Genomic microsatellite characteristics analysis of *Dysomma anguillare* (Anguilliformes, Dysommidae), based on high-throughput sequencing technology

Ziyan Zhu[‡], Yuping Liu[‡], Shufei Zhang[§], Sige Wang[‡], Tianyan Yang[‡]

‡ Zhejiang Ocean University, Zhoushan, China
§ Guangdong Provincial Key Laboratory of Fishery Ecology and Environment, South China Sea Fisheries Research Institute, Guangzhou, China

Corresponding author: Tianyan Yang (hellojelly1130@163.com)

Academic editor: Yahui Zhao

Abstract

Microsatellite loci were screened from the genomic data of *Dysomma anguillare* and their composition and distribution were analysed by bioinformatics for the first time. The results showed that 4,060,742 scaffolds with a total length of 1,562 Mb were obtained by high-throughput sequencing and 1,160,104 microsatellite loci were obtained by MISA screening, which were distributed on 770,294 scaffolds. The occurrence frequency and relative abundance were 28.57% and 743/Mb, respectively. Amongst the six complete microsatellite types, dinucleotide repeats accounted for the largest proportion (592,234, 51.05%), the highest occurrence frequency (14.58%) and the largest relative abundance (379.27/Mb). A total of 1488 microsatellite repeats were detected in the genome of *D. anguillare*, amongst which the hexanucleotide repeat motifs were the most abundant (608), followed by pentanucleotide repeat motifs (574), tetranucleotide repeat motifs (232), trinucleotide repeat motifs (59), dinucleotide repeat motifs (11) and mononucleotide repeat motifs (4). The abundance of microsatellites of the same repeat type decreased with the increase of copy numbers. Amongst the six types of nucleotide repeats, the preponderance of repeated motifs are A (191,390, 43.77%), CA (150,240, 25.37%), AAT (13,168, 14.05%), CACG (2,649, 8.14%), TAATG (119, 19.16%) and CCCTAA (190, 19.16%, 7.65%), respectively. The data of the number, distribution and abundance of different types of microsatellites in the genome of *D. anguillare* were obtained in this study, which would lay a foundation for the development of high-quality microsatellite markers of *D. anguillare* in the future.

Keywords

Introduction

Shortbelly eel (*Dysomma anguillare* Barnard, 1923) is a small-sized warm water eel that is widely distributed in the Indian Ocean and the western Pacific Ocean (Nelson et al. 2016). In China, it is also one of the preponderant bycatch in the offshore waters of the southern East China Sea (Zhao et al. 2016). As an intermediate to high trophic-level species in the coastal food webs, it is of great significance in the offshore marine ecosystem and biodiversity. However, the limited studies of *D. anguillare* were mainly focused on the nutrition and feeding habits (Zhang and Tang 2003), the spatial-temporal pattern of community structure (Liu and Xian 2009) and the effects of lipid removal on the stable isotopes (Yang et al. 2020).

The explicit germplasm genetic characteristics of fishery species are considered to be the indispensable prerequisite for effective fisheries management (Hemmer-Hansen et al. 2018). However, the available genetic data for this species are still scarce and only partial mitochondrial and nuclear gene sequences have hitherto been reported and analysed (Chen et al. 2014, Chang et al. 2016, Wang et al. 2019). Microsatellite DNA, also named simple sequence repeats (SSRs) are short tandem duplications (typically 1-6 nucleotide repeats and mostly less than 100 bp in length), ubiquitous occurring in eukaryotic organisms. Besides, the repetitions vary drastically amongst different genotype of the same species (Tautz and Renz 1984). The co-dominant microsatellite molecular markers, based on polymerase chain reaction (PCR) techniques, have overriding advantages in high polymorphism, good repeatability, simple operation and low experimental cost. Therefore, it has possessed important applied worth in gene mapping and QTL analysis, population genetics and evolutionary research, as well as molecular marker-assisted breeding (Messier et al. 1996, Schlötterer 2000). At present, the conventional development strategies of representative microsatellite loci mainly include anchored-PCR-based method, selective hybridisation enrichment method, database search and relative species selection method (Sun et al. 2009). Nevertheless, these above-mentioned technical means not only are time-consuming and expensive, but also reflect incomplete distribution of microsatellites and develop limited molecular markers.

In recent years, along with the rapid progress of high-throughput sequencing (HTS) technology and the reduction of sequencing cost, developing numerous high-polymorphism SSR markers from multi-omics data has become more and more convenient. In this study, the genome-wide sequences of *Dysomma anguillare* were obtained, based on HiSeq[TM] 4500 platform for the first time; meanwhile, the SSR loci distribution and characteristics were also analysed by bioinformatics tools. The findings

will help to provide useful references and basic information for germplasm resources conservation, population genetic evaluation and phylogenetic relationships analysis amongst related species of Anguilliformes.

