caldarazzo, I., Vera, M., Garcia-Marin, J.L. and Roldan, M.I., 2016. Development and characterization of novel microsatellite markers by next generation sequencing for the blue and red shrimp *aristeus antennatus*. *Peer J*., **4**: 14. https://doi.org/10.7717/peerj.2200

Kang, J.H., Kang, H.S., Noh, E.S., Park, J.Y. and An, C.M., 2014. Isolation and characterization of novel microsatellite markers for the northern mauxia shrimp, *Acetes chinensis*, using pyrosequencing. *Mar. Genom*., **18**: 67-69. https://doi.org/10.1016/j.margen.2014.05.005

Kong, J. and Gao, H.A., 2005. Analysis of tandem repeats in the genome of chinese shrimp *Fenneropenaeus chinensis*. *China Sci. Bull*., **50**: 1462-1469. https://doi.org/10.1360/982005-270

Li, R.Q., Yu, C., Li, Y.R., Lam, T.W., Yiu, S.M., Kristiansen, K. and Wang, J., 2009. Soap2: An improved ultrafast tool for short read alignment. *Bioinformatics*, **25**: 1966-1967. https://doi.org/10.1093/bioinformatics/btp336

Li, Y.H., Guo, X.W., Cao, X.J., Deng, W., Luo, W. and Wang, W.M., 2012. Population genetic structure and post-establishment dispersal patterns of the red swamp crayfish *Procambarus clarkii* in china. *PLoS One*, **7**: 8. https://doi.org/10.1371/journal.pone.0040652

Liao, X.L., Ma, H.Y., Xu, G.B., Shao, C.W., Tian, Y.S., Ji, X.S., Yang, J.F. and Chen, S.L., 2009. Construction of a genetic linkage map and mapping of a female-specific DNA marker in half-smooth tongue sole (*Cynoglossus semilaevis*). *Mar. Biotechnol*., **11**: 699-709. https://doi.org/10.1007/

s10126-009-9184-3

Liming, L.M., Guannan, G.N., Quanqi, Q.Q., Zhigang, Z.G., Xubo, X.B., Jieming, J.M. and Yu, H.Y., 2013. A microsatellite genetic linkage map of half smooth tongue sole (*Cynoglossus semilaevis*). *Mar. Genom.*, **9**: 17-23. https://doi.org/10.1016/j.margen.2012.07.002

Loughnan, S.R., Beheregaray, L.B. and Robinson, N.A., 2015. Microsatellite marker development based on next-generation sequencing for the smooth marron (*Cherax cainii*, austin) and cross-species amplification in other cherax species. *BMC Res. Notes*, **8**: 370. https://doi.org/10.1186/s13104-015-1345-z

Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., Yuan, J.Y., He, G.Z., Chen, Y.X., Pan, Q., Liu, Y.J., Tang, J.B., Wu, G.X., Zhang, H., Shi, Y.J., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C.L., Cheung, D.W., Yiu, S.M., Peng, S.L., Zhu, X.Q., Liu, G.M., Liao, X.K., Li, Y.R., Yang, H.M., Wang, J., Lam, T.W. and Wang, J., 2012. Soapdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *Giga Sci.*, **1**: 6. https://doi.org/10.1186/2047-217X-1-18

Marcais, G. and Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**: 764-770. https://doi.org/10.1093/bioinformatics/btr011

Miah, G., Rafii, M.Y., Ismail, M.R., Puteh, A.B., Rahim, H.A., Islam, K.N. and Latif, M.A., 2013. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *Int. J. mol. Sci.*, **14**: 22499-22528. https://doi.org/10.3390/ijms141122499

Motalebipour, E.Z., Kafkas, S., Khodaeiaminjan, M., Coban, N. and Gozel, H., 2016. Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: Development of novel SSR markers and genetic diversity in *Pistacia* species. *BMC Genomics*, **17**: 14. https://doi.org/10.1186/s12864-016-3359-x

Pevzner, P.A., Tang, H.X. and Waterman, M.S., 2001. An eulerian path approach to DNA fragment assembly. *Proc. natl. Acad. Sci. USA.*, **98**: 9748-9753. https://doi.org/10.1073/pnas.171285098

Queller, D.C., Strassmann, J.E. and Hughes, C.R., 1993. Microsatellites and kinship. *Trends Ecol. Evol.*, **8**: 285. https://doi.org/10.1016/0169-5347(93)90256-O

Ross, C.L., Dyer, K.A., Erez, T., Miller, S.J., Jaenike, J. and Markow, T.A., 2003. Rapid divergence of microsatellite abundance among species of drosophila. *Mol. Biol. Evol.*, **20**: 1143-1157. https://doi.org/10.1093/molbev/msg137

Rozen, S. and Skaletsky, H., 2000. Primer 3 on the www for general users and for biologist programmers. *Methods Mol. Biol.* (*Clifton, N.J.*)., **132**: 365-386. https://doi.org/10.1385/1-59259-192-2:365

