Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus*

Sheng-yong Xu¹, Na Song², Shi-jun Xiao³, Tian-xiang Gao¹*

¹ Fishery College, Zhejiang Ocean University, Zhoushan 316022, P.R. China, ² Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, P.R. China, ³ School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, P.R. China

*Correspondence: Tian-xiang Gao (gaotianxiang0611@163.com)

Abstract: The marbled rockfish *Sebastiscus marmoratus* is an ecologically and economically important marine fish species distributed along the northwestern Pacific coast from Japan to the Philippines. Here, next-generation sequencing was used to generate a whole genome survey dataset to provide fundamental information of its genome and develop genome-wide microsatellite markers for *S. marmoratus*. The genome size of *S. marmoratus* was estimated as approximate 800 Mb by using K-mer analyses and its heterozygosity ratio and repeat sequence ratio were 0.17% and 39.65%, respectively. The preliminary assembled genome was nearly 609 Mb with GC content of 41.3% and the data was used to develop microsatellite markers. A total of 191,592 microsatellite motifs were identified. The most frequent repeat motif was dinucleotide with a frequency of 76.10%, followed by 19.63% trinucleotide, 3.91% tetranucleotide, and 0.36% pentanucleotide motifs. The AC, GAG, and ATAG repeats were the most abundant motifs of dinucleotide, trinucleotide, and tetranucleotide motifs, respectively. In summary, a wide range of candidate microsatellite markers were identified and characterized in the present study using genome survey analysis. High-quality whole genome sequence based on the “Illumina+PacBio+Hi-C” strategy is warranted for further comparative genomics and evolutionary biology studies in this species.

Keywords: marbled rockfish, microsatellite marker, genome size, genome survey
Introduction

The assessment of genetic diversity and structure is one of the major goals of population management and conservation biology [1]. This assessment should ideally be achieved by utilizing polymorphic and informative markers. Microsatellites or simple sequence repeats (SSRs) are short tandem repeated motif (1~6 bases) that are found in both non-coding and coding regions of the genome and are characterized by a high degree of length polymorphism [2]. Microsatellite markers have become one of the most popular molecular markers and have been widely used in genetic studies due to their ubiquitous occurrence, high reproducibility, multiallelic nature, and codominant mode [2-3]. The advantages of microsatellite markers have made them one of the most useful tools for detecting genetic diversity, genetic linkage mapping, genetic structure, and germplasm and evolution analysis. However, conventional approaches to isolate and develop microsatellite primers were time- and cost-consuming because it is necessary to create enriched microsatellite libraries [2]. Until recently, next generation sequencing (NGS) has provided a new perspective for development studies of microsatellite markers, owing to its high throughput and speed of data generation. So far, NGS has been applied to genomics-based strategies to discover sequences for new microsatellite markers in animals and plants, in a time- and cost-effective manner [4-8]. Genome survey sequencing (GSS) based on the NGS platform, has been proven particularly useful in identifying genome-wide microsatellite markers in non-model species. Microsatellite markers development studies from GSS were performed in numbers of species [9-13]. Genome survey studies also provide information about genome structure of organisms, including estimates of genome size, levels of heterozygosity, and repeat contents.

The marbled rockfish (Sebastiscus marmoratus, Cuvier, 1829) is an ecologically and economically important ovoviviparous marine species inhabiting littoral rocky bottoms along the northwest Pacific coast from Japan to the Philippines [14]. S. marmoratus has strong site fidelity and appears within narrow home ranges [14]. Several studies have been conducted on S. marmoratus germplasm resources due to the decline in wild populations [15-17]. However, inconsistent results were demonstrated given the insufficient resolution of molecular markers. Till now, limited microsatellite marker resources are publically available for S. marmoratus using
The use of microsatellite markers in molecular studies is limited and more microsatellite markers are needed for further studies. In the present study, we aimed to characterize and develop genome-wide microsatellite markers in *S. marmoratus* by genome survey sequencing. The newly identified microsatellites would be useful for extending our current knowledge of *S. marmoratus* genome organization and for genome mapping, marker-aided selection, and population genetics.

**Material and methods**

**Ethics Statement**

Ethical approval was not required for this study because no endangered or alive animals were involved. The specimen used in this study was caught by hook fishing and was dead when collected. All handling of *Sebastiscus marmoratus* specimens was conducted in strict accordance with Animal Care Quality Assurance in China and Zhejiang Ocean University.

**Sample collection and genome survey sequencing**

One male adult *S. marmoratus* was collected from Rushan (36°43′N, 121°39′E), China in October 2015. Muscle tissue was stored in 95% ethanol at -80 °C. Total genomic DNA was extracted using a standard phenol-chloroform method for muscle tissue. DNA was treated with RNase A to produce pure, RNA-free DNA. Two paired-end DNA libraries were constructed with insert size of 350 bp, and then sequenced using the Illumina HiSeq2500 platform following the manufacturer’s protocol. The library construction and sequencing were performed at Novogene in Beijing.

**Data analysis**

After removing low quality reads, all clean data were used to perform K-mer analysis. Based on the results of the K-mer analysis, information on peak depth and the number of predicted best K-mer were obtained and used to estimate the size of the genome. Its relationship was expressed by using the following algorithm: Genome size = K-mer_num/peak_depth, where K-mer_num is the total number of predicted best K-mer, and peak_depth is the expected value of the K-mer depth. Also, the heterozygosity ratio and repeat sequence ratio were estimated following the description in [23], based on the K-mer analysis. K-mer analyses were performed using software GCE v1.0.0 [24] and KmerGenie v1.7039 [25], respectively. The clean reads were assembled into contigs in software SOAPdenovo v2.01 [26] with a K-mer of 21 by applying the de Bruijn graph.
structure. The paired-end information was then used to join the unique contigs into scaffolds.

