

Research Article

Genomic characteristics and profile of microsatellite primers for *Acanthogobius ommaturus* by genome survey sequencing

Bingjie Chen¹, Zhicheng Sun¹, Fangrui Lou¹,  Tian-xiang Gao² and  Na Song¹¹Key Laboratory of Mariculture (Ocean University of China), Ministry of Education, Qingdao 266003, China; ²Fishery College, Zhejiang Ocean University, Zhoushan 316022, China**Correspondence:** Na Song (songna624@163.com)

Acanthogobius ommaturus is one of the suitable species to study the genetic mechanism of adaptive evolution, but there are few reports on its genetics. In the present study, the genomic survey sequencing method was used to analyze the genome characters of *A. ommaturus*. A total of 50.50 G high-quality sequence data were obtained in the present study. From the 19-mer distribution frequency, the estimated genome size was 928.01 Mb. The calculated sequence repeat rate was about 38.31%, the heterozygosity was approximately 0.17%, and the GC% content was approximately 40.88%. Moreover, 475,724 simple sequence repeats (SSRs) were identified. Among them, dinucleotide repeats were the most (53.70% of the total SSRs), followed by tri- (35.36%), hexa- (4.59%), tetra- (4.57%) and penta- (1.77%) nucleotide repeats type. This is the first genome-wide feature of this species to be reported.

Introduction

Acanthogobius ommaturus belongs to the order Perciformes, Gobioidae, Gobiidae and *Acanthogobius*. It is a large-body goby fish distributed in the subsea or brackish waters of China, Korea, Japan and Indonesia [1]. The suborder Gobioidae is the largest group of fish species, with more than 2000 species reported [2]. Most goby fish have a wide distribution range, strong adaptability and high diversity, which are suitable for studying the genetic mechanism of adaptive evolution [3]. *Acanthogobius ommaturus* is an economically important species that has been reported to be the dominant species across its distribution area [4–6]. In recent years, many studies focusing on its ecological habits, morphological traits, biochemical determinations and feeding strategy in aquaculture industry have been conducted. For example, Feng et al. studied the morphological character, individual fecundity and ovary histology [1]. Huang et al. explored the effects of cryopreservation on several immune and antioxidative enzyme activities [7]. The characterization of Lactate Dehydrogenase in *A. ommaturus* was also analyzed [8]. Insoo and Hong investigated its feeding ecology in a tidal flat in Korea [9]. However, only few studies have involved its genetic diversity and phylogeographic patterns. Until now NCBI has only published the genome information of seven species of Gobioidae, and it cannot provide us valid information of *A. ommaturus* because of distinct genetic relationship between them [10]. In order to understand the genetic background of *A. ommaturus* and conduct further genetic studies, it is necessary to understand genome characteristics of *A. ommaturus*.

At present, high-throughput next-generation sequencing (NGS) is the main method of genome investigation, which is an important and effective strategy to generate genetic and genomic information [11]. Genome survey sequencing bases on the low-depth sequencing data of the small fragment library [12], and can effectively evaluate the genome size, GC content and heterozygosity through *K*-mer analysis [13,14].

Received: 01 May 2020
Revised: 27 October 2020
Accepted: 27 October 2020Accepted Manuscript online:
28 October 2020
Version of Record published:
13 November 2020

Compared with traditional methods, it is also an effective method to characterize microsatellite markers of target species. As a codominant molecular marker, microsatellite marker has been widely used in studies related to genetic diversity and population structure [15,16]. Although there were some population genetic studies based on different molecular markers, such as mitochondrial DNA and AFLP [17,18], there were no studies on population genetics of *A. ommaturus* by using microsatellite markers. Therefore, in the present study, based on high-throughput sequencing, *K*-mer analysis was used to investigate the genomic characteristics of *A. ommaturus* and also give profile of its microsatellite primers. The results of the present study will provide theoretical basis for the subsequent development of the whole genome *de novo* sequencing assembly strategy [11,19], and can be used to describe the genetic background of many other marine animals [11,20–22]. In addition, the draft genome that is obtained by this sequencing is also conducive to the detecting of SNP loci.