Schug, M.D., Wetterstrand, K.A., Gaudette, M.S., Lim, R.H., Hutter, C.M. and Aquadro, C.F., 1998. The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. *Mol. Ecol.*, **7**: 57-70. https://doi.org/10.1046/j.1365-294x.1998.00304.x

Shen, H.S., Hu, Y.C., Ma, Y.C., Zhou, X., Xu, Z.H., Shui, Y., Li, C.Y., Xu, P. and Sun, X.W., 2014. In-depth transcriptome analysis of the red swamp crayfish *Procambarus clarkii*. *PLoS One.*, **9**: 13. https://doi.org/10.1371/journal.pone.0110548

Shi, L.L., Yi, S.K. and Li, Y.H., 2018. Genome survey sequencing of red swamp crayfish *Procambarus clarkii*. *Mol. Biol. Rep.*, **45**: 799-806. https://doi.org/10.1007/s11033-018-4219-3

Tamura, G. and Satodate, R., 1996. Microsatellite analysis: Nihon rinsho. *Jap. J. clin. Med.*, **54**: 928-932.

Toth, G., Gaspari, Z. and Jurka, J., 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.*, **10**: 967-981. https://doi.org/10.1101/gr.10.7.967

Wang, W., Gu, W., Ding, Z.F., Ren, Y.L., Chen, J.X. and Hou, Y.Y., 2005. A novel spiroplasma pathogen causing systemic infection in the crayfish *Procambarus clarkii* (crustacea : Decapod), in china. *FEMS Microbiol. Lett.*, **249**: 131-137. https://doi.org/10.1016/j.femsle.2005.06.005

Weissenbach, J., 1993. Microsatellite polymorphisms and the genetic-linkage map of the human genome. *Curr. Opin. Genet. Dev.*, **3**: 414-417. https://doi.org/10.1016/0959-437X(93)90114-5

Xiong, L.W., Wang, Q. and Qiu, G.F., 2012. Large-scale isolation of microsatellites from chinese mitten crab *Eriocheir sinensis* via a solexa genomic survey. *Int. J. mol. Sci.*, **13**: 16333-16345. https://doi.org/10.3390/ijms131216333

Yu, Y., Zhang, X.J., Yuan, J.B., Li, F.H., Chen, X.H., Zhao, Y.Z., Huang, L., Zheng, H.K. and Xiang, J.H., 2015. Genome survey and high-density genetic map construction provide genomic and genetic resources for the pacific white shrimp *Litopenaeus vannamei*. *Sci. Rep.*, **5**: 14. https://doi.org/10.1038/srep15612

Yue, G.H., Li, J.L., Wang, C.M., Xia, J.H., Wang, G.L. and Feng, J.B., 2010. High prevalence of multiple paternity in the invasive crayfish species,

Procambarus clarkii. *Int. J. biol. Sci*., **6**: 107-115. https://doi.org/10.7150/ijbs.6.107

Yue, G.H., Wang, G.L., Zhu, B.Q., Wang, C.M., Zhu, Z.Y. and Lo, L.C., 2008. Discovery of four natural clones in a crayfish species *Procambarus clarkii*. *Int. J. biol. Sci*., **4**: 279-282. https://doi.org/10.7150/ijbs.4.279

Zhang, X.J., Yuan, J.B., Sun, Y.M., Li, S.H., Gao, Y., Yu, Y., Liu, C.Z., Wang, Q.C., Lv, X.J., Zhang, X.X., Ma, K.Y., Wang, X.B., Lin, W.C., Wang, L., Zhu, X.L., Zhang, C.S., Zhang , J.Q., Jin, S.J., Yu, K.J., Kong, J., Xu, P., Chen, J., Zhang, H.B., Sorgeloos, P., Sagi, A., Alcivar-Warren, A., Liu, Z.J., Wang, L., Ruan, J., Chu, K.H., Liu, B., Li, F.H. and Xiang, J.H., 2019. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat. Commun*., **10**: 14. https://doi.org/10.1038/s41467-018-08197-4

Zhao, M., Wang, W., Chen, W., Ma, C.Y., Zhang, F.Y., Jiang, K.J., Liu, J.G., Diao, L., Qian, H., Zhao, J.X., Wang, T. and Ma, L.B., 2019. Genome survey, high-resolution genetic linkage map construction, growth-related quantitative trait locus (QTL) identification and gene location in *Scylla paramamosain*. *Sci. Rep*., **9**: 10. https://doi.org/10.1038/s41598-019-39070-z

Zhu, Z.Y. and Yue, G.H., 2008. Eleven polymorphic microsatellites isolated from red swamp crayfish, *Procambarus clarkii*. *Mol. Ecol. Resour*., **8**: 796-798. https://doi.org/10.1111/j.1755-0998.2007.02067.x