The Perl script MicroSA
tellite (MISA, http://pgrc.ipk-gatersleben.de/misa/) was used to identify microsatellite motifs in the de novo draft genome. The search parameters were set for the detection of di-, tri-, tetra-, penta-, and hexanucleotide microsatellite motifs with a minimum of 6, 5, 5, 5, and 5 repeats, respectively. The microsatellite loci were subjected to primer design using Primer3 v2.3.7 software [27-28] with the standard parameters.

Results and discussion

Genome size prediction and sequence assembly

The experimental design, sequencing and analysis pipeline was shown in Fig. 1. A total of 35.1 Gb raw data were generated by sequencing genome survey library with 350 bp inserts. The effective rate, error rate, Q20, Q30 and GC content of raw data was shown in Table 1 and Fig. 2. A total of 34.8 Gb clean data were obtained after filtering and used for K-mer analysis. When employing KmerGenie, the predicted best K for K-mer analysis was 107 and the predicted genome size was about 812.86 Mb. Comparatively, when using GCE, the 21-mer frequency distribution derived from the sequencing reads was plotting in Fig. 3; the peak of the 21-mer distribution was 38, and the total K-mer count was 29,998,866,801. As a result, the genome size was estimated as 796.25 Mb and the heterozygosity ratio and repeat sequence ratio were 0.17% and 39.65%, respectively. The development of NGS technology has provided researchers with an affordable way of addressing a wide range of questions, especially in non-model species such as S. marmoratus [11]. In addition, the K-mer method has been successfully applied for the estimation of genome size using NGS reads without prior knowledge of the genome size [29]. Here, for the first time, we reported a genome survey of S. marmoratus using whole genome shotgun sequencing. The K-mer analyses suggested that the genome size is about 800 Mb, which is 87% of the size (920 Mb) previously estimated for S. marmoratus using flow cytometry [30].

Assembly was performed using 34.8-Gb Illumina PE clean reads. The length of contig N50 was 674 bp, and the Scaffold N50 was 4362 bp. The total length of scaffolds was 609.46 Mb. The GC content of scaffolds was 41.3%. The number of scaffolds >100 bp was 412,901 (98.99%) and >1 kb was 188,316 (45.15%) (Table 2). Information about the genome size of S. marmoratus from this study may be useful for further genomic studies in this species.

Identification and characteristic of microsatellite motifs in genome survey
From the 609,456,819 bp genome survey sequence, a total of 191,592 microsatellite motifs were identified, which included 140,801 microsatellite-containing sequences. However, only 67,846 sequences contained more than one microsatellite motifs, and 16,325 microsatellites were present in compound formation. Therefore, the microsatellite distribution frequency in this genome was estimated to be about 314.6 microsatellite per Mb. The motif types of microsatellites included 76.10% dinucleotide, 19.63% trinucleotide, 3.91% tetranucleotide, 0.36% pentanucleotide, and few hexanucleotide repeats (Fig. 4a, supplementary Table S1). The number of dinucleotide repeats was the highest, which was similar to previous studies on the distributions and characteristics of the microsatellites in S. marmoratus [22]. The frequency of repeats in most eukaryotes decreases exponentially with repeat length because mutation rates are higher in longer repeats [31]. Chen et al. [32] also reported that the number of repeats is inversely correlated with repeat length, and our present results confirmed this pattern. The relative abundances of specific repeat motifs were highly variable among the repeats. The frequency distribution range of microsatellite repeats ranged from 6 to 11 repeats for dinucleotide, from 5 to 8 repeats for trinucleotide, from 5 to 6 repeats for tetranucleotide. Of the dinucleotide repeats, the AC, TG, CA, and GT repeats were the first four repeats in abundance, accounting for 19.8% (28,853), 18.4% (26,797), 16.5% (24,093), and 14.3% (20,896), respectively (Fig. 4b). Of the trinucleotide repeats, the GAG repeat was the most abundant, accounting for 5.4% (2030), whereas the ACG repeat was the least, accounting for 0.07% (25). In terms of the frequency of repeats, the five-fold repeat was the most frequent of all trinucleotide repeats (Fig. 4c). Of the tetranucleotide repeats, the ATAG repeat was the most abundant, accounting for 3.2% (239) (Fig. 4d). However, only five-fold and six-fold repeats were identified and the five-fold repeat was predominant of all tetranucleotide repeats. Compared to the results of Song et al. [22], in which the distributions and characteristics of the microsatellites in S. marmoratus were analyzed on the basis of 454 FLX pyrosequencing technique, our results showed high-efficiency in microsatellite loci identification. The number of microsatellites in the present study, as well as the kinds of microsatellite motifs, were much higher than previous study using 454 FLX pyrosequencing technique [22]. This difference might be due to the higher throughput of Illumina sequencing than 454 pyrosequencing.

In the present study, primers were designed for the di- to pentanucleotide repeats to
develop genome-wide microsatellite markers in *S. marmoratus*. With the exceptions of compound repeats, primers were successfully designed for 65.43%, 73.40%, 72.15%, and 63.45% of the di-, tri-, tetra-, and pentanucleotide loci, respectively, proving themselves to be promising candidates for PCR amplification (Fig. 4a).

Genomic microsatellite markers, which are reliable, highly polymorphic, multi-allelic, and easy to amplify, are widely used in population genetics, linkage analysis, evolutionary studies and so on [33]. Queirós et al. [34] suggested that reliable and accurate estimates of genetic diversity can be obtained using random microsatellites distributed throughout the genome because selecting the most polymorphic markers will generally overestimate parameters of genetic diversity, leading to misinterpretations of the actual genetic diversity, which is particularly important for managed and threatened populations. In this study, we provided various candidate genomic microsatellites for *S. marmoratus* that will enhance the range of markers for this species after amplification and testing in various populations. This is the first study to analyze the genome size and the characteristics of *S. marmoratus* microsatellites using genome survey sequencing. The results will be helpful for future population genetics and germplasm resource conservation. In addition, we suggested further studies should generate high-quality whole genome sequence of *S. marmoratus* based on the combination of “Illumina+PacBio+Hi-C” techniques, to provide robust information for genomic and evolutionary biology studies.