Materials and methods

We have read the policies relating to animal experiments and confirmed the present study complied (ARRIVE guidelines; EU Directive 2010/63/EU for animal experiments). All procedures performed in the present study were approved by the Institutional Animal Care and Use Committee of Ocean University of China.

Sample preparation and DNA extraction

The samples were obtained by trawl net by fishermen from coastal water of Qingdao, China in October 2019, and we bought the lively and undamaged individuals in Qingdao wharf. Then, these fish were quickly transported to our laboratory for recovery. Approximately 7 days later, an active, healthy and undamaged *A. ommaturus* (overall length: 15.80 cm; body length: 12.9 cm; body weight: 30.50 g) was selected as the sample for sequencing. Selected *A. ommaturus* were killed to death with an overdose of anesthetic (tricaine methanesulfonate, MS-222, 100 mg/l). Subsequently, the fish were rapidly dissected, and part of the muscle tissues on the caudal peduncle were collected and preserved in 95% alcohol in ice box. The tissue samples were then quickly transformed to our lab where the samples were stored in -80°C until analysis. The traditional phenol–chloroform method was used to extract genomic DNA. The total DNA was treated with RNase to obtain DNA with high purity and without RNA pollution.

Library construction and sequencing

After DNA extraction, library construction and sequencing were conducted. The qualified genomic DNA was broken to the target fragment (350 base pairs) by physical fragmentation (ultrasonic vibration), and then the small-fragment sequencing library was constructed through the steps of terminal repair, addition of A, addition of connector, selection of target fragment and PCR. Agilent 2100 and Q-PCR were used to detect the size of library fragments and library quantification to determine whether the library meets the sequencing standard. The library was fixed to the sequencing chip by bridge PCR. After library quality inspection and chip fixation, two paired-end DNA libraries were constructed with insert size of 350 base pairs, and then sequenced using the Illumina HiSeq X ten platform following the manufacturer's protocol. The data generated by sequencing was used for the next step of bioinformatics analysis after quality control. Entire read sets were deposited in the short read archive (SRA) databank (<http://www.ncbi.nlm.nih.gov/sra/>) and are available under accession number PRJNA658176.

Analysis method

After GC distribution statistics and evaluation of quality values Q20 (ratio of bases with sequencing quality value above 20) and Q30 (proportion of bases with sequencing quality values above 30), clean reads were obtained after filtering the data of paired-end sequencing, which were used for a *K*-mer analysis. The *K*-mer analysis can be used to obtain the genome length, heterozygosity and repeat content.

The information of peak depth and number of predicted best *K*-mer could be obtained according to the results of the *K*-mer analysis, and subsequently this information was used to estimate the size of the genome. The formula was as follows in eqn (1):

$$\text{Genome size} = \frac{\text{K-mer number}}{\text{K-mer peak depth}} \quad (1)$$

where *K*-mer number was the total number of predicted best *K*-mer, and peak depth was the expected value of the *K*-mer depth. The heterozygosity ratio and repeat sequence ratio were determined according to the methods described in Song et al. [23]. The software Jellyfish (v2.0, <http://www.genome.umd.edu/jellyfish.html>) and GenomeScope (v1.0, <http://qb.cshl.edu/genomescope/analysis.php?code=example6>) were used to conduct *K*-mer

