Genome-wide assessment and characterization of simple sequence repeats (SSRs) makers for *Capoeta aculeata* (Valenciennes, 1844) using NGS data

Gandomkar H.[1]; Nazari S.[1*]; Hosseini P.[2]; Abdolhay H.A.[2]

Received: January 2021 Accepted: June 2021

Abstract

The species *Capoeta aculeata* (Valenciennes, 1844) is one of the most important freshwater species endemics to Iran. However, the investigation of a population genetic structure of this species is limited by the low number of molecular markers currently described. In this study, we implemented next generation sequencing technology to identify polymorphic microsatellite markers and investigate the population genetic structure of *C. aculeata* sampled from three geographical sites in Iran. We sequenced 60 individuals from three populations occurring in the Zagros basin. We characterized and developed 36 novel polymorphic microsatellite markers and these loci were examined in 120 individuals from three populations occurring in the Zagros basin. The average number of alleles per locus varied from 1.7 to 16. (average=7.89). The results showed that, the polymorphism information content (PIC) of these SSR loci varied from 0.254 to 0.888. The observed heterozygosity (*Ho*) per locus ranged from 0.170 to 0.881, while the expected heterozygosity (*He*) per locus was from 0.170 to 0.881. Among these SSR loci, 20 loci deviated significantly from the Hardy–Weinberg equilibrium after Bonferroni correction ($p<0.05$). These microsatellite markers could provide a valuable tool for future population and conservation genetics studies of *C. aculeate* populations and other closely related species.

Keywords: *Capoeta aculeate*, Microsatellite markers, Next generation sequencing (NGS), Genetic structure

1-Shahid Motahary Cold-water Fishes Genetic and Breeding Research Center, Iranian Fisheries Sciences Research Institute, Agricultural Research, Education and Extension Organization, Yasouj, Iran

2-Iranian Fisheries Sciences Research Institute, Agricultural Research, Education and Extension Organization, Tehran, Iran

[*]Corresponding author's Email: Sajadnazari13@gmail.com

Introduction

The genus *Capoeta* is a herbivorous cyprinid genus and highly diversified with 14 species and widely distributed in water bodies of Western Iran. Kohgiluyeh and Boyer-Ahmad Province in the southwest part of Iran is a region with high amount of endemism in some freshwater fish species including *Capoeta aculeata* (Valenciennes, 1844) (Zareian *et al*., 2016). However, populations of this species have experienced significant declines over the last 20 years, due primarily to habitat loss and fragmentation caused by water storage and land clearing. The previous phylogenetic and phylogeographic studies found that populations of *C. aculeata* are different from the others (Freyhof *et al*., 2014).

Simple sequence repeat (SSR), also known as microsatellite markers are highly variable DNA sequences, distributed throughout eukaryotic genomes (Sunnucks, 2000; Zane *et al*., 2002). Given its co-dominant inheritance, bi-parental mode of inheritance, polymerase chain reaction (PCR)-based, high amplification success rate and levels of polymorphism, microsatellite markers have been used in a wide range of applications in population genetics, molecular ecology, conservation genetics and quantitative trait loci (QTL) identification studies (Chang *et al*., 2009; Xue *et al*., 2014; Montanari *et al*., 2016; Cheng *et al*., 2020). Nevertheless, the conventional methods for developing SSR markers are believed to be the key element which limits the application of microsatellite markers due to they are expensive, low efficiency, laborious and time-consuming (Jia *et al*., 2013; Perina *et al*., 2016). In fact, SSR markers are tremendously effective tools in population genetic studies because they might reveal the distinct population Segments (DPSs) even in fine-scale genetic structure studies. Until now, microsatellite loci have not been reported in this species (Gandomkar *et al*., 2020).

The major advantage of Next generation sequencing (NGS) technique is their ability to produce large amounts of sequence data and it can provide wide genetic information of simple sequence repeat (SSR) or microsatellite loci. Restriction-site associated DNA sequencing (RAD-seq) is a powerful tool to develop microsatellite and single nucleotide polymorphism (SNP) markers, which uses Illumina next-generation sequencing (NGS) platforms to discover informative volume of independent markers in a lot of specimens at the same time (Feng *et al*. 2014; Xiao *et al.,* 2018; Khoshkholgh and Nazari, 2020).