**Acknowledgements**

Not applicable

**Author contribution**

T.X.G. conceived the study. S.Y.X., N.S. and S.J.X. collected the samples and extracted the genomic DNA. S.Y.X. performed genome assembly and bioinformatics analyses. S.Y.X. and T.X.G. wrote the original draft manuscript and all authors reviewed the manuscript.

**Competing interests**

The authors declare no competing interests.

**Funding**

This study was supported by National Natural Science Foundation of China (41176117, 41776171) to T.X.G. and Zhoushan Science and Technology Project (2019C21027) to S.Y.X.

**References**

1. Moritz, C. (1994) Defining ‘evolutionarily significant units’ for conservation. Trends in Ecology and Evolution. 9, 373-375. doi: 10.1016/0169-5347(94)90057-4
2. Zane, L., Bargelloni, L., and Patarnello, T. (2002) Strategies for microsatellite isolation: a review. Molecular Ecology. 11, 1-16. doi: 10.1046/j.0962-1083.2001.01418.x
3. Zhang, D.X., and Hewitt, G.M. (2003) Nuclear DNA analyses in genetic studies of populations:
practice, problems and prospects. Molecular Ecology. 12, 563-584. doi: 10.1046/j.1365-294X.2003.01773.x

Cheng, L., Liao, X., Yu, X., and Tong, J. (2007) Development of EST-SSRs by an efficient FIASCO-based strategy: a case study in rare minnow (Gobiocephalus Rarus). Animal Biotechnology. 18, 143-152. doi: 10.1080/10495390601054980

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research. 17, 240-248. doi: 10.1101/gr.5681207

Triwitayakorn, K., Chatkulkawin, P., Kanjanawattanawong, S., Sarphet, S., Yoocha, T., Sangsrukru, D., et al. (2011) Transcriptome sequencing of Hevea brasiliensis for development of microsatellite markers and construction of a genetic linkage map. DNA Research. 18, 471-482. doi: 10.1093/dnares/dsr034

Castoe, T.A., Poole, A.W., de Koning, A.P.J., Jones, K.L., Tombak, D.F., Oyer-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. doi: 10.1371/journal.pone.0030953

Capobianchi, M.R., Giombini, E., and Rozera, G. (2013) Next-generation sequencing technology in clinical virology. Clinical Microbiology and Infection. 19, 15-22. doi: 10.1111/1469-0691.12056

Zhou, W., Hu, Y.Y., Sui, Z.H., Fu, F., Wang, J.G., Chang, L.P., et al. (2013) Genome survey sequencing and genetic background characterization of Gracilaria lemaneiformis (Rhodophyta) based on next-generation sequencing. PLoS ONE. 8, e69909. doi: 10.1371/journal.pone.0069909

Adelynna, M.A.N., Jung, H., Chand, V., Mather, P.B., and Azizah, M.N.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. doi: 10.1016/j.mgene.2016.10.005

Motalebipour, E.Z., Kafkas, S., Khodaei-Minjan, M., Coban, 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. BMC Genomics. 17, 998. doi: 10.1186/s12864-016-3359-x

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. doi: 10.1371/journal.pone.0162841

Lu, X., Luan, S., Kong, J., Hu, L.Y., Mao, Y., and Zhong, S.P. (2017) Genome-wide mining, characterization, and development of microsatellite markers in Marsupenaeus japonicus by genome sequencing survey. Chinese Journal of Oceanology and Limnology. 35, 203-214. doi: 10.1007/s00343-016-5250-7

Fujita, H., and Kohda, M. (1998) Timing and sites of parturition of the viviparous scorpionfish, Sebastiscus marmoratus. Environmental Biology of Fishes. 52, 225-229. doi: 10.1023/A:1007471919373

Sun, D.Q., Shi, G., Liu, X.Z., Wang, R.X., and Xu, T.J. (2011) Genetic diversity and population structure of the marbled rockfish, Sebastiscus marmoratus, revealed by SSR markers. Journal of Genetics. 90, e21-e24. doi: 10.1007/s12041-011-0026-5

Zhang, H. (2013) Molecular phylogeography of two marine ovoviviparous fishes in
Northwestern Pacific. PhD thesis, Qingdao, China: Ocean University of China. (In Chinese)

Xu, S.Y., Sun, D.R., Song, N., Gao, T.X., Han, Z.Q., and Shui, B.N. (2017) Local adaptation shapes pattern of mitochondrial population structure in Sebastiscus marmoratus. Environmental Biology of Fishes. 100, 763-774. doi: 10.1007/s10641-017-0602-5

Xu, T.J., Quan, X.Q., Sun, Y.N., Zhao, K.C., and Wang, R.X. (2010) A first set of polymorphic microsatellite loci from the marbled rockfish, Sebastiscus marmoratus. Biochemical Genetics. 48, 680-683. doi: 10.1007/s10528-010-9349-9

Yin, L.N., Zhang, H., Yanagimoto, T., and Gao, T.X. (2012) Isolation and characterization of nine polymorphic microsatellite markers of the marbled rockfish Sebastiscus marmoratus (Scorpaeniformes, Scorpaenidae). Russian Journal of Genetics. 48, 1264-1266. doi: 10.1134/S1022795412120174

Luo, H.B., Liu, S.F., Ye, J.B., Yuan, Y.J., Ding, S.X., and Zhuang, Z.M. (2014) Polymorphic microsatellite markers in the false kelpfish Sebastiscus marmoratus: isolation, characterization, and cross-species amplification. Genetics and Molecular Research. 13, 134-138. doi: 10.4238/2014.January.10.4

Deng, H.W., Li, Z.B., Dai, G., Yuan, Y., Ning, Y.F., Shangguan, J.B., et al. (2015) Isolation of new polymorphic microsatellite markers from the marbled rockfish Sebastiscus marmoratus. Genetics and Molecular Research. 14, 758-762. doi: 10.4238/2015.January.30.19