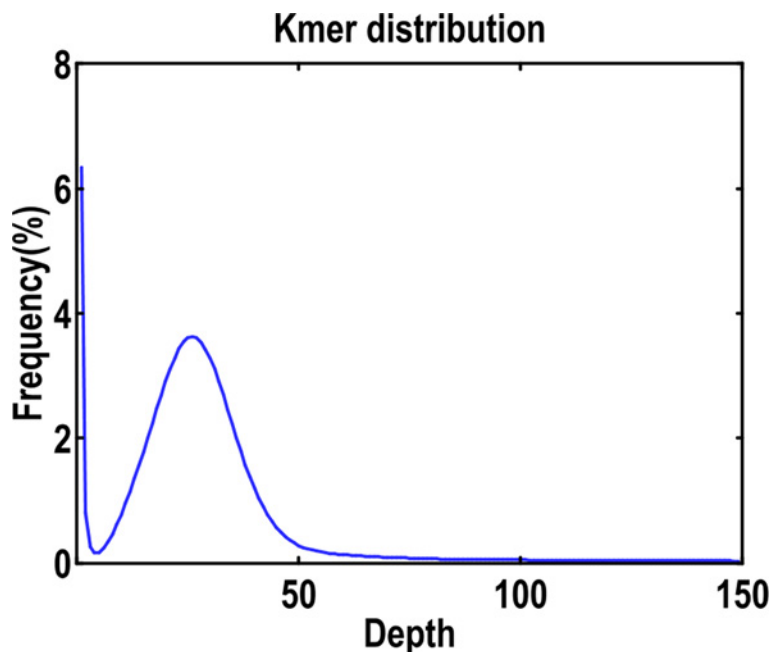


Figure 1. *K*-mer ($K = 19$) analysis for estimation of the genome size of *A. ommaturus*. The X-axis was depth and the Y-axis was the proportion that represented the frequency at that depth.

analyses [24,25]. The clean reads were assembled into contigs in software SOAPdenovo (v2.01, <https://github.com/aquaskyline/SOAPdenovo2/>) by applying the de Bruijn graph structure [26]. The paired-end information was then used to join the unique contigs into scaffolds [11]. To assess the completeness of the assembly, a Benchmarking Universal Single-Copy Orthologs (BUSCO, v.4.0, <https://busco.ezlab.org/>) evaluation was performed using the Eukaryota vertebrata_odb10 database [27]. The Perl script MISA was performed to identify microsatellite motifs in the *de novo* draft genome, and search parameters were set as minimum of 6, 5, 5, 5 and 5 repeats for detecting di-, tri-, tetra-, penta- and hexanucleotide microsatellite motifs, respectively. Primer3 (v2.3.7) software with standard parameters was used to identify the microsatellite loci [28,29].

Results and discussion

Genome sequencing and sequence quality estimation

A 350-bp insert library was constructed using the genomic DNA of the *A. ommaturus* sample, which was sequenced and filtered on the Illumina sequencing platform. A total of high-quality data of 50.50 Gb was obtained, with a total sequencing depth of approximate $54\times$. The Q20 and Q30 values were over 96% and 90%, respectively.

Genome size and repeat ratio, and heterozygosity assessment

Genome survey sequencing yielded a total of 50.50 Gb high-quality data, which were used for subsequent *K*-mer analysis. Using a 350-bp library data to construct a *K*-mer distribution map of $K = 19$ (Figure 1), and the genome size, repeat sequence ratio and heterozygosity were evaluated. The average *K*-mer depth that corresponded to the main peak, was 44, and the total number of *K*-mers obtained from the sequencing data was 44,434,973,013. After removing the anomalies of depth, a total of 41,080,854,391 *K*-mers were used for genome length estimation.

The calculated genome length was about 928.01 Mb. According to the reported genome data, the genome size of most fish is generally small and less than 1 Gb, except for a few fish species such as *Cyprinus carpio*, *Megalobrama amblycephala* and *Sinocyclocheilus graphami* [30]. Moreover, very few fish, like *Salmo salar* [31] and *Oncorhynchus mykiss* [32], are around 1.9-3.0 Gb because of a round of full gene duplication and the expansion of repeat sequence. The genome sizes of four Mudskippers (Gobioidei) were 0.983, 0.806, 0.780 and 0.739 Gb for *Periophthalmodon schlosseri*, *Periophthalmus magnuspinnatus*, *Boleophthalmus pectinirostris* and *Scartelaos histophorus*, respectively [10]. In the present study, genome size of *A. ommaturus* was estimated to be 928.01 Mb (0.928 Gb), which was in accordance with most fish.