Material and Methods

Sample collection and genomic DNA extraction

Fifty-three samples of *Dysomma anguillare* were collected by trawling in the coastal waters of Zhoushan, Zhejiang Province in September 2022. After preliminary morphological identification, muscle tissues from five male and five female individuals were randomly selected for the genomic DNA extraction by the traditional Tris-saturated phenol method (Maniatis et al. 1982). Subsequently, the DNA barcode method, based on the mitochondrial COI sequence, was further conducted to ensure the species accuracy. The 1% agarose-gel electrophoresis and NanoDrop 2000 ultraviolet spectrophotometer (USA, Thermo Fisher Scientific) were performed to detect the integrity and purity of the genomic DNA, respectively. The obtained DNA samples were stored at -20°C for further analysis.

Library construction and high-throughput sequencing

Equal amounts of DNA (2 μg each) were mixed for library construction and next-generation sequencing by Onemore Technology (Wuhan) Co., Ltd. The genomic DNA was randomly fragmented using Covaris Ultrasonic Processor into small 200 to 350 bp fragments. Two pair-end DNA libraries were constructed through terminal repair, adding Poly-A tails and sequencing adapters, purification and PCR amplification and then sequenced using the Illumina HiSeq^{TM} 4500 sequencing technology.

Sequence cleaning and genome assembly

Raw data output from Illumina platform were firstly transformed into sequence reads by base calling and recorded in a FASTQ format. Subsequently, clean reads were obtained after filtering adaptor sequences and low quality read by Cutadapt v.1.16 (Martin 2011). SOAPdenovo v.2.04 was used to assemble the clean data with the setting parameters "-K 53 -R -M 3 -d 1", which employed the *de Bruijn* graph-based assembly strategy (Kajitani et al. 2014). First, reads sequenced from the small-fragment library were divided into smaller substrings (*K-mers*) to construct a preliminary *de Bruijn* diagram. Then, the simplified *de Bruijn* graph was obtained after removing the low-coverage branches and branches that cannot be connected further due to sequencing errors and the sequences at every bifurcation locus were truncated to obtain the initial contigs. By mapping the paired-end reads back to the contigs, the connectivity relationships between the reads and the information of the inserted fragment size were used to further assemble the contigs into scaffolds and obtain the primary genomic sequence.

Screening and identification of SSRs

MicroSatellite identification tool (MISA) software (http://pgrc.ipk-gatersleben.de/misa/) written by Perl script was implemented to scan the assembled scaffolds to identify the genome-wide microsatellite repeat units and to analyse the length, location and quantity of the SSRs (Thiel et al. 2003). The occurrence frequency of SSR loci, average distribution distance and density of microsatellites, type and length of repeat motifs were calculated using Microsoft Excel 2019. The default parameters of MISA were set as follows: the repeat motif length was from 1 to 6 nucleotides and the minimum thresholds of repeat counts were 1-10, 2-6, 3-5, 4-5, 5-5 and 6-5, which meant the number of mononucleotide repeats was less than 10, number of dinucleotide repeats was less than 6 and numbers of remaining repeats were all less than 5, respectively. Besides, the number of bases interrupting two SSRs in a compound microsatellite should be less than 100. Considering the Watson-Crick complementary condition and the difference in the base arrangement, the repeat sequences and their complementary sequences were grouped together. For example, the $(AC)_n$, $(CA)_n$, $(TG)_n$ and $(GT)_n$ were treated as the same SSR repeat types.

Results

Genome sequencing and assembly

The information of contigs and scaffolds of the *Dysomma anguillare* genome was listed in the Table 1. About 11,805,379 contigs with the total length 1,960 Mb were obtained after splicing and the average GC content was about 42.2%. The number of scaffolds produced by the SOAPdenovo v.2.0 assembly was 4,060,742 and the full length was 1,561 Mb, with the average GC content 39.6%.

N50 value is a widely used metric for measuring the quality of sequences by the assembly algorithms' output. It refers to the contig or scaffold length value when the accumulated fragment length (from long to short) exceeds 50% of the total length of all contigs or scaffolds for the first time. The greater the N50 value, the smaller the quantity and the better the assembly quality. In this study, the N50 values of contig and scaffold assembly were 272 bp and 709 bp, respectively. Compared with the assembled genomes of related species *Anguilla japonica* (Henkel et al. 2012), *A. anguilla* (Jansen et al. 2017) and *A. rostrate* (Pavey et al. 2017), the assembly effect of *Dysomma anguillare* was relatively good and developing microsatellite markers could reflect the genome-wide characteristics of SSRs.

SSR repeat types and distribution

A total of 1,160,104 microsatellites with 1-6 bp nucleotide motifs were detected in 770,294 unigenes and 234,959 of them contained more than one SSR locus, with the

occurrence frequency (total number of SSRs detected/total number of unigenes) of 28.57%. The density of distribution (total length of unigenes/total number of SSRs screened) was on average 1/1.35 kb and the relative abundance (total number of SSRs screened/total length of unigenes) was 743/Mb.