Song, N., Chen, M.Y., Gao, T.X., and Yanagimoto, T. (2017) Profile of candidate microsatellite markers in Sebastiscus marmoratus using 454 pyrosequencing. Chinese Journal of Oceanology and Limnology. 35, 198-202. doi: 10.1007/s00343-016-5103-4

Li, G., Song, L., Jin, C., Li, M., Gong, S.P., and Wang, Y.F. (2019) Genome survey and SSR analysis of Apocynum venetum. Bioscience Reports. 39, BSR20190146. doi: 10.1042/BSR20190146

Liu, B.H., Shi, Y.J., Yuan, J.Y., Hu, X.S., Zhang, H., Li, N., et al. (2013) Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv. 1308.2012. https://arxiv.org/abs/1308.2012

Chikhi, R., and Medvedev, P. (2014) Informed and automated k-mer size selection for genome assembly. Bioinformatics. 30, 31-37. doi: 10.1093/bioinformatics/btt310

Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., Yuan, J.Y., et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 1, 18. doi: 10.1186/2047-217X-1-18

Koressaar, T., and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics. 23, 1289-1291. doi: 10.1093/bioinformatics/btm091

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., et al. (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research. 40, e115. doi: 10.1093/nar/gks596

Lu, M., An, H., and Li, L. (2016) Genome survey sequencing for the characterization of the genetic background of Rosa roxburghii Tratt and leaf ascorbate metabolism genes. PLoS ONE. 11, e0147530. doi: 10.1371/journal.pone.0147530

Ojima, Y., and Yamamoto, K. (1990) Cellular DNA contents of fishes determined by flow cytometry. La Kromosomo II. 57, 1871-1888.

Katti, M.V., Ranjekar, P.K., and Gupta, V.S. (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution. 18, 1161-1167.
32. Chen, M., Tan, Z.Y., Zeng, G.M., and Peng, J. (2010) Comprehensive analysis of simple sequence repeats in pre-miRNAs. Molecular Biology and Evolution. 27, 2227-2232. doi: 10.1093/molbev/msq100

33. Varshney, R.K., Graner, A., Sorrells, M.E. (2005) Genic microsatellites markers in plants: features and application. Trends in Biotechnology. 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005

34. Queirós, J., Godinho, R., Lopes, S., Gortazar, C., de la Fuente, J., Alves, P.C. (2015) Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. Molecular Ecology Resources. 15, 747-760. doi: 10.1111/1755-0998.12349
Table 1 Quality control information of Illumina sequencing data

| Lib ID | Raw data (bp) | Clean data (bp) | Effective rate (%) | Error rate (%) | Q20 (%) | Q30 (%) | GC content (%) |
|--------|---------------|-----------------|--------------------|----------------|---------|---------|----------------|
| DES_L5 | 35,057,094,600 | 34,843,246,022  | 99.39              | 0.02; 0.03     | 98.02; 95.13 | 94.91; 89.18 | 42.86; 43.04   |

Note: The two statistics of error rate, Q20, Q30, and GC content were for pair-end read 1 and read 2, respectively.

Table 2 The result of assembly in *S. marmoratus* using 34.8-Gb Illumina clean data

| Contigs | Scaffolds |
|---------|-----------|
| Size (bp) | Number | Size (bp) | Number |
| N90 | 145 | 995,699 | 1117 | 179,431 |
| N80 | 241 | 680,067 | 1877 | 128,640 |
| N70 | 373 | 485,655 | 2589 | 94,551 |
| N60 | 518 | 353,199 | 3413 | 69,208 |
| N50 | 674 | 254,586 | 4362 | 49,699 |
| Total size | 583,830,195 | - | 609,456,819 | - |
| GC content | 41.39% | - | 41.30% | - |
| Total number (> 100 bp) | 1,467,661 | 412,901 |
| Total number (> 1 kb) | 127,823 | 188,316 |

Figure 1. Overview of the experimental design and analysis pipeline.
Figure 2 Distribution figure of error rate, sequencing quality and GC content of raw data.

Figure 3 K-mer (21-mer) analysis for estimating the genome size of *S. marmoratus*. The X-axis is depth and the Y-axis is the proportion that represents the frequency at that depth. Data produced from 350 bp insert library. The peak K-mer frequency was 38.
Figure 4 The distribution and frequency of microsatellite motifs. (a), Frequency of different microsatellite repeat types. ALL, all of the identified microsatellites, PAL, potentially amplifiable loci; (b), Frequency of different dinucleotide microsatellite motifs; (c) Frequency of different trinucleotide microsatellite motifs; (d), Frequency of different tetranucleotide microsatellite motifs.

Supplementary Table S1 Frequency of identified microsatellite motifs
Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus*

Sheng-yong Xu¹, Na Song², Shi-jun Xiao³, Tian-xiang Gao¹*¹

¹ Fishery College, Zhejiang Ocean University, Zhoushan 316022, P.R. China, ² Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao 266003, P.R. China, ³ School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, P.R. China

*Correspondence: Tian-xiang Gao (gaotianxiang0611@163.com)

Abstract: The marbled rockfish *Sebastiscus marmoratus* is an ecologically and economically important marine fish species distributed along the northwestern Pacific coast from Japan to the Philippines. Here, next-generation sequencing was used to generate a whole genome survey dataset to provide fundamental information of its genome and develop genome-wide microsatellite markers for *S. marmoratus*. The genome size of *S. marmoratus* was estimated as approximate 800 Mb by using K-mer analyses and its heterozygosity ratio and repeat sequence ratio were 0.17% and 39.65%, respectively. The preliminary assembled genome was nearly 609 Mb with GC content of 41.3% and the data was used to develop microsatellite markers. A total of 191,592 microsatellite motifs were identified. The most frequent repeat motif was dinucleotide with a frequency of 76.10%, followed by 19.63% trinucleotide, 3.91% tetranucleotide, and 0.36% pentanucleotide motifs. The AC, GAG, and ATAG repeats were the most abundant motifs of dinucleotide, trinucleotide, and tetranucleotide motifs, respectively. In summary, a wide range of candidate microsatellite markers were identified and characterized in the present study using genome survey analysis. High-quality whole genome sequence based on the “Illumina+PacBio+Hi-C” strategy is warranted for further comparative genomics and evolutionary biology studies in this species.