Table 1 Statistics of *A. ommaturus* assembled genome sequences

K-mer depth (bp)	Contigs					Scaffolds				
	Total length (bp)	Total number	Max length (bp)	N50 (bp)	N90 (bp)	Total length (bp)	Total number	Max length (bp)	N50 (bp)	N90 (bp)
45	705,413,224	1,522,798	35,678	1120	203	720,996,115	598,674	51,912	2458	436
55	765,231,759	1,421,279	43,980	1237	216	776,582,763	717,328	60,036	2528	343
65	819,722,014	1,395,163	40,568	1267	227	827,753,551	848,081	66,118	2443	301
75	861,431,452	1,387,872	54260	1282	236	867,370,160	944,567	67,355	2343	283

Table 2 A assembly evaluation summary by BUSCO

Term	Results
Complete (C)	1300 (38.76%)
Complete (C) and single-copy (S)	1285 (38.31%)
Complete (C) and duplicated (D)	15 (0.45%)
Fragmented (F)	1150 (34.29%)
Missing (M)	904 (26.95%)
Total	3354

Table 3 Simple sequence repeat (SSR) distribution statistics for *A. ommaturus*

Statistics	Di-	Tri-	Tetra-	Penta-	Hexa-
SSR number	187,501	123,452	15,969	6,182	16,034
Percentage	53.70%	35.36%	4.57%	1.77%	4.59%

According to the *K*-mer distribution, it was estimated that the repeat sequence content was about 38.31%. Most teleost fish exhibit a low repeat content, with an average of less than 30% [33], such as *Gadus morhua* (25.4%) [34], *Gasterosteus aculeatus santaeannae* (25.2%) [35] and *Larimichthys crocea* (16.38%) [36]. A few fish, such as *Danio rerio* and *Salmo salar*, have higher repeat sequences, reaching 52.2% [33] and 60% [31], respectively.

In the present study, there was no obvious heterozygous peak. Overall, the heterozygosity was low, and the estimated heterozygosity was about 0.17%. According to the characteristics of fish genome, draft genome can be obtained by next-generation sequencing technology.

Genome assembly and GC content

With 45, 55, 65 and 75 bp *K*-mers, *de novo* assembly was performed using all of the clean reads (Table 1). When 55 bp *K*-mer value was used, the length of N50 could get the maximum value. The BUSCO evaluation results showed a low score caused by lacking of sufficient genomic data of related species. The intrinsic defects of the second-generation sequencing technology may be also one of the reasons for the poor assembly effect (Table 2). Genomic GC content had a significant effect on the randomness of second-generation genome sequencing. Too high (>65%) or too low (<25%) GC content will lead to sequencing bias and seriously affect the results of genomic analysis. By analyzing the sequencing data of the survey library, the GC content of *A. ommaturus* genome was approximately 40.88%, which was suitable for further analysis and genome assembly.

Identification of SSR

Based on the genome survey sequence, the total number of identified microsatellite or simple sequence repeats (SSRs) was 349,138 (Table 3). Among them, 89,834 sequences contained more than one microsatellite site. The dinucleotide repeats were dominant (53.70%), followed by trinucleotide repeats (35.36%), hexanucleotide repeats (4.59%), tetranucleotide repeats (4.57%) and pentanucleotide repeats (1.77%). According to the distribution and frequency of microsatellite motifs, the dinucleotide repeats had the highest number and type of repeats, which was similar to the previous studies of other fish, such as *Pseudosciaena crocea* [37] and *Megalobrama amblycephala* [38]. The frequency of repetitions decreases exponentially with the length of the repetitions, because long mutations have high mutation rates [39]. This is consistent with the result that the number of repetitions is inversely proportional to the length of repetitions reported by Chen et al. (2010) [40].