The reads created by RAD-seq can be grouped according to the enzyme recognition sequence, so this procedure could develop the accuracy and precision of contigs assembly, and consistently expand the success rate of polymorphic microsatellite markers development (Wei *et al.,* 2014; Ewers-Saucedo, 2016). In comparison with other techniques, restriction site-associated DNA sequencing (RAD-seq)

gradually become the most promising method and can give genome-wide genotypic and sequence data adjacent to single restriction enzyme cut sites and generate reduced representation genomic libraries (Davey *et al*., 2011; Zhou *et al*., 2017). Recently, the characterization and development of RAD-seq have been used to report microsatellite markers in many cyprinid species, such as *Ctenopharyngodon Idella* (Yu *et al*., 2014), *Schizothorax prenanti* (Luo *et al*., 2016) and *Xenocypris argentea (*Peng *et al*., 2018). However, the genetic background of *C. aculeata* is limited and no microsatellite locus has been characterized for this species. In this study, RAD-seq was used to isolate and develop the first resource of microsatellite loci for *C. aculeata* from Iran.

Materials and methods

In this study, fin clips of 120 individuals were collected from three wild *C. aculeata* populations inhabiting the Beshar, Khersan, Maroun rivers in Kohgiluyeh and Boyer-Ahmad Province in Iran. All procedures were approved by the Committee on Animal Care and Use, and Committee on the Ethics of Animal Experiments of Agricultural Research , Education and Extension Organization (AREEO) and Islamic Azad University. The modified method (Nazari *et al*., 2016) was used to extract genomic DNA using the commercial DN-easy blood and tissue kit (Qiagen™, Germany) following the manufacturer's instructions. The quality of DNA was assessed using a Nanodrop spectrophotometer (Nanodrop ND1000) and 1% agarose-gel electrophoresis, after which DNA concentration of each sample was diluted to 100 ng in a 100 μL volume and stored at –20°C (Khoshkholgh and Nazari, 2019).

Screening of microsatellite loci by PCR

DNA was isolated and purified from fin clips (about 30–50 mg) or the muscle samples of each fish. DNA was extracted using the commercial DN-easy blood and tissue kit (Qiagen™, Germany) following the manufacturer's instructions and standardized the final DNA concentration of each sample to 100 ng in a 10 μL volume (Nazari *et al*., 2016). DNA Extraction was examined for concentration using spectrophotometer (Nanodrop ND1000) and standardized to a specific concentration (for example, 100 ng/μL for Polymerase Chain Reaction (PCR)). Quality and concentrations of DNA samples were proved on 1.0% agarose gels and evaluated using a Nanodrop 1000 spectrophotometer.

RAD library construction

RAD library preparation was carried out following the protocol of (Wang *et al*., 2016). Briefly, genomic DNA from each individual was digested for 2 h at 37°C in a 40 μL reaction at specific sites with *EcoRI-HF* restriction enzymes according to the RAD protocol (Bian *et al*., 2018). Adapter P1 containing individual-specific nucleotide barcodes was ligated to the digested product. Ligation reaction was performed in a 40 μL volume along with 5 μL of ligation

buffer, 0.85 μL of P1 adapter, and 300 units of ligase for 2 h at 25°C. DNA was subsequently ligated to a second adapter (P2), which contains T overhangs. The ligated materials were then pooled, purified, eluted and subjected to PCR enrichment. Total genomic DNA samples were sheared to fragment DNA to an average size of 350 bp and a pool of adapter-ligated molecules with random, variable ends was generated. DNA fragments *with* lengths of 250–500 bp were isolated using the Qiagen Min Elute Gel Extraction kit (Qiagen™, Germany). The libraries constructed were sequenced on the Illumina HiSeq 4000 platform (BGI, China) using 150 bp paired-end reads.

The quality control of the raw sequences was initially evaluated with the program FastQC (Babraham Bioinformatics) (Andrews 2010). Reads were controlled with Cutadapt to eliminate potential ligated materials of Illumina adapters, permitting a 10% mismatch rate in the adapter sequence (Martin 2011). During this process, the *programs process_radtags* module and *clone_filter* from the *software Stacks* version 2.52 (http://catchenlab.life.illin ois.edu/stacks/) were *used* to demultiplex data and trim adapter sequences and remove over-represented sequences, respectively (Catchen *et al*., 2013; Rochette *et al*., 2019). To avoid low-quality reads with artificial bias, the following thresholds were set (Li and Godzik 2006), reads with adapter contamination were removed; reads with ≥10% unidentified nucleotides were removed; reads with >50% bases having phred quality <5 were removed; putative duplication reads generated by PCR amplification in library construction were discarded; reads were checked for presence of the partial 5 bp *EcoRI* motif (AATTC). De novo assembly was performed using software SPAdes (Bankevich *et al*., 2012). The contigs shorter than 100 bp were removed from the assembly. To evaluate the quality of assembly, the reads involved in alignment were re-mapped onto the assembled genome using BWA program (version 0.7.17) (Li and Durbin, 2009).