Keywords: marbled rockfish, microsatellite marker, genome size, genome survey
Introduction

The assessment of genetic diversity and structure is one of the major goals of population management and conservation biology [1]. This assessment should ideally be achieved by utilizing polymorphic and informative markers. Microsatellites or simple sequence repeats (SSRs) are short tandem repeated motif (1~6 bases) that are found in both non-coding and coding regions of the genome and are characterized by a high degree of length polymorphism [2]. Microsatellite markers have become one of the most popular molecular markers and have been widely used in genetic studies due to their ubiquitous occurrence, high reproducibility, multiallelic nature, and codominant mode [2-3]. The advantages of microsatellite markers have made them one of the most useful tools for detecting genetic diversity, genetic linkage mapping, genetic structure, and germplasm and evolution analysis. However, conventional approaches to isolate and develop microsatellite primers were time- and cost-consuming because it is necessary to create enriched microsatellite libraries [2]. Until recently, next generation sequencing (NGS) has provided a new perspective for development studies of microsatellite markers, owing to its high throughput and speed of data generation. So far, NGS has been applied to genomics-based strategies to discover sequences for new microsatellite markers in animals and plants, in a time- and cost-effective manner [4-8]. Genome survey sequencing (GSS) based on the NGS platform, has been proven particularly useful in identifying genome-wide microsatellite markers in non-model species. Microsatellite markers development studies from GSS were performed in numbers of species [9-13]. Genome survey studies also provide information about genome structure of organisms, including estimates of genome size, levels of heterozygosity, and repeat contents.

The marbled rockfish (Sebastiscus marmoratus, Cuvier, 1829) is an ecologically and economically important ovoviviparous marine species inhabiting littoral rocky bottoms along the northwest Pacific coast from Japan to the Philippines [14]. S. marmoratus has strong site fidelity and appears within narrow home ranges [14]. Several studies have been conducted on S. marmoratus germplasm resources due to the decline in wild populations [15-17]. However, inconsistent results were demonstrated given the insufficient resolution of molecular markers. Till now, limited microsatellite marker resources are publically available for S. marmoratus using
different methods [18-22]. The use of microsatellite markers in molecular studies is limited and more microsatellite markers are needed for further studies. In the present study, we aimed to characterize and develop genome-wide microsatellite markers in *S. marmoratus* by genome survey sequencing. The newly identified microsatellites would be useful for extending our current knowledge of *S. marmoratus* genome organization and for genome mapping, marker-aided selection, and population genetics.

**Material and methods**

*Ethics Statement*

Ethical approval was not required for this study because no endangered or alive animals were involved. The specimen used in this study was caught by hook fishing and was dead when collected. All handling of *Sebastiscus marmoratus* specimens was conducted in strict accordance with Animal Care Quality Assurance in China and Zhejiang Ocean University.

*Sample collection and genome survey sequencing*

One male adult *S. marmoratus* was collected from Rushan (36°43'N, 121°39'E), China in October 2015. Muscle tissue was stored in 95% ethanol at -80 °C. Total genomic DNA was extracted using a standard phenol-chloroform method for muscle tissue. DNA was treated with RNase A to produce pure, RNA-free DNA. Two paired-end DNA libraries were constructed with insert size of 350 bp, and then sequenced using the Illumina HiSeq2500 platform following the manufacturer’s protocol. The library construction and sequencing were performed at Novogene in Beijing.

*Data analysis*

After removing low quality reads, all clean data were used to perform K-mer analysis. Based on the results of the K-mer analysis, information on peak depth and the number of predicted best K-mer were obtained and used to estimate the size of the genome. Its relationship was expressed by using the following algorithm: Genome size = K-mer_num/peak_depth, where K-mer_num is the total number of predicted best K-mer, and peak_depth is the expected value of the K-mer depth. Also, the heterozygosity ratio and repeat sequence ratio were estimated following the description in [23], based on the K-mer analysis. K-mer analyses were performed using software GCE v1.0.0 [24] and KmerGenie v1.7039 [25], respectively. The clean reads were assembled into contigs in software SOAPdenovo v2.01 [26] with a K-mer of 21 by applying the de Bruijn graph.
structure. The paired-end information was then used to join the unique contigs into scaffolds.

The Perl script MicroSA
tellite (MISA, http://pgrc.ipk-gatersleben.de/misa/) was used to identify microsatellite motifs in the de novo draft genome. The search parameters were set for the detection of di-, tri-, tetra-, penta-, and hexanucleotide microsatellite motifs with a minimum of 6, 5, 5, 5, and 5 repeats, respectively. The microsatellite loci were subjected to primer design using Primer3 v2.3.7 software [27-28] with the standard parameters.

Results and discussion

Genome size prediction and sequence assembly

The experimental design, sequencing and analysis pipeline was shown in Fig. 1. A total of 35.1 Gb raw data were generated by sequencing genome survey library with 350 bp inserts. The effective rate, error rate, Q20, Q30 and GC content of raw data was shown in Table 1 and Fig. 2. A total of 34.8 Gb clean data were obtained after filtering and used for K-mer analysis. When employing KmerGenie, the predicted best K for K-mer analysis was 107 and the predicted genome size was about 812.86 Mb. Comparatively, when using GCE, the 21-mer frequency distribution derived from the sequencing reads was plotting in Fig. 3; the peak of the 21-mer distribution was 38, and the total K-mer count was 29,998,886,801. As a result, the genome size was estimated as 796.25 Mb and the heterozygosity ratio and repeat sequence ratio were 0.17% and 39.65%, respectively. The development of NGS technology has provided researchers with an affordable way of addressing a wide range of questions, especially in non-model species such as S. marmoratus [11]. In addition, the K-mer method has been successfully applied for the estimation of genome size using NGS reads without prior knowledge of the genome size [29]. Here, for the first time, we reported a genome survey of S. marmoratus using whole genome shotgun sequencing. The K-mer analyses suggested that the genome size is about 800 Mb, which is 87% of the size (920 Mb) previously estimated for S. marmoratus using flow cytometry [30].