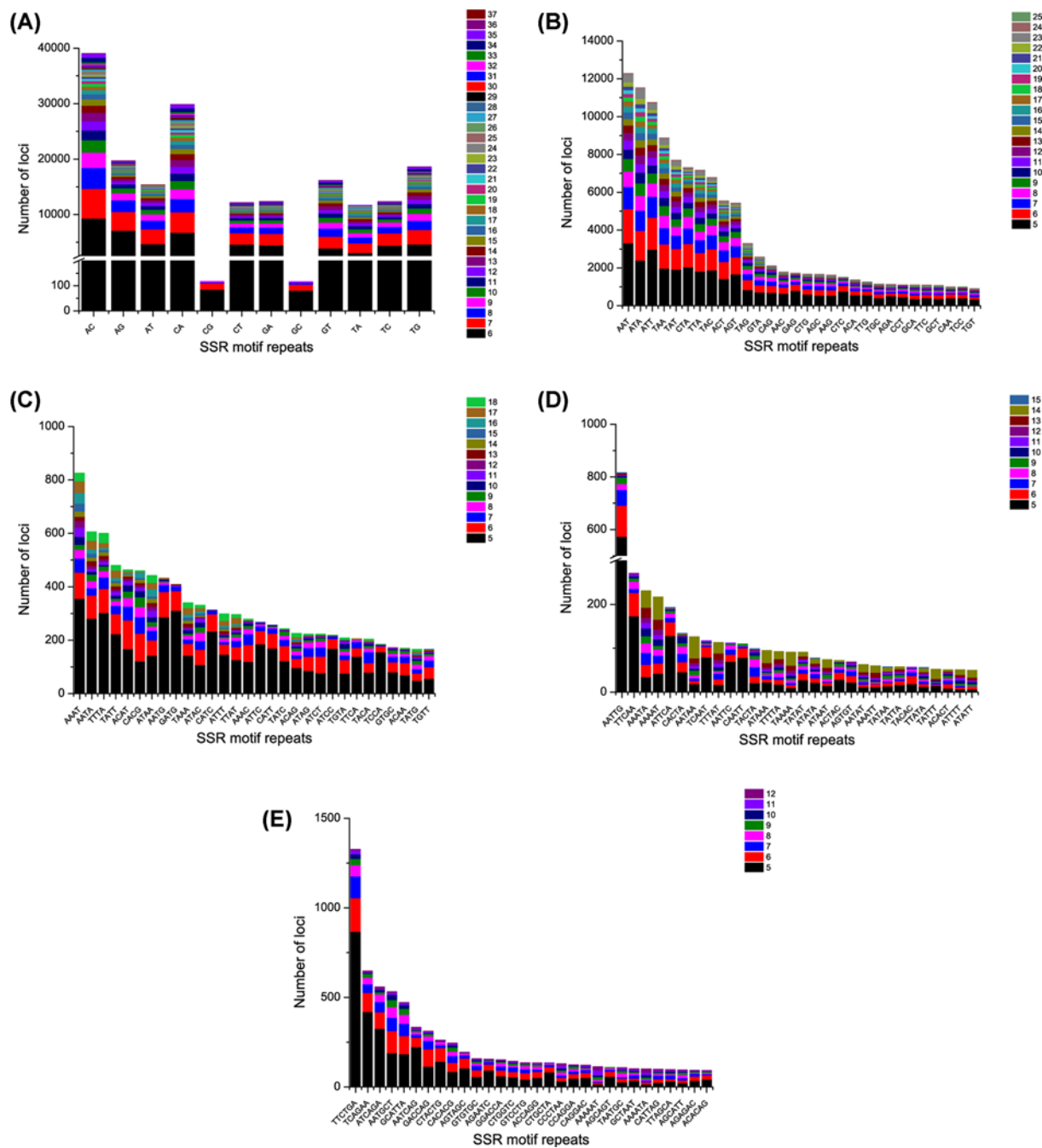


Figure 2. The distribution and frequency of microsatellite motifs

(A) Frequency of different dinucleotide microsatellite motifs. (B) Frequency of different trinucleotide microsatellite motifs. (C) Frequency of different tetranucleotide microsatellite motifs. (D) Frequency of different pentanucleotide microsatellite motifs. (E) Frequency of different hexanucleotide microsatellite motifs.