All clean sequences were aligned to the reference assembly using BWA program. Microsatellite motif repeats of mono to hexa-nucleotide microsatellites were identified from the assembled genome using MIcroSAtellite identification tool (MISA) (http://pgrc.ipk-gater slebe n.de/misa/). The following criteria of motifs were required to be met: at least 10 repetitive motifs for mononucleotide, 6 repetitive motifs for di-nucleotide, and 5 repetitive motifs for tri- to hexa-nucleotide. The number of microsatellite loci, repeat motifs, number of repeats, motifs sequence length, repeat type, start and end location of the repeat sequence, and microsatellite sequence, were analyzed using MISA.

A total of 36 polymorphic loci were selected for validation by PCR. Polymerase chain reaction (PCR) reactions were performed in a final volume of 25 μL consisting of 1.2 μL of template DNA (100 ng), 2.5 μL of

10×PCR buffer, 2.5 µL of Mg2+ (20 mM), 0.85 µL of dNTP mixture (2.5 mM each), 0.30 µL of each primer (10 pmol/µl), 17.35 µL of ddH2O, and 0.3 µL of 1 U/µL Taq DNA polymerase (Qiagen™, Germany). Amplified products were checked on 1% agarose gel electrophoresis and sequenced using ABI Prism 3730 automated DNA analyzer (Applied Biosystems). The PCR amplification procedure was performed in a thermal cycler (Bio-Rad, USA) using following program: first denaturation 94°C, 30 s, one cycle; denaturation 94°C, 30 s, 12 cycles; a specified annealing temperature for each primer pair for 30 s, 60 s, 12 cycles; then the products were extended at 72°C for 7 min (for annealing temperatures of each primer pair, see Table 3). The PCR products were electrophoresed on 1.2% agarose gels, stained with ethidium bromide followed by visualization under UV light. SSR loci with single bands were chosen as candidates for additional validation.

Table 1: Description of 36 microsatellite markers in *Capoeta aculeata* (Valenciennes, 1844) (F=forward; R=reverse)

Name	**Primer sequence (F and R, 5′-3′)**	**Allele size (bp)**	**Repeat motif**	**N_A**	**TM (°C)**	H_O	H_E	F_{IS}	PIC	***P*HWE**
Ca01	F: GGCGAACTTGTCTGAGTAAC R: GCATTGCAGATTCCAATA	202–206	(ATT)8	3.8	54	0.318	0.432	0.275	0.635	0.295
Ca02	F: CTCATTCCTTGGACTTGA R: CGTGCATAGATCCCATGCA	228–238	(GAT)9	4.3	56	0.854	0.879	-0.028	0.672	0.000*
Ca03	F: AGCTGGCTTGGGAAGCTG R: AGCGGTAGTAGTGAGTACTC	142–164	(CTT)4	6.8	56	0.722	0.745	-0.048	0.254	0.926
Ca04	F: TCTGCTAAGCACATCTACAGA R: GTCTGTCTCGCTTTCAGTCTC	98–106	(CGCT)4	1.7	55	0.212	0.268	-0.214	0.701	0.000*
Ca05	F: TGTACAGTATGTCGTCGTCTCA R: AGTGTTAGCGTTCGTAAGGG	274–298	(GAT)7	3.4	56	0.852	0.867	0.086	0.504	0.000*
Ca06	F: GCGTAGTTACGCGTTAGTCTA R: TGTACGAAGCTAAAGCAGCTG	94–110	(AT)10	5	58	0.766	0.784	-0.116	0.869	0.122
Ca07	F: TTGCTGAGCTTCGTTACGTGA R: CGAAATTCACACTATTCTGACT	100–116	(CA)7	6.6	58	0.484	0.642	-0.127	0.397	0.000*
Ca08	F: ACCGCGCTGATTCTGCTGATCG R: GACCGTCCACTATCTCTCGGA	192–206	(CT)8	2.4	54	0.214	0.261	0.115	0.588	0.048
Ca09	F: TAGGAGCCATAGTCTGCTCCTA R: ATCGTTACTCAGAAGTGCTCTG	242–274	(TG)10	3.8	54	0.224	0.298	0.365	0.766	0.000*