These SSR loci can be classified into six repeat types: mononucleotide, dinucleotide, trinucleotide, tertranucleotide, pentanucleotide and hexanucleotide. The most abundant type of repeat motif was dinucleotide, accounting for 51.05% in the all SSR loci and then followed by mononucleotide (37.69%), trinucleotide (8.08%), tertranucleotide (2.71%) and pentanucleotide (0.25%), while hexanucleotide was the minimum (0.21%) of all (Fig. 1). The occurrence frequency of dinucleotide repeats was highest, while hexanucleotide was observed the lowest, representing 14.58% and 0.06% of the total genome, respectively. The relative abundance of dinucleotide reached 379.27/Mb, with an average of one SSR locus per 2.64 kb and the next was mononcleotide (280.00/Mb). By comparison, the relative abundance of hexanucleotide was the lowest (1.59/Mb) (Table 2).

Repeat numbers of different SSRs

The number of repeats of SSR loci mainly ranged from 5 to 24. The predominant repeat number of the SSR loci was 10 times, comprising 17.52% of the total number of SSR loci. In general, the number of repeat types decreased with the increase in repeat numbers (Fig. 2). The repeats of mononucleotide, dinucleotide and trinucleotide were mainly distributed in 10-19 times (96.83%), 6-15 times (95.15%) and 5-9 times (85.34%), respectively. However, the repeat times of the rest of the repeat types were all within 13 times, which were mainly in the range of 5-8 times and separately accounted for 92.40%, 96.70% and 99.56% (Table 3).

In summary, the repeat numbers of SSR loci were mainly concentrated in 10-15 times and 5-8 times, with a total number of 1,016,359 (87.61%). Few SSR loci with more than 25 repeats were identified and the type of base repeats was monotonous, only composing of mononucleotide repeat.

Copy numbers of repeat units

Amongst the detected 1,488 repeat units, hexanucleotide repeats possessed the most types and pentanucleotide repeats took second place. Nevertheless, the type of mononucleotide repeats was the least limited to the base number (Table 4). Amongst all these repeats, the dominant repeat motifs in mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide were A (191,390, 43.77%), CA (150,240, 25.37%), AAT (13,168, 14.05%), CACG (2,649, 8.14%), TAATG (119, 19.16%) and CCCTAA (190, 7.65%), respectively (Fig. 3, Table 4).

SSR length distribution and polymorphism evaluation

The sequence length amongst different types of SSRs varied a lot, from 10 to 54 bp (Fig. 4). The minimum and maximum variations in length were detected in hexanucleotide and mononucleotide repeats, respectively. The former was in the range of 30-54 bp with the total length of 1,774 bp, while the latter was in the range of 10-51 bp with total length of 379,455 bp, which constituted approximately 49.14% of the total length of SSRs. Amongst the six types of nucleotide repeat, dinucleotide and trinucleotide were dominant in the distribution of microsatellites from the perspective of sequence length, which were 677,805 bp in total and accounting for 87.78% in all SSRs.

The length of the microsatellite was one of the main factors affecting its polymorphism. Temnykh et al. (2001) divided SSR sequences into two categories: the high-polymorphic type I (length ≥ 20 bp) and the moderate-polymorphic type II (12 bp ≤ length < 20 bp). The microsatellites with length less than 12 bp owned lower polymorphism, but higher mutation potential. In the present study, there were 21,347 type I SSRs (19%) and 294,373 type II SSRs (54%), respectively. SSR loci with low mutation potential accounted for 27%.

Discussion

Number and relative abundance of microsatellites in the genome of *Dysomma anguillare*

The bioinformatics software was used to search and analyse the various types and numbers of six perfect microsatellites in the genome of *Dysomma anguillare*. Approximately 1,160,104 microsatellite loci were revealed across the 1.56 Gb genome sequence, with a total length of 24,707,980 bp (occupying 58% of the full genome length). In contrast to other published genomes of bony fishes, it was higher than *Takifugu rubripes* (0.77%) (Cui et al. 2006), *Scleropages formosus* (0.78%) (Duan et al. 2019) and *Bagarius yarrelli* (1.23%) (Yang et al. 2021), but lower than *Pelteobagrus fulvidraco* (1.80%) (Xu et al. 2020) and *Harpadon nehereus* (2.01%) (Yang et al. 2021), indicating that genome-wide microsatellites content was not directly related to the genetic relationship and the reasons might involve different retrieval tools, parameter settings and databases (He et al. 2015). Hancock (1996) speculated that the numbers of microsatellites increased with the chromosome length and the disproportional relationship between the genome size and microsatellite numbers was also confirmed in our study.

Relative abundance was an important feature to measure microsatellite richness. It was calculated to be 743/Mb of *Dysomma anguillare*, which was much higher than that of other marine fishes, such as *Scatophagus argus* (653/Mb) (Wang et al. 2020), *Cociella crocodilus* (428/Mb) (Zhao et al. 2021), *Tridentiger bifasciatus* (347/Mb) (Zhao et al. 2022

) and four species of pufferfishes (365/Mb in *Takifugu rubripes*, 369/Mb in *Takifugu flavidus*, 397/Mb in *Takifugu bimaculatus* and 525/Mb in *Tetraodon nigroviridis*) (Xu et al. 2021). The above result showed that abundant microsatellites existed in the genome of *D. anguillare*, which would provide sufficient molecular markers for the further germplasm identification and genetic diversity studies.