The AC repeat motif was the most frequent among all twelve types of dinucleotide repeat, whereas GC was the least frequent (Figure 2A). The AAT repeat motif was the most frequent among all sixty types of trinucleotide repeat (Figure 2B). The AAAT repeat motif was the most frequent among all types of tetranucleotide repeat (Figure 2C). The AATTG repeat motif was the most frequent pentanucleotide repeat (Figure 2D), and the TTCTGA was the most frequent hexanucleotide repeat (Figure 2E). There were only eight types of repeats detected in hexanucleotide

repeats, and the number of repeats detected in dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats and pentanucleotide repeats were higher than 10, even up to 35.

To provide more accurate information for SSR primer verification in future research, we selected 4996 pairs of sequences suitable for synthesizing microsatellite primers from the 89,834 sequences, with the following criteria: repetition times of SSR units are more than 6; length of SSR units is 3–10 bp; expected length of PCR products is 130–300 bp; primers that repeat the same four bases continuously are excluded. The selected SSR primers information were provided in the Supplementary Material.

Conclusions

In summary, for *A. ommaturus*, the genome size was approximately 928.01 Mb, the proportion of repeats was approximately 38.31%, the heterozygosity rate was approximately 0.17%, and the GC content of the genome was approximately 40.88%. From the perspective of the basic structural characteristics of the genome, it was a simple genome. This paper could be conducive to the construction of subsequent fine map of the genome.

Data Availability

All research data are included in the main text of the manuscript. Entire read sets were deposited in the short read archive (SRA) databank (<http://www.ncbi.nlm.nih.gov/sra/>) and are available under accession number PRJNA658176.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This study was supported by the National Key R & D Program of China [grant number 2018YFD0900905] and National Natural Science Foundation of China [grant numbers 41506158 and 41776171].

Author Contribution

Conceptualization: B.C. Software: B.C. and F.R. Validation: N.S. and T.G. Investigation: B.C. and Z.S. Resources: B.C. and N.S. Writing—original draft preparation: B.C. Writing—review and editing: B.C. and N.S. Project administration: Z.S. Funding acquisition: N.S.

Acknowledgements

We thank Dr Shengyong Xu and Mr. Jian Zheng for English editing.

Abbreviations

NGS, next-generation sequencing; SRA, short read archive; SSR, simple sequence repeat.

References

- Feng, G., Zhang, L., Zhuang, P., Chen, J., Zhang, T. and Hou, J. (2010) Morphological character, individual fecundity and ovary histology of *Acanthogobius ommaturus* in Yangtze River estuary. *Mar. Fish.* **3**, 257–263
- Wu, H. and Zhong, J. (2008) Osteichthys, Perciformes, Gobioidae. *Fauna Sinica*, 1st ed., Science Press, Beijing, China, Osteichthys, Perciformes 5, 136–718
- Song, C., Feng, Z., Li, C., Sun, Z., Gao, T., Song, N. et al. (2019) Profile and development of microsatellite primers for *Acanthogobius ommaturus* based on high-throughput sequencing technology. *J. Oceanol. Limnol.*, <https://doi.org/10.1007/s00343-019-9154-1>
- Tong, C. (2012) Characteristics of the fish assemblages in the intertidal salt marsh zone and adjacent mudflat in the Yangtze Estuary. *Acta Ecol. Sin.* **32**, 6501–6510, <https://doi.org/10.5846/stxb201203220391>
- Li, F., Zhang, H., Lv, Z., Xu, B. and Zheng, L. (2013) Species composition and community diversity of nekton in Laizhou Bay, China. *Biodivers. Sci.* **21**, 537–546
- Liu, X., Li, Y., Wang, W., Duan, Y., Chai, Y. and Dong, J. (2015) The fish community structure and diversity in the north of Liaodong Bay. *J. Fish. China* **39**, 1155–1165
- Huang, X., Zhang, L., Zhuang, P., Jiang, Q., Yao, Z. and Liu, J. (2010) Effects of cryopreservation on enzyme activity in *Acanthogobius ommaturus* ovum. *Mar. Fish.* **32**, 48–53
- Yum, J. (2008) Characterization of Lactate Dehydrogenase in *Acanthogobius hasta*. *J. Life Sci.* **18**, 264–272, <https://doi.org/10.5352/JLS.2008.18.2.264>