Assembly was performed using 34.8-Gb Illumina PE clean reads. The length of contig N50 was 674 bp, and the Scaffold N50 was 4362 bp. The total length of scaffolds was 609.46 Mb. The GC content of scaffolds was 41.3%. The number of scaffolds >100 bp was 412,901 (98.99%) and >1 kb was 188,316 (45.15%) (Table 2). Information about the genome size of S. marmoratus from this study may be useful for further genomic studies in this species.

Identification and characteristic of microsatellite motifs in genome survey
From the 609,456,819 bp genome survey sequence, a total of 191,592 microsatellite motifs were identified, which included 140,801 microsatellite-containing sequences. However, only 67,846 sequences contained more than one microsatellite motifs, and 16,325 microsatellites were present in compound formation. Therefore, the microsatellite distribution frequency in this genome was estimated to be about 314.6 microsatellite per Mb. The motif types of microsatellites included 76.10% dinucleotide, 19.63% trinucleotide, 3.91% tetranucleotide, 0.36% pentanucleotide, and few hexanucleotide repeats (Fig. 4a, supplementary Table S1). The number of dinucleotide repeats was the highest, which was similar to previous studies on the distributions and characteristics of the microsatellites in *S. marmoratus* [22]. The frequency of repeats in most eukaryotes decreases exponentially with repeat length because mutation rates are higher in longer repeats [31]. Chen et al. [32] also reported that the number of repeats is inversely correlated with repeat length, and our present results confirmed this pattern. The relative abundances of specific repeat motifs were highly variable among the repeats. The frequency distribution range of microsatellite repeats ranged from 6 to 11 repeats for dinucleotide, from 5 to 8 repeats for trinucleotide, from 5 to 6 repeats for tetranucleotide. Of the dinucleotide repeats, the AC, TG, CA, and GT repeats were the first four repeats in abundance, accounting for 19.8% (28,853), 18.4% (26,797), 16.5% (24,093), and 14.3% (20,896), respectively (Fig. 4b). Of the trinucleotide repeats, the GAG repeat was the most abundant, accounting for 5.4% (2030), whereas the ACG repeat was the least, accounting for 0.07% (25). In terms of the frequency of repeats, the five-fold repeat was the most frequent of all trinucleotide repeats (Fig. 4c). Of the tetranucleotide repeats, the ATAG repeat was the most abundant, accounting for 3.2% (239) (Fig. 4d). However, only five-fold and six-fold repeats were identified and the five-fold repeat was predominant of all tetranucleotide repeats. Compared to the results of Song et al. [22], in which the distributions and characteristics of the microsatellites in *S. marmoratus* were analyzed on the basis of 454 FLX pyrosequencing technique, our results showed high-efficiency in microsatellite loci identification. The number of microsatellites in the present study, as well as the kinds of microsatellite motifs, were much higher than previous study using 454 FLX pyrosequencing technique [22]. This difference might be due to the higher throughput of Illumina sequencing than 454 pyrosequencing.

In the present study, primers were designed for the di- to pentanucleotide repeats to
develop genome-wide microsatellite markers in *S. marmoratus*. With the exceptions of compound repeats, primers were successfully designed for 65.43%, 73.40%, 72.15%, and 63.45% of the di-, tri-, tetra-, and pentanucleotide loci, respectively, proving themselves to be promising candidates for PCR amplification (Fig. 4a).

Genomic microsatellite markers, which are reliable, highly polymorphic, multi-allelic, and easy to amplify, are widely used in population genetics, linkage analysis, evolutionary studies and so on [33]. Queirós et al. [34] suggested that reliable and accurate estimates of genetic diversity can be obtained using random microsatellites distributed throughout the genome because selecting the most polymorphic markers will generally overestimate parameters of genetic diversity, leading to misinterpretations of the actual genetic diversity, which is particularly important for managed and threatened populations. In this study, we provided various candidate genomic microsatellites for *S. marmoratus* that will enhance the range of markers for this species after amplification and testing in various populations. This is the first study to analyze the genome size and the characteristics of *S. marmoratus* microsatellites using genome survey sequencing. The results will be helpful for future population genetics and germplasm resource conservation. In addition, we suggested further studies should generate high-quality whole genome sequence of *S. marmoratus* based on the combination of “Illumina+PacBio+Hi-C” techniques, to provide robust information for genomic and evolutionary biology studies.

**Acknowledgements**

Not applicable

**Author contribution**

T.X.G. conceived the study. S.Y.X., N.S. and S.J.X. collected the samples and extracted the genomic DNA. S.Y.X. performed genome assembly and bioinformatics analyses. S.Y.X. and T.X.G. wrote the original draft manuscript and all authors reviewed the manuscript.

**Competing interests**

The authors declare no competing interests.

**Funding**

This study was supported by National Natural Science Foundation of China (41176117, 41776171) to T.X.G. and Zhoushan Science and Technology Project (2019C21027) to S.Y.X.