Distribution characteristics of microsatellites in the genome of *Dysomma anguillare*

Varied microsatellite types composing of 1-6 nucleotide repeats were discovered in the genome of *Dysomma anguillare* and dinucleotide repeats were the most frequent, followed by mononucleotide repeats, while the percentages of SSRs containing 3-6 nucleotide repeats were no more than 10%. Therefore, priority should be given to dinucleotide repeats when designing SSR primers of *D. anguillare*. Mononucleotide and dinucleotide repeats were regarded as the most abundant types of SSRs in most species. It was reported that mononucleotide repeats tended to dominate in the genomes of higher grade organisms (Gao and Kong 2005). However, dinucleotide repeats contained higher proportions in fish genomes, which probably related to the differences in gene expression and regulation.

The CA repeat motif was the most abundant amongst dinucleotide repeats and occupied 25.37% of them, which was consistent with *Scophthalmus maximus* (Ruan 2009) and pufferfishes (Cui et al. 2006, Xu et al. 2021), but different from *Ictalurus punctatus* (Tang et al. 2022), while the number of GC repeat motifs was the least. The base sliding might generate microsatellites more easily at the low melting temperature (T_m). Two hydrogen bonds between A-T base pairs were more likely to be broken than three hydrogen bonds between G-C base pairs, resulting in reduction of the GC repeats (Huang et al. 2020). Some other scholars pointed out that the methylation of CpG might cause the spontaneous deamination of cytosine to thymine in order to maintain the thermodynamic stability of the DNA molecule. In this study, the proportion of GC repeats motif was only 0.1% and from this aspect, the lower GC content in the whole genome also reflected the small amount of GC repeats (Schorderet and Gartler 1992).

The structural instability and composition of trinucleotide repeats were closely related to some genetic diseases in humans (Sinden et al. 2002). It was found that AAT repeat motif was the most numerous of the trinucleotide repeats in the *Dysomma anguillare* genome, the same as for humans and primates (Kelkar et al. 2008). Therefore, in-depth analysis of trinucleotide repeats would contribute to predict some gene loci associated with human diseases and thereby reduces the occurrence of certain illness by changing gene expression.

Copy numbers and length variations in the genome of *Dysomma anguillare*

The repeat unit length was in inverse proportion to the copy number of microsatellite DNA (Harr and Schlötterer 2000). Commonly, the higher the copy number of SSRs meant the

more alleles and the richer polymorphism. The number of microsatellite repeats in the *Dysomma anguillare* genome was mainly in the range of 5 to 25. Motifs that showed more than 25 reiterations were very rare (only 2,712 SSRs) and all of them were composed of mononucleotide repeats. Previous studies proved that the mutation rate of microsatellites was positively correlated to the copy number of the repeat motif (Wierdl et al. 1997) and longer microsatellites were expected to have higher mutation rate owing to more chances of replication slippage (Calabrese and Sainudiin 2005). The results demonstrated that the number of SSRs decreased as the repeat number increased. In addition, tetranucleotide, pentanucleotide and hexanucleotide microsatellites might have higher mutation rates than those of the mononucleotide, dinucleotide and trinucleotide microsatellites.

The length of microsatellites in the *Dysomma anguillare* genome was generally 10-18 bp and the number of microsatellites was inversely proportional to the repeat motif length. The structure and its characteristics analysis of a parthenogenic gastropod *Melanoides tuberculate* concluded that the longer the repeat sequence length was, the greater the selection pressure undergoing and the lower numbers of repeats was (Samadi et al. 1998). This phenomenon had been verified by various kinds of plants and animals, for instance, *Juglans regia* (Liao et al. 2014), *Patinopecten yessoensis* (Ni et al. 2018) and *Phrynocephalus axillaris* (Song et al. 2019). According to Temnykh et al. (2001), SSR polymorphism could be considered low, medium and high and the SSRs with lengths longer than 12 bp were potential molecular markers with high polymorphism. In the study, the type I and type II SSRs in the *D. anguillare* genome occupied about 73% of the total, showing great potential for polymorphism microsatellite development.

Conclusions

In conclusion, MISA software was used for the first time to search and analyse six types of perfect microsatellite loci from the whole genome survey data of *Dysomma anguillare*. The results showed that both the relative abundance and density of various microsatellite types were very high. Amongst the 1,160,104 SSR loci, the number of different repeat types presented a trend as: dinucleotide > mononucleotide > trinucleotide > tetranucleotide > pentanucleotide > hexanucleotide. The dominant repeat motifs of them were A, CA, AAT, CACG, TAATG and CCCTAA, respectively. The results supplemented the genetic marker database of marine fishes and provided valuable information resources for further genetic analysis of *D. anguillare*.