- 9 Insoo, S. and Hong, J. (2006) Feeding Ecology of Gavelin Goby (*Acanthogobius hasta*) and Fine Spot Goby (*Chaeturichthys stigmatias*) in the Jangbong Tidal Flat, Incheon, Korea. *Korean J. Fisheries Aquat. Sci.* **39**, 165–179
- 10 You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J. et al. (2014) Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nat. Commun.* **5**, 1–8, <https://doi.org/10.1038/ncomms6594>
- 11 Xu, S., Song, N., Xiao, S. and Gao, T. (2020) Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus*. *Biosci. Rep.* **5**, 1–8
- 12 Lu, X., Luan, S., Kong, J., Hu, L., Mao, Y. and Zhong, S. (2017) Genome-wide mining, characterization, and development of microsatellite markers in *Marsupenaeus japonicus* by genome survey sequencing. *Chin. J. Oceanol. Limn.* **35**, 203–214, <https://doi.org/10.1007/s00343-016-5250-7>
- 13 Capobianchi, M.R., Giombini, E. and Rozera, G. (2013) Next-generation sequencing technology in clinical virology. *Clin. Microbiol. Infec.* **19**, 15–22, <https://doi.org/10.1111/1469-0691.12056>
- 14 Motalebipour, E.Z., Kafkas, S., Khodaeiaminjan, M., Çoban, N. and Gözel, H. (2016) Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: development of novel SSR markers and genetic diversity in Pistacia species. *Genomics* **17**, 998
- 15 Englbrecht, C.C., Freyhof, J., Nolte, A., Rassmann, K., Schliwien, U. and Tautz, D. (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-Pleistocene origin of the major central European populations. *Mol. Ecol.* **9**, 709–722, <https://doi.org/10.1046/j.1365-294x.2000.00912.x>
- 16 Whitehead, A., Anderson, S.L., Kuivila, K.M., Roach, J.L. and May, B. (2003) Genetic variation among interconnected populations of *Catostomus occidentalis*: implications for distinguishing impacts of contaminants from biogeographical structuring. *Mol. Ecol.* **12**, 2817–2833, <https://doi.org/10.1046/j.1365-294x.2003.01933.x>
- 17 Song, N., Zhang, X.M., Sun, X.F., Yanagimoto, T. and Gao, T.X. (2010) Population genetic structure and larval dispersal potential of spottedtail goby *Synechogobius ommaturus* in the north-west Pacific. *J. Fish Biol.* **77**, 388–402, <https://doi.org/10.1111/j.1095-8649.2010.02694.x>
- 18 Song, N., Zhang, X. and Gao, T. (2010) Genetic diversity and population structure of spottedtail goby (*Synechogobius ommaturus*) based on AFLP analysis. *Biochem. Syst. Ecol.* **38**, 1089–1095, <https://doi.org/10.1016/j.bse.2010.10.007>
- 19 Portis, E., Portis, F., Valente, L., Moglia, A., Barchi, L., Lanteri, S. et al. (2016) A genome-wide survey of the microsatellite content of the globe artichoke genome and the development of a web-based database. *PLoS ONE* **11**, e0162841, <https://doi.org/10.1371/journal.pone.0162841>
- 20 Adelyna, M.N., Jung, H., Chand, V., Mather, P.B. and Azizah, M.S. (2016) A genome survey sequence (GSS) analysis and microsatellite marker development for Indian mackerel, *Rastrelliger kanagurta*, using ion torrent technology. *Meta Gene* **10**, 67–72, <https://doi.org/10.1016/j.mgene.2016.10.005>
- 21 Castoe, T.A., Poole, A.W., de Koning, A.J., Jones, K.L., Tomback, D.F., Oyler-McCance, S.J. et al. (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE* **7**, e30953, <https://doi.org/10.1371/journal.pone.0030953>