**References**

1. Moritz, C. (1994) Defining ‘evolutionarily significant units’ for conservation. Trends in Ecology and Evolution. 9, 373-375. doi: 10.1016/0169-5347(94)90057-4

2. Zane, L., Bargelloni, L., and Patarnello, T. (2002) Strategies for microsatellite isolation: a review. Molecular Ecology. 11, 1-16. doi: 10.1046/j.0962-1083.2001.01418.x

3. Zhang, D.X., and Hewitt, G.M. (2003) Nuclear DNA analyses in genetic studies of populations:
practice, problems and prospects. Molecular Ecology. 12, 563-584. doi:
10.1046/j.1365-294X.2003.01773.x
4. Cheng, L., Liao, X., Yu, X., and Tong, J. (2007) Development of EST-SSRs by an efficient
FIASCO-based strategy: a case study in rare minnow (Gobiocypris Rarus). Animal
Biology. 18, 143-152. doi: 10.1080/10495390601054980
5. Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007) Rapid and
cost-effective polymorphism identification and genotyping using restriction site associated
DNA (RAD) markers. Genome Research. 17, 240-248. doi: 10.1101/gr.5681207
6. Triwitayakorn, K., Chatulkawin, P., Kanjanawattanawong, S., Sraphet, S., Yoocha, T.,
Sangsakru, D., et al. (2011) Transcriptome sequencing of Hevea brasiliensis for development
of microsatellite markers and construction of a genetic linkage map. DNA Research. 18,
471-482. doi: 10.1093/dnares/dsr034
7. Castoe, T.A., Poole, A.W., de Koning, A.P.J., Jones, K.L., Tombak, 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. doi: 10.1371/journal.pone.0030953
8. Caporcihanhi, M.R., Giombini, E., and Rozera, G. (2013) Next-generation sequencing
technology in clinical virology. Clinical Microbiology and Infection. 19, 15-22. doi:
10.1111/1469-0691.12056
9. Zhou, W., Hu, Y.Y., Sui, Z.H., Fu, F., Wang, J.G., Chang, L.P., et al. (2013) Genome survey
sequencing and genetic background characterization of Gracilaria lemaneiformis
(Rhodophyta) based on next-generation sequencing. PLoS ONE. 8, e69909. doi:
10.1371/journal.pone.0069909
10. Adelynna, M.A.N., Jung, H., Chand, V., Mather, P.B., and Azizah, M.N.S. (2016) A genome
survey sequence (SSS) analysis and microsatellite marker development for Indian mackerel,
Rastrelliger kanagurta, using Ion Torrent technology. Meta Gene. 10, 67-72. doi:
10.1016/j.mgen.2016.10.005
11. Motalebipour, E.Z., Kafkas, S., Khodaeaminjan, M., Coban, 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. BMC Genomics. 17, 998. doi:
10.1186/s12864-016-3359-x
12. 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. doi: 10.1371/journal.pone.0162841
13. Lu, X., Luan, S., Kong, J., Hu, L.Y., Mao, Y., and Zhong, S.P. (2017) Genome-wide mining,
characterization, and development of microsatellite markers in Marsupenaeus japonicus by
genome survey sequencing. Chinese Journal of Oceanology and Limnology. 35, 203-214. doi:
10.1007/s00343-016-5250-7
14. Fujita, H., and Kohda, M. (1998) Timing and sites of parturition of the viviparous scorpionfish,
Sebastiscus marmoratus. Environmental Biology of Fishes. 52, 225-229. doi:
10.1023/A:1007471919373
15. Sun, D.Q., Shi, G., Liu, X.Z., Wang, R.X., and Xu, T.J. (2011) Genetic diversity and population
structure of the marbled rockfish, Sebastiscus marmoratus, revealed by SSR markers. Journal
of Genetics. 90, e21-e24. doi: 10.1007/s12041-011-0026-5
16. Zhang, H. (2013) Molecular phylogeography of two marine ovoviviparous fishes in
Northwestern Pacific. PhD thesis, Qingdao, China: Ocean University of China. (In Chinese)

17. Xu, S.Y., Sun, D.R., Song, N., Gao, T.X., Han, Z.Q., and Shui, B.N. (2017) Local adaptation shapes pattern of mitochondrial population structure in Sebastiscus marmoratus. Environmental Biology of Fishes. 100, 763-774. doi: 10.1007/s10641-017-0602-5

18. Xu, T.J., Quan, X.Q., Sun, Y.N., Zhao, K.C., and Wang, R.X. (2010) A first set of polymorphic microsatellite loci from the marbled rockfish, Sebastiscus marmoratus. Biochemical Genetics. 48, 680-683. doi: 10.1007/s10528-010-9349-9

19. Yin, L.N., Zhang, H., Yanagimoto, T., and Gao, T.X. (2012) Isolation and characterization of nine polymorphic microsatellite markers of the marbled rockfish Sebastiscus marmoratus (Scorpaeniformes, Scorpaenidae). Russian Journal of Genetics. 48, 1264-1266. doi: 10.1134/S1022795412120174

20. Liu, H.B., Liu, S.F., Ye, J.B., Yuan, Y.J., Ding, S.X., and Zhuang, Z.M. (2014) Polymorphic microsatellite markers in the false kelpfish Sebastiscus marmoratus: isolation, characterization, and cross-species amplification. Genetics and Molecular Research. 13, 134-138. doi: 10.4238/2014.January.10.4

21. Deng, H.W., Li, Z.B., Dai, G., Yuan, Y., Ning, Y.F., Shangguan, J.B., et al. (2015) Isolation of new polymorphic microsatellite markers from the marbled rockfish Sebastiscus marmoratus. Genetics and Molecular Research. 14, 758-762. doi: 10.4248/2015.January.30.19

22. Song, N., Chen, M.Y., Gao, T.X., and Yanagimoto, T. (2017) Profile of candidate microsatellite markers in Sebastiscus marmoratus using 454 pyrosequencing. Chinese Journal of Oceanology and Limnology. 35, 198-202. doi: 10.1007/s00343-016-5103-4

23. Li, G., Song, L., Jin, C., Li, M., Gong, S.P., and Wang, Y.F. (2019) Genome survey and SSR analysis of Apocynum venetum. Bioscience Reports. 39, BSR20190146. doi: 10.1042/BSR20190146

24. Liu, B.H., Shi, Y.J., Yuan, J.Y., Hu, X.S., Zhang, H., Li, N., et al. (2013) Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv. 1308.2012. https://arxiv.org/abs/1308.2012

25. Chikhi, R., and Medvedev, P. (2014) Informed and automated k-mer size selection for genome assembly. Bioinformatics. 30, 31-37. doi: 10.1093/bioinformatics/btt310

26. Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., Yuan, J.Y., et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 1, 18. doi: 10.1186/2047-217X-1-18

27. Koressaar, T., and Remm, M. (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics. 23, 1289-1291. doi: 10.1093/bioinformatics/btm091

28. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., et al. (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research. 40, e115. doi: 10.1093/nar/gks596

29. Lu, M., An, H., and Li, L. (2016) Genome survey sequencing for the characterization of the genetic background of Rosa roxburghii Tratt and leaf ascorbate metabolism genes. PLoS ONE. 11, e0147530. doi: 10.1371/journal.pone.0147530

30. Ojima, Y., and Yamamoto, K. (1990) Cellular DNA contents of fishes determined by flow cytometry. La Kromosomo Il. 57, 1871-1888.