Acknowledgements

Author contributions

Tianyan Yang conceived and designed the study. Shufei Zhang collected and provided samples. Yuping Liu preformed the DNA extraction and bioinformatics analysis. Ziyan Zhu wrote and edited the manuscript. All authors contributed to the preparation of the manuscript.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Calabrese P, Sainudiin R (2005) Models of microsatellite evolution. In: Loughin TM (Ed.) Statistical Methods in Molecular Evolution. Springer, New York, 289-305 pp. [ISBN 978-0-38722-333-9].
- Chang CH, Shao KT, Lin HY, Chiu YC, Lee MY, Liu SH, Lin PL (2016) DNA barcodes of the native ray-finned fishes in Taiwan. Molecular Ecology Resources 17 (4): 796-805. https://doi.org/10.1111/1755-0998.12601
- Chen JN, López JA, Lavoué S, Miya M, Chen WJ (2014) Phylogeny of the Elopomorpha (Teleostei): Evidence from six nuclear and mitochondrial markers. Molecular Phylogenetics and Evolution 70: 152-161. https://doi.org/10.1016/j.ympev.2013.09.002
- Cui JZ, Shen XY, Yang GP, Gong QL, Gu QQ (2006) The analysis of simple sequence repeats in *Takifugu rubripes* genome. Periodical of Ocean University of China 36 (2): 249-254. https://doi.org/10.3969/j.issn.1672-5174.2006.02.015
- Duan YN, Liu Y, Hu YC, Liu C, Song HM, Wang XJ, Sun JH, Mu XD (2019) Distribution regularity of microsatellites in *Scleropages formosus* genome. Chinese Agricultural Science Bulletin 35 (23): 152-158. https://doi.org/10.11924/j.issn. 1000-6850.casb18030101
- Gao H, Kong J (2005) Distribution characteristics and biological function of tandem repeat sequences in the genomes of different organisms. Zoological Research 26 (5): 555-564. https://doi.org/10.3321/j.issn:0254-5853.2005.05.017
- Hancock JM (1996) Simple sequences and the expanding genome. BioEssays 18 (5): 421-425. https://doi.org/10.1002/bies.950180512
- Harr B, Schlötterer C (2000) Long microsatellite alleles in drosophila melanogaster have a downward mutation bias and short persistence times, which cause their genome-wide under representation. Genetics 155 (3): 1213-1220. https://doi.org/ 10.1016/1350-4533(94)90010-8
- Hemmer-Hansen J, H&uuml s, Baktoft H, Huwer B, Eero M (2018) Genetic analyses reveal complex dynamics within a marine fish management area. Evolutionary Applications 12 (4): 830-844. https://doi.org/10.1111/eva.12760
- Henkel CV, Dirks RP, de Wijze DL, Minegishi Y, Aoyama J, Jansen HJ, Turner B, Knudsen H, Bundgaard M, Hvam KL, Boetzer M, Pirovano W, Weltzien FA, Dufour S,

Tsukamoto K, Spaink HP, van den Thillart GE (2012) First draft genome sequence of the Japanese eel, *Anguilla japonica*. Gene 511: 195-201. https://doi.org/10.1016/j.gene.2012.09.064

- He XD, Zheng JW, He KY, Wang BS (2015) Comparison of different search tools to find microsatellites sites in unigene sequences of *Salix babylonica*. Molecular Plant Breeding 13 (1): 197-204. https://doi.org/10.13271/j.mpb.013.000197
- Huang J, Liu L, Yang B, Yang CZ (2020) Distribution regularities of microsatellites in the genome of great cormorant (*Phalacrocorax carbo*). Chinese Journal of Wildlife 41 (1): 108-114. https://doi.org/10.3969/j.issn.1000-0127.2020.01.015
- Jansen HJ, Jansen HJ, Jong-Raadsen SA, Dufour S, Weltzien FA, Swinkels W, Koelewijn A, Palstra AP, Palstra B, Spaink HP, van den Thillart GE, Dirks RP, Henkel CV (2017) Rapid de novo assembly of the European eel genome from nanopore sequencing reads. Scientific Reports 7: 1-13. https://doi.org/10.1038/s41598-017-07650-6
- Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Research 24 (8): 1384-1395. https://doi.org/10.1101/gr.170720.113
- Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Research 18 (1): 30-38. https://doi.org/10.1101/gr.7113408
- Liao ZY, Ma QY, Dai XG, Zhang DF, Li SX (2014) Microsatellite characters in *Juglans regia* L. genome by high throughput sequencing technology. Journal of Northeast Forestry University 42 (2): 65-68. https://doi.org/10.13759/j.cnki.dlxb.2014.02.016
- Liu SD, Xian WW (2009) Temporal and spatial patterns of the ichthyoplankton community in the Yangtze Estuary and its adjacent waters. Biodiversity Science 17 (2): 151-159. https://doi.org/10.3724/SP.J.1003.2009.08194
- Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York, 157-163 pp. [ISBN 0-87969-136-0]
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17 (1): 10-12. https://doi.org/10.14806/ej.17.1.200
- Messier W, Li SH, Stewart CB (1996) The birth of microsatellites. Nature 381 (6582): 483. https://doi.org/10.1038/381483a0
- Nelson JS, Grande TC, Wilson MVH (2016) Fishes of the World. 5th Edition. John Wiley & Sons, Inc, Hoboken. [ISBN 978-1-11834-233-6] https://doi.org/10.1002/9781119174844
- Ni SS, Yang Y, Liu SF, Zhuang ZM (2018) Microsatellite analysis of *Patinopecten yessoensis* using next-generation sequencing method. Progress in Fishery Sciences 39 (1): 107-113. https://doi.org/10.11758/yykxjz.20161209001
- Pavey SA, Laporte M, Normandeau E, Gaudin J, Letourneau L, Boisvert S, Corbeil J, Audet C, Bernatchez L (2017) Draft genome of the American eel (*Anguilla rostrata*). Molecular Ecology Resources 17: 806-811. https://doi.org/10.1111/1755-0998.12608
- Ruan XH (2009) Development, characterization and application of microsatellite markers in Turbot. The dissertation of a doctor of pharmaceutical chemistry. Medical College, Ocean University of China, Qingdao.
- Samadi S, Artiguebielle E, Estoup A, Pointier JP, Silvain JF, Heller J, Cariou ML, Jarne P (1998) Density and variability of dinucleotide microsatellites in the *parthenogenetic polyploid* snail *Melanoides tuberculata*. Molecular Ecology 7 (9): 1233-1236. https://doi.org/10.1046/j.1365-294x.1998.00405.x

- Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109 (6): 365-371. https://doi.org/10.1007/s004120000089
- Schorderet DF, Gartler SM (1992) Analysis of CpG suppression in methylated and nonmethylated species. Proceedings of the National Academy of Sciences 89 (3): 957-961. https://doi.org/10.1073/pnas.89.3.957
- Sinden RR, Potaman VN, Oussatcheva EA, Pearson CE, Lyubchenko YL, Shlyakhtenko LS (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. Journal of Bioscience 27 (Suppl 1): 53-65. https://doi.org/10.1007/BF02703683
- Song Q, Guo XG, Chen DL (2019) Characterization of microsatellite DNA loci and design of candidate primers to amplify these regions for *Phrynocephalus forsythii* by using 454 GS FLX. Sichuan Journal of Zoology 38 (5): 512-520. https://doi.org/10.11984/j.issn.1000-7083.20190010
- Sun B, Bao YX, Zhao QY, Zhang LL, Hu ZY (2009) Methods for obtaining microsatellite loci: A review. Chinese Journal of Ecology 28 (10): 2130-2137. URL: http://www.cjae.net/EN/Y2004/V15/I09/1580
- Tang RY, Su MY, Yang WS, Xu JJ, Wang T, Yin SW (2022) Analysis of microsatellite distribution characteristics in the channel catfish (*Ictalurus punctatus*) genome. Progress in Fishery Sciences 43 (2): 89-97. https://doi.org/10.19663/j.issn2095-9869.20210126002
- Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research 12 (10): 4127-4138. https://doi.org/10.1093/nar/12.10.4127
- Temnykh S, DeClerck G, Lukashova A, McCouch S, Lipovich L, Cartinhour S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11 (8): 1441-1452. https://doi.org/10.1101/gr.184001
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 106: 411-422. https://doi.org/10.1007/s00122-002-1031-0
- Wang YM, Zhang FY, Zhao M, Ma CY, Zhang LZ, Ma LB (2019) The complete mitochondrial genome of *Dysomma anguillare* (Anguilliformes, Synaphobranchidae) with phylogenetic consideration. Mitochondrial DNA Part B 4 (1): 1688-1689. https://doi.org/10.1080/23802359.2019.1604103
- Wang YR, Yang W, Ren XL, Jiang DN, Deng SP, Chen HP, Zhu CH, Li GL (2020) Distribution patterns of microsatellites and development of polymorphic markers from *Scatophagus argus* genome. Journal of Guangdong Ocean University 40 (4): 7-14. https://doi.org/10.3969/j.issn.1673-9159.2020.04.002
- Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146 (3): 769-779. https://doi.org/10.1093/genetics/146.3.769
- Xu JJ, Zheng X, Li J, Yin SW, Wang T (2020) Distribution characteristics of whole genome microsatellite of *Pelteobagrus fulvidraco*. Genomics and Applied Biology 39 (12): 5488-5498. https://doi.org/10.13417/j.gab.039.005488
- Xu JJ, Zheng X, Zhang XY, Wang T, Yin SW (2021) Analysis of distribution characteristics of microsatellites in four genomes of puffer fish. Genomics and Applied Biology 40 (4): 1441-1451. https://doi.org/10.13417/j.gab.040.001441