- 22 Zhou, W., Hu, Y., Sui, Z., Fu, F., Wang, J., Chang, L. et al. (2013) Genome survey sequencing and genetic background characterization of *Gracilariopsis lemaneiformis* (Rhodophyta) based on next-generation sequencing. *PLoS ONE* **8**, e69909, <https://doi.org/10.1371/journal.pone.0069909>
- 23 Li, G., Song, L., Jin, C., Li, M., Gong, S. and Wang, Y. (2019) Genome survey and SSR analysis of *Apocynum venetum*. *Biosci. Rep.* **39**, BSR20190146, <https://doi.org/10.1042/BSR20190146>
- 24 Guillaume, M. and Carl, K. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770
- 25 Gregory, W.V., Fritz, J.S., Maria, N., Charles, J.U., Han, F., James, G. et al. (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204
- 26 Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J. et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 2047–217X, <https://doi.org/10.1186/2047-217X-1-18>
- 27 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, <https://doi.org/10.1093/bioinformatics/btv351>
- 28 Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. et al. (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115, <https://doi.org/10.1093/nar/gks596>
- 29 Koressaar, T. and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291, <https://doi.org/10.1093/bioinformatics/btm091>
- 30 Chen, S., Xu, W. and Liu, Y. (2019) Fish genomic research: Decade review and prospect. *J. Fish China* **43**, 1–14
- 31 Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T. et al. (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–212, <https://doi.org/10.1038/nature17164>
- 32 Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B. et al. (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat. Commun.* **5**, 3657, <https://doi.org/10.1038/ncomms4657>
- 33 Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M. et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503, <https://doi.org/10.1038/nature12111>
- 34 Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T.F. et al. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207–210, <https://doi.org/10.1038/nature10342>
- 35 Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J. et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61, <https://doi.org/10.1038/nature10944>
- 36 Wu, C., Zhang, D., Kan, M., Lv, Z., Zhu, A., Su, Y. et al. (2014) The draft genome of the large yellow croaker reveals well-developed innate immunity. *Nat. Commun.* **5**, 5227, <https://doi.org/10.1038/ncomms6227>
- 37 Li, Q., Li, Z., Dai, G., Cao, Y., Chen, X., Chen, L. et al. (2014) Isolation and characterization of eleven microsatellite loci in the marbled rockfish, *Sebastiscus marmoratus* (Scorpaenidae). *Conserv. Genet. Resour* **6**, 53–55, <https://doi.org/10.1007/s12686-013-0001-y>

- 38 Zeng, C., Gao, Z., Luo, W., Liu, X., Wang, W., Zhang, X. et al. (2013) Characteristics of microsatellites in blunt snout bream (*Megalobrama amblycephala*) EST sequences using 454 FLX. *Acta Hydrobiol. Sin.* **37**, 982–988
- 39 Katti, M.V., Ranjekar, P.K. and Gupta, V.S. (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* **18**, 1161–1167, <https://doi.org/10.1093/oxfordjournals.molbev.a003903>
- 40 Chen, M., Tan, Z., Zeng, G. and Peng, J. (2010) Comprehensive analysis of simple sequence repeats in pre-miRNAs. *Mol. Biol. Evol.* **27**, 2227–2232, <https://doi.org/10.1093/molbev/msq100>