Table 3 (continued):

Name	Primer sequence (F and R, 5′-3′)	Allele size (bp)	Repeat motif	NA	TM (°C)	HO	HE	FIS	PIC	PHWE
Ca10	F: TTCGAATAGCTCATCACACCGC R: TCGTACACACGGAGGACTAGT	210–242	(AT)12	12	55	0.170	0.213	-0.031	0.656	0.273
Ca11	F: CTTGGTCCTGGCATTGAGTCT R: ATACTGGGCATGGGGCGTGA	264–286	(CA)8	3.2	58	0.702	0.746	-0.117	0.828	0.000*
Ca12	F: GTAGCTAGTCGACTCGATTACT R: ATGTAATCATATGTGACGCGA	202–236	(TGA)8	2.0	58	0.318	0.389	0.488	0.771	0.000*
Ca13	F: CACTTGCGCTAGGATTATGAT R: GTATCTGAGCTTCTGATCTGTC	238–276	(TGA)8	4.1	50	0.278	0.265	0.163	0.758	0.428
Ca14	F: ACTGCGTGGTATGGCTGATCC R: TTCGCGGCGACTATCTAGGGA	210–224	(TC)7	4.2	50	0.881	0.884	0.436	0.857	0.000*
Ca15	F: GACTACGTGCCCAGGCCAGTAC R: ATCTGTATGATGCCATTTGCG	246–282	(AAG)6	6.0	50	0.633	0.679	-0.130	0.835	0.000*
Ca16	F: TGCTCGTGAGTGGTGTAGTAC R: GTAAGTCTCCATGTGTGCGTAC	158–184	(ATC)6	8.4	52	0.593	0.489	0.015	0.747	0.178
Ca17	F: TATCAGTGATAGAGCTGATAG R: GTCATAGCTATCGAGTAGGTATC	86–96	(TGA)8	2.6	54	0.571	0.743	0.228	0.825	0.000*
Ca18	F: GGTTGCATCGTACTGTCAACAAT R: GGCTCGATAAGTGATACCTCA	254–278	(TG)6	3.8	54	0.774	0.627	0.286	0.867	0.387
Ca19	F: TTCCGCTAGCTATCAGCGGGA R: TACACAGCTCATATGGATGTTAG	112–128	(TCC)8	10	58	0.633	0.649	0.523	0.341	0.000*
Ca20	F: GTCTGACATACAGCATCCGGTC R: TCACTCTGACTGCTCTCACTCT	95–106	(AG)9	4.4	60	0.562	0.476	-0.123	0.643	0.188
Ca21	F: ACTTGAGCTACTGACATCGGTA R: TCCGATAGCTCTATGCGTGTG	180–204	(TC)8	4.0	58	0.419	0.664	0.606	0.830	0.000*
Ca22	F: CGATATGGAACGCTGGCGTTTG R: TCGGTTGCTCGGAGGTCCCCTA	227–263	(ATAC)6	3.1	50	0.318	0.390	-0.044	0.682	0.028
Ca23	F: ACTCGCCTAATCCGATCGGATC R: TTGCTAGTCATTCCAGTAGTA	158–194	(TC)12	5.4	52	0.579	0.805	-0.171	0.797	0.000*
Ca24	F: GCATGACTATCTCATCGGAGTA R: GTGTACGATGCGATCTTGGATC	102–136	(TATC)14	4.0	52	0.290	0.478	0.345	0.692	0.000*
Ca25	F: ACTAGCATACGGATTCAGTGTA R: TCGGTTATCGGTCCATATATGC	312–374	(TA)5	16	53	0.370	0.584	0.244	0.568	0.000*
Ca26	F: TTGCCATTAATCGTAGCGACA R: GCTACTTAAGCAGCCATCGTG	224–238	(AT)6	2.6	57	0.675	0.785	0.326	0.731	0.005
Ca27	F: CGTCGCCGGCTTTACATCTGGA R: GCTCGCTGCTCGCCATGGTCG	124–154	(AT)6	4.1	51	0.325	0.843	0.494	0.745	0.000*
Ca28	F: GATTCTCTAAGTACCTTTGTGT R: TAGACACTTGACATTCTATCG	114–140	(GT)5	14.0	57	0.525	0.713	0.283	0.634	0.000*

Table 3 (continued):