- Yang R, Tian SQ, Gao CX, Dai LB, Wang SC (2020) Effects of lipid removal on the stable isotopes of *Dysomma anguillaris* in the offshore waters of southern Zhejiang. Journal of Fishery Sciences of China 27 (9): 1085-1094. https://doi.org/10.3724/SP.J.1118.2020.20027
- Yang TY, Huang XX, Ning ZJ, Gao TX (2021) Genome-wide survey reveals the microsatellite characteristics and phylogenetic relationships of *Harpadon nehereus*. Current Issues in Molecular Biology 43 (3): 1282-1292. https://doi.org/10.3390/cimb43030091
- Yang WS, Tang RY, Su MY, Xu JJ, Wang T, Yin SW (2021) Analysis of microsatellite distribution characteristics in the whole genome of *Bagarius yarrelli*. Journal of Nanjing Normal University, Engineering and Technology Edition 21 (3): 62-68. https://doi.org/10.3969/j.issn.1672-1292.2021.03.009
- Zhang B, Tang QS (2003) Feeding habits of six species of eels in East China Sea and Yellow Sea. Journal of Fisheries of China 27 (4): 307-314. URL: https://www.china-fishery.com/scxuebao/article/abstract/20030404?st=article_issue
- Zhao RR, Lu ZC, Cai SS, Gao TX, Xu SY (2021) Whole genome survey and genetic markers development of crocodile flathead *Cociella crocodilus*. Animal Genetics 52 (6): 891-895. https://doi.org/10.1111/age.13136
- Zhao SL, Xu HX, Zhong JS (2016) Zhejiang marine ichthyology. Zhejiang Science and Technology Press, Hangzhou. [ISBN ISBN 978-7-53417-152-9]
- Zhao X, Liu YX, Du XQ, Ma SY, Song N, Zhao LL (2022) Whole-genome survey analyses provide a new perspective for the evolutionary biology of shimofuri goby, *Tridentiger bifasciatus*. Animals 12 (15): 1914. https://doi.org/10.3390/ani12151914
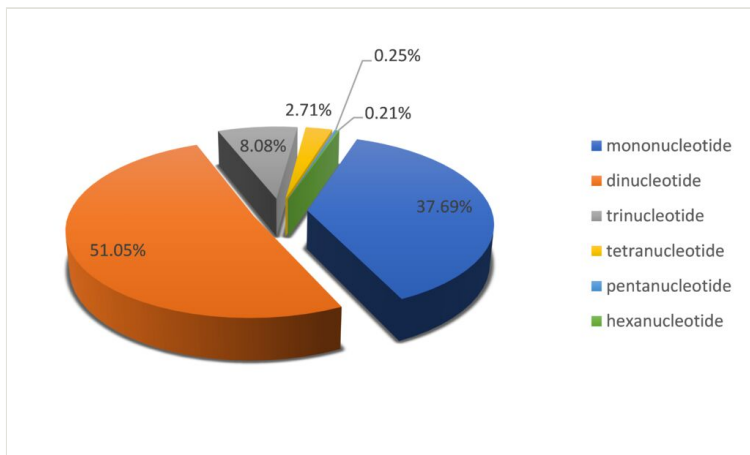
Figure 1.

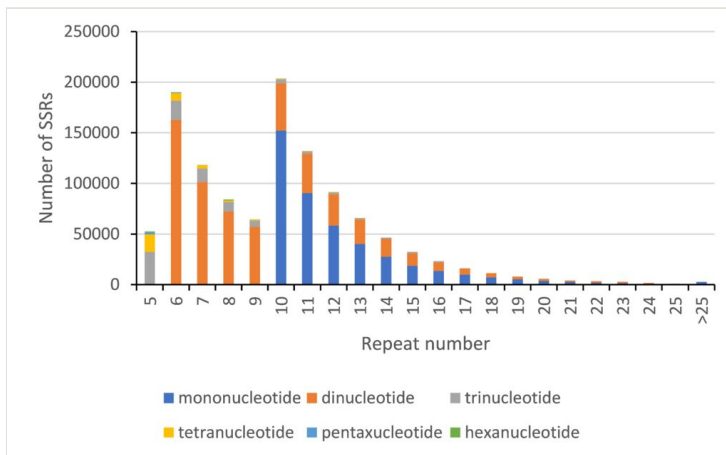Distribution of SSRs repeat types in genomes of *D. anguillare*.

Figure 2.

SSR repeats distribution of *D. anguillare*.