31. Katti, M.V., Ranjekar, P.K., and Gupta, V.S. (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution. 18, 1161-1167.
doi: 10.1093/molbev.molbev.a003903

32. Chen, M., Tan, Z.Y., Zeng, G.M., and Peng, J. (2010) Comprehensive analysis of simple sequence repeats in pre-miRNAs. Molecular Biology and Evolution. 27, 2227-2232. doi: 10.1093/molbev/msq100

33. Varshney, R.K., Graner, A., Sorrells, M.E. (2005) Genic microsatellites markers in plants: features and application. Trends in Biotechnology. 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005

34. Queirós, J., Godinho, R., Lopes, S., Gortazar, C., de la Fuente, J., Alves, P.C. (2015) Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. Molecular Ecology Resources. 15, 747-760. doi: 10.1111/1755-0998.12349
Tables and Figures

Table 1 Quality control information of Illumina sequencing data

| Lib ID | Raw data (bp) | Clean data (bp) | Effective rate (%) | Error rate (%) | Q20 | Q30 | GC content (%) |
|--------|---------------|-----------------|--------------------|----------------|-----|-----|----------------|
| DES_L5 | 35,057,094,600 | 34,843,246,022 | 99.39 | 0.02; 98.02; 94.91; 42.86; 0.03 | 95.13 | 89.18 | 43.04 |

Note: The two statistics of error rate, Q20, Q30, and GC content were for pair-end read 1 and read 2, respectively.

Table 2 The result of assembly in *S. marmoratus* using 34.8-Gb Illumina clean data

| Contigs | Scaffolds |
|---------|-----------|
| Size (bp) | Number | Size (bp) | Number |
| N90 | 145 | 995,699 | 1117 | 179,431 |
| N80 | 241 | 680,067 | 1877 | 128,640 |
| N70 | 373 | 485,655 | 2589 | 94,551 |
| N60 | 518 | 353,199 | 3413 | 69,208 |
| N50 | 674 | 254,586 | 4362 | 49,699 |
| Total size | 583,830,195 | - | 609,456,819 | - |
| GC content | 41.39% | - | 41.30% | - |
| Total number (> 100 bp) | 1,467,661 | 412,901 |
| Total number (> 1 kb) | 127,823 | 188,316 |

Figure 1. Overview of the experimental design and analysis pipeline.
Figure 2 Distribution figure of error rate, sequencing quality and GC content of raw data.

Figure 3 K-mer (21-mer) analysis for estimating the genome size of *S. marmoratus*. The X-axis is depth and the Y-axis is the proportion that represents the frequency at that depth. Data produced from 350 bp insert library. The peak K-mer frequency was 38.
Figure 4 The distribution and frequency of microsatellite motifs. (a), Frequency of different microsatellite repeat types. ALL, all of the identified microsatellites, PAL, potentially amplifiable loci; (b), Frequency of different dinucleotide microsatellite motifs; (c) Frequency of different trinucleotide microsatellite motifs; (d), Frequency of different tetranucleotide microsatellite motifs.

Supplementary Table S1 Frequency of identified microsatellite motifs
List of Responses

Dear Editors and Reviewers:

Thank you very much for your letter and for the reviewers’ comments concerning our manuscript entitled “Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus*” (ID: BSR-2019-2252). Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet the approval. Revised portion are marked in red in the paper. The main corrections in the paper and the responses to the reviewer’s comments are as following:

Responds to the reviewer’s comments:

Reviewer #1:

1. Response to comment: add the estimation parameters on sequencing quality, such as Q20 and Q30.
   
   **Response:** According to the reviewer’s suggestion, we have revised the manuscript (line 90-94) to add the sequencing quality information.

2. Response to comment: add the figure of GC content (%).
   
   **Response:** According to the reviewer’s suggestion, we have added the quality control information of raw sequencing data in the manuscript as Table 1 and Figure 2.

3. Response to comment: one of highlights of this article is the identification of SSR, I would like to see some candidate SSR loci or primers can be reported in Supplementary table, that would make this article more valuable.
   
   **Response:** We are very thankful for the reviewer’s suggestion. However, only identifying microsatellite motifs could be implemented based on the genome survey sequencing data. The validation of microsatellites would be processed in further studies by using population-based approaches. As a result, we could not show some candidate SSR loci or primers in the manuscript. We are very sorry for that and hope the reviewers and editors understand.

4. Response to comment: Grammar correction, line 10, “northwestern”; line 11, change “employed” to “used”, line 14, change “nearly” to “approximate”.
   
   **Response:** We are very thankful for the reviewer’s suggestion and we have revised these words in the manuscript.

   Special thanks to you for your good comments!

Reviewer #2:

1. Response to comment: In addition to the comments made by reviewer 1, which are necessary to be completed for the manuscript to be considered for publication, I suggest addition of a further figure that explains more clearly the workflow that was carried out to reach the conclusions of the paper.
   
   **Response:** Thank you very much for your suggestion. According the reviewer’s suggestion, we have added the experimental design and analysis pipeline (Figure 1) in our revised version.

   Special thanks to you for your good comments!

We appreciate for editors and reviewers’ warm work earnestly, and hope that the correction will meet with approval.
Once again, thank you very much for your comments and suggestions.
Sample collection → DNA extraction → Illumina library construction → Illumina HiSeq2500 sequencing → Raw data → Data quality control → Data filtering → Clean data → Genome assessment → de novo assembly → microsatellite identification