Name	Primer sequence (F and R, 5′-3′)	Allele size (bp)	Repeat motif	NA	TM (°C)	HO	HE	FIS	PIC	PHWE
Ca29	F: GGTGTCATCTCTCATACTAGC R: CTGTGTCGACATTGCTAGCCG	192–210	(AT)5	4.2	58	0.472	0.624	− 0.283	0.888	0.000*
Ca30	F: TGGTCGGAGTTAACGCTAGC R: GAGTACGTCGTGGCTCTAACA	134–174	(TA)5	2.0	54	0.551	0.643	− 0.215	0.582	0.094
Ca31	F: GGCTAACTACCTGAGGTTATG R: GTCGGGATATCGCATGTATCA	142–148	(CT)8	2.2	53	0.188	0.680	− 0.549	0.265	0.343
Ca32	F: AGGTGCATGCACAATCGCGTGC R: TAGCTCACTCACCGTCATGAG	162–168	(TA)14	4.4	55	0.207	0.642	− 0.074	0.468	0.288
Ca33	F: TCAGAGATGCTCGGATACTGCA R: ACCTTGGTGTATTGTATGAGCG	110–131	(TG)6	5.8	56	0.625	0.685	-0.495	0.726	0.009
Ca34	F: GCATACGCAGACTGTATTCGTG R: TACAGTCGATGCTAATCATGTC	121–154	(AATC)4	9.2	55	0.567	0.802	0.368	0.387	0.000*
Ca35	F: GTCTCACATCAATCGATGTGTG R: GAACGGATACTCACTTGAGTC	212–224	(AG)9	4.2	56	0.438	0.442	0.166	0.368	0.098
Ca36	F: AGTCCTAGTTCATCATCGCTAG R: CCGAGTTAGAATGAGCAGTAT	264–268	(TG)5	4.4	57	0.231	0.587	0.216	0.467	0.000*

Ho: observed heterozygosity, He: expected heterozygosity, N_A: number of alleles, F_{IS}: inbreeding coefficient, PIC: polymorphism information content; PHW: *P*-values for Hardy–Weinberg equilibrium. *Indicates significant deviation from Hardy–Weinberg Equilibrium after Bonferroni correction.

Validation of *microsatellite polymorphism*

For validated SSR loci, each of the selected primer sequences was validated with 120 single individual of *C. aculeata* from three sampling locations in Iran. From fragment sizing and reporting, GeneMapper (v4.1) was used to analyze genotype data. The negative controls with nuclease-free sterile water were included in each PCR experiment.

For validated loci, statistics including the number of alleles (*Na*), the effective number of alleles (*Ne*), the observed heterozygosity (H_O), expected heterozygosity (H_E) indices and inbreeding coefficient (F_{IS}) were calculated in GenAlex v6.5 (Peakall and Smouse 2012). The Hardy–Weinberg distribution (HWE) with Bonferroni correction was estimated at the population level using Arlequin v3.0 (Excoffier, 2005). The inbreeding coefficient index (F_{IS}) and the polymorphism information content (PIC) were estimated using Genepop V4.7 (Rousset, 2008).

Results

In the present study, the total GBS raw reads varied from 11.4 million to 22.38 million; with total raw nucleotides ranging from 0.93 Gb to 2.96Gb. GBS sequence analysis and microsatellite mining were performed using a modified GBS analysis workflow without using a reference genome (Table 1 and 2). After filtering low-quality reads and reads that

lacked enzyme cutting sites by the *process_radtags* component of Stacks v1.39 software, the total number of clean reads kept in the libraries varied from 10.81 to 21.74 million, with an average of 15.72 million. The summary of sequencing reads generated for samples of *C. aculeata* were listed in Table 2.

Of the 15,476 sequences, 16,706 microsatellite loci were found by MISA in *C. aculeata* genome sequences. The number of 926 sequences contains more than one microsatellite. Among the number of microsatellite loci, dinucleotide repeats were the most abundant (6843, 51.91%), followed by trinucleotide repeats (2991, 22.69%), mononucleotide repeats (2756, 20.91%) tetranucleotide repeats (478, 3.62%), pentanucleotide repeats (94, 0.71%), and hexanucleotide repeats (18, 0.13%) (Table 2).

In the present study, the average of the allele numbers inspected at each locus varied between 1.7 for locus *Ca04* to 16 for locus *Ca25*. The results showed that, the polymorphism information content (PCI) of these SSR loci varied from 0.254 to 0.888. The observed heterozygosity (*Ho*) per locus ranged from 0.170 to 0.881, while the expected heterozygosity (*He*) per locus was from 0.213 to 0.884. The final set of all microsatellite loci were most informative (PIC>0.50). Across all samples, 20 SSR loci showed significant deviations from the Hardy–Weinberg equilibrium after Bonferroni correction, ($p<0.05$) (Fig 1, Table 3).