Figure 3.
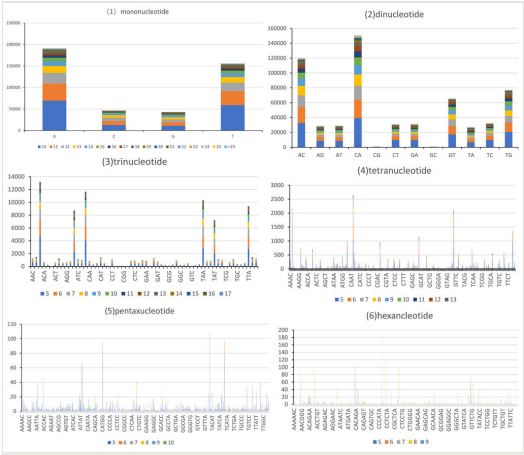
The distribution of microsatellite repeats in genome of *D. anguillare.*
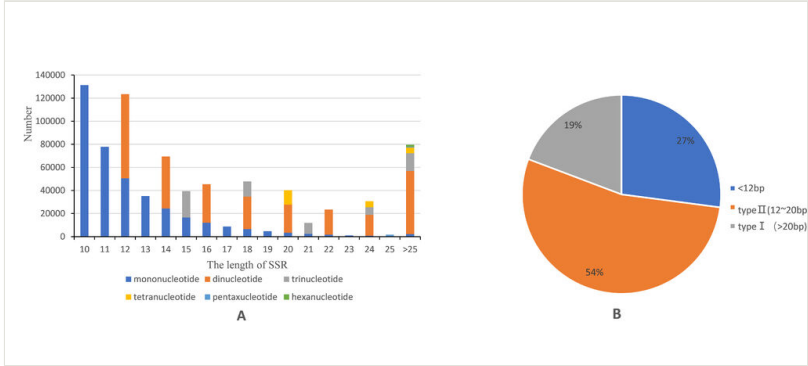
Figure 4.

Length distribution of genes in *D. anguillare.*

A SSR length distribution; **B** Distribution types of SSR (type I and type II).

Table 1.

The contig and scaffold assembly results statistics.

Assembly level	The total length (bp)	The sequence number	Length number of sequences ≥ 2Kb	The maximum length (bp)	N50 (bp)	N90 (bp)	GC content (%)
Contig	1,960,673,378	11,805,379	30,667	9,646	272	60	42.2
Scaffold	1,561,530,495	4,060,742	95,727	23,878	709	134	39.6

Table 2.

Proportions of each SSR repeat types in the genome of *D. anguillare*.

Repeat type	Number	Occurrence frequency (%)	Relative abundance (per · Mb $^{-1}$)	Average length (bp)	Total length (bp)
Mononucleotide	437,234	10.77%	280.00	0.87	379,455
Dinucleotide	592,234	14.58%	379.27	0.50	298,350
Trinucleotide	93,734	2.31%	60.03	0.72	67,533
Tetranucleotide	31,481	0.78%	20.16	0.72	22,680
Pentanucleotide	2,936	0.07%	1.88	0.82	2,409
Hexanucleotide	2,485	0.06%	1.59	0.71	1,774
Total	1,160,104	28.57%	742.93	4.35	772,201

Table 3.

Distribution interval of the copy number in different microsatellite motif for *D. anguillare.*

Repeat number	Mononu cleotide	Dinu cleotide	Trinu cleotide	Tetranu cleotide	Pentanu cleotide	Hexanu cleotide	Total	Proportion (%)
5	0	0	32,413	17,143	2,071	972	52,599	4.53%
6	0	162,916	18,834	7,300	498	394	189,942	16.37%
7	0	101,359	13,353	3,184	176	270	118,342	10.20%
8	0	72,287	9,325	1,460	94	838	84,004	7.24%
9	0	57,111	6,070	895	46	11	64,133	5.53%
10	152,127	46,524	3,962	594	51	0	203,258	17.52%
11	90,414	38,631	2,619	422	0	0	132,086	11.39%
12	58,458	30,798	1,896	430	0	0	91,582	7.89%
13	40,161	23,987	1,488	53	0	0	65,689	5.66%
14	27,543	17,686	1,203	0	0	0	46,432	4.00%
15	18,717	12,237	1,451	0	0	0	32,405	2.79%
16	13,469	8,578	1,069	0	0	0	23,116	1.99%
17	9,925	5,919	51	0	0	0	15,895	1.37%
18	7,271	3,962	0	0	0	0	11,233	0.97%
19	5,276	2,745	0	0	0	0	8,021	0.69%
20	3,800	2,077	0	0	0	0	5,877	0.51%
21	2,605	1,480	0	0	0	0	4,085	0.35%
22	1,889	1,344	0	0	0	0	3,233	0.28%
23	1,297	1,670	0	0	0	0	2,967	0.26%
24	853	878	0	0	0	0	1,731	0.15%
25	697	45	0	0	0	0	742	0.06%
>25	2,712	0	0	0	0	0	2,712	0.23%

Table 4.

Dominant base types and the proportion in genome of *D. anguillare*.

Repeat type	Number of types	Maximum			Minimum		
		Repeat motif	Number	Proportion (%)	Repeat motif	Number	Proportion (%)
Mononucleotide	4	A	191,390	43.77	G	43,065	9.85
Dinucleotide	12	CA	150,240	25.37	GC	604	0.1
Trinucleotide	59	AAT	13,168	14.05	ACG	26	0.03
Tetranucleotide	232	CACG	2,649	8.41	ACCC/ACTT/ AGGT/CCAC /CCGA/CGAT/ TACG/TGGG	1	0.00
Pentanucleotide	574	TAATG	110	19.16	-	1	0.17
Hexanucleotide	608	CCCTAA	190	7.65	-	1	0.04