Table 2: Number of raw reads, number and percentage of retained reads after filtering of the three population samples of *Capoeta aculeata* analysed by 2b-RAD.

Region	Raw reads	Filtered reads	%Of reads lost
Beshar	3,385,425	2,024,643	25.8
Khersan	3,285,345	2,965,851	19.8
Maroun	3,433,692	2,832,037	36.1

Table 3. Genetic diversity of the three population samples of *Capoeta aculeata* analysed by 2b-RAD. Per each sample the following information are reported: the relatedness expressed in terms of mean and variance using the Wang estimator, the observed and expected heterozygosity (Ho and He) and fixation index (F_{IS}) in the neutral and outlier loci datasets.

Sample ID	Relatedness Wang Mean & Variance	Neutral Loci			Outlier loci		
		Ho	He	F_{IS}	Ho	He	F_{IS}
Beshar	-0.36 (0.01)	0.18	0.21	0.08(NS)	0.23	0.27	0.11(NS)
Khersan	-0.34 (0.02)	0.19	0.24	0.12)	0.23	0.26	0.1
Maroun	-0.37 (0.03)	0.25	0.28	0.06(NS)	0.29	0.38	0.07(NS)

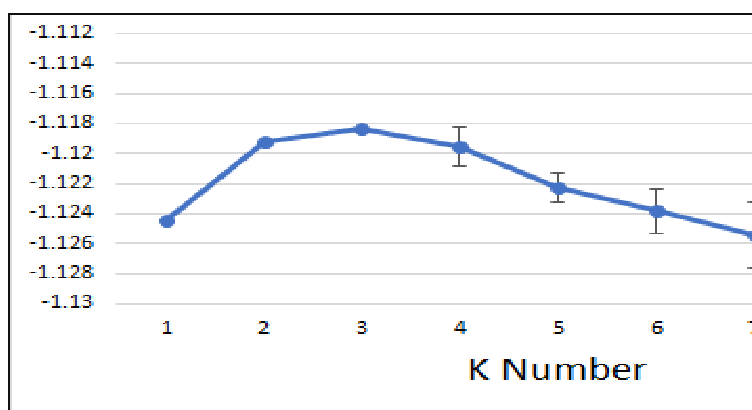

Figure 1: Population structure results using markers. The graph at k = 3 indicates the minimum number of subgroups possible in the diversity panel.

Discussion

For non-model organisms, microsatellite loci were generally identified by 5' anchor PCR (Ling *et al.*, 2013), transcriptome sequence analysis (Ariede *et al.*, 2018; Han *et al.*, 2018), and microsatellite library screening (Bazsalovicsová *et al.*, 2018) and development of microsatellite loci through cross-species amplification of closely related species (Gravley *et al.*, 2018). In the present study, RAD-seq method was used for isolating microsatellite loci in *C. aculeata*. In contrast with conventional approaches, RAD-seq approach is considered to be faster and profitable. Furthermore, this method can produce a large number of microsatellites markers for one time (Peng *et al.*, 2018; Xiao *et al.*, 2018).

Population structure genetic is very important for the sustainability of many species (Khoshkholgh and Nazari, 2019). Conservation management plans with no prior information of the genetic background could result in disturbance to the genetic structure with adverse effects on the gene pools of wild populations (Khoshkholgh and Nazari, 2020). Until now, the microsatellite loci have not been carefully developed for *C. aculeata* which has posed a serious obstacle to conservation and management of this species (Gandomkar *et al.*, 2020). The current study detected 16,706 microsatellite markers in *C. aculeata* genome using MISA, which accounted for 2.12% of the total genome sequence. The relative abundance of microsatellite sequences was calculated at 1.16 loci per kb of *C. aculeata* genomes. In general, the frequencies of microsatellite loci are expected to decrease with increasing repeat length due to longer repeats have a higher possibility of being mutated. (Pathak *et al.*, 2013), and this tendency has been identified in many organisms (Guichoux *et al.*, 2011; Zalapa *et al.*, 2012; Ma *et al.*, 2013) . In summary, this is the first report of SSR loci in *C. aculeate* developed using RAD-seq and therefore these SSR markers would provide an invaluable resource for population genetics and natural resource conservation in *C. aculeata*.

References

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Ariede, R.B., Freitas, M.V., Hata, M.E., Matrochirico-Filho, V.A., Utsunomia, R., Mendonça, F.F., Foresti, F., Porto-Foresti, F., Hashimoto, D.T. 2018. Development of microsatellite markers using next-generation sequencing for the fish *Colossoma macropomum*. *Molecular Biology Reports*, 45, 9–18. doi: 10.1007/s11033-017-4134-z

Bankevich, A., Nurk S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal in Computational Biology*, 19, 455–477

Bazsalovicsová, E., Koleniˇcová, A., Králová-Hromadová, I., Minárik, G., Šoltys, K., Kuchta, R. and Štefka, J., 2018. Development of microsatellite loci in zoonotic tapeworm *Dibothriocephalus latus* (Linnaeus, 1758), Lühe, 1899 (syn. *Diphyllobothrium latum*) using microsatellite library screening. *Molecular and Biochemical Parasitology*, 2018, 225, 1–3.

Catchen, J.P., Hohenlohe, S.B., Amores, A. and Cresko, W., 2013. *Stacks:* an analysis tool set for population genomics. *Molecular Ecology*, 22(**11**), 3124-3140.

Chang, Y., Feng, Z., Yu, J. and Ding, J., 2009. Genetic variability analysis in five populations of the sea cucumber *Stichopus* (*Apostichopus*) *japonicus* from China, Russia, South Korea and Japan as revealed by microsatellite markers. *Marine Ecology,* 30, 455–461.

Cheng, J., Hui, M., Li, Y. and Sha, Z., 2020. Genomic evidence of population genetic differentiation in deep-sea squat lobster *Shinkaia crosnieri* (crustacea: Decapoda: Anomura) from Northwestern Pacific hydrothermal vent and cold seep. *Deep-Sea Research Part I*, 156, https://doi.org/10.1016/j.dsr.2019.103188.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510

Ewers-Saucedo, C., Zardus, J.D. and Wares, J.P., 2016. Microsatellite loci discovery from next-generation sequencing data and loci characterization in the epizoic barnacle *Chelonibia testudinaria* (Linnaeus, 1758). *PeerJ,* 4, e2019.

Excoffier, L., 2005. Arlequin (version 3.0), an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, https ://doi.org/10.1143/JJAP.34. L 418

Feng, N., Ma, H., Ma, C. Shujuan, L, Wei, J. Yuexing, L, Lingbo, M. 2014. Characterization of 40 single nucleotide polymorphism (SNP) via T m-shift assay in the mud crab (*Scylla paramamosain*). *Molecular Biology Reports*, 41, 5467–5471. https ://doi. org/10.1007/s1103 3-014-3420-2

Freyhof, J., Esmaeili, H.R., Sayyadzadeh, G. and Geiger, M., 2014. Review of the crested loaches of the genus Paracobitis from Iran and Iraq with the description of four new species (Teleostei:Nemacheilidae). *Ichthyol ogical Exploration of Freshwaters,* 25(**1**),11–38

Gandomkar, H., Shekarabi, S.P.H., Abdolhay, H.A., Nazari, S. and Mehrjan, M.S., 2020. Genetic structure of the *Capoeta aculeata* populations inferred from microsatellite DNA loci. *Biodiversitas*, 21, 4565-4570. https://doi.org/10.13057/biodiv/d211014

Gravley, M.C., Pierson, B.J., Sage, G.K., Ramey, A.M. and Talbot, S.L., 2018. DNA microsatellite markers for northern fulmar (*Fulmaris glacialis*) and cross-species amplifcation in select seabird species, Alaska 2018. U.S. Geological Survey data release.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F. and Petit, R.J., 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources*,11, 591–611

Han, Z., Xiao, S., Li, W., Ye, K. and Wang, Z.Y., 2018. The identification of growth, immune related genes and marker discovery through transcriptome in the yellow drum (*Nibea albiflora*). *Genes Genomics*, 40, 881–891. DOI: 10.1007/s13258-018-0697-x

Jia, S.W., Liu, P., Li, J., Li, J.T. and Pan, L.Q., 2013. Isolation and characterization of polymorphic microsatellite loci in the ridgetail white prawn *Exopalaemon carinicauda*. *Genetics and Molecular Research*, **12**, 2816–2820.

Khoshkholgh, M. and Nazari, S., 2019. The genetic diversity and differentiation of narrow-clawed crayfish *Pontastacus leptodactylus* (Eschscholtz, 1823) (Decapoda: Astacidea: Astacidae) in the Caspian Sea Basin, Iran as determined with mitochondrial and microsatellite DNA markers. *Journal of Crustacean Biology*, 39(**2**), 112–120.

Khoshkholgh, M. and Nazari, S., 2020. Characterization of single nucleotide polymorphism markers for the narrow-clawed crayfish *Pontastacus leptodactylus* (Eschscholtz, 1823) based on RAD sequencing. *Conservation Genet Resour* . https://doi.org/10.1007/s12686-020-01154-8

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with

Burrows-Wheeler Transform. *Bioinformatics*, 25, 1754–1760

Li. W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics. 22. 1658-1659. DOI: 10.1093/bioinformatics/btl158

Ling, L.P., Adibah, A.B., Tan S. G., Christianus, A. and Faridah, Q. Z. 2013. Isolation by the 5′ anchored PCR technique and characterization of eighteen microsatellite loci in horseshoe crab (*Tachypleus gigas*). *Journal of Genetics,* 92**,** 101–104. https://doi.org/10.1007/s12041-011-0115-5

Luo, H., Ye, H., Xiao, S.J., *et al*. 2016. Development of SNP markers associated with immune-related genes of *Schizothorax prenanti*. *Conservation Genetics Resources*, 8, 223–226.DOI:10.1007/s12686-016-0539-6.

Ma, H., Zou, X., Ji, X., Ma, C., Lu, J., Jiang, W., Xia, L., Li, S., Liu, Y., Gong, Y. and Ma, L., 2013. Discovery and characterization of a first set of polymorphic microsatellite markers in red crab (*Charybdis feriatus*). *Journal of Genetics*, 92, e113–e115

Martin, M., 2011. Sequencing Reads. *EMBnet Journal*, 17, 10–12

Montanari, S., Perchepied, L., Renault, D., Frijters, L., Velasco, R., Horner, M., Gardiner, S.E., Chagné, D., Bus, V.G.M., Durel, C.E. and Malnoy, M., 2016. A QTL detected in an inter specific pear population confers stable fire blight resistance across different environments and genetic backgrounds. *Molecular Breeding*, 36, 1–16.

Nazari, S., Jafari, V., Pourkazemi, M., Kolangi Miandare, H. and Abdolhay, H., 2016. Association between myostatin gene (MSTN-1) polymorphism and growth traits in domesticated rainbow trout (*Oncorhynchus mykiss*). *Agri Gene*, 1, 109–115.

Pathak, R.U., Mamillapalli, A., Rangaraj, N., Kumar, R.P., Vasanthi, D., Mishra, K. and Mishra, R.K., 2013. AAGAG repeat RNA is an essential component of nuclear matrix in Drosophila. *RNA Biology*, 10(**4**),564–71.

Peakall, R. and Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research_an update. *Bioinformatics*, 28,2537–2539.

Peng, X., Zhao, L., Liu, J. and Guo, X., 2018. Development of SNP markers for *Xenocypris argentea* based on transcriptomics. *Conservation Genetics Resources*, 10(**4**), 679–684. https ://doi.org/10.1007/s12686-017-0900-4

Perina, A., Gonzلlez-Tiz◌́n, A.M., Vizcano, A., Gonzلlez-Orteg◌́n, E. and Martinez-Lage, A., 2016. Isolation and characterization of 20 polymorphic microsatellite loci in *Palaemon serratus* and cross-amplification in *Palaemon* species by 454 pyrosequencing. *Conservation Genetics Resources,* 8, 169–196

Rochette N.C. Rivera-Colón, A. G. Catchen, J. M. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. Resour., 28, 4737–4754. https://doi.org/10.1111/mec.15253

Rousset, F., 2008. Genepop'007: a complete re-implementation of the genepop software for windows and linux. *Molecular Ecology Resources*, 8, 103–106.

Sunnucks, P., 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, 15,199–203.

Wang, S., Liu, P., Lv, J., Li, Y., Cheng, T. and Zhang, L., 2016. Serial sequencing of iso-length rad tags for cost-efficient genome-wide profiling of genetic and epigenetic variations. *Nature Protocols*, 11(**11**),2189–2200. DOI:10.1038/nprot .2016.133

Wei, N., Bemmels, J.B. and Dick, C.W., 2014. The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. *Molecular Ecology Resources*, 14, 953–965.

Xiao, M., Hu, Q., Zhao, Y., Bao, F., Cui, F. and Zheng, R., 2018. Development of 36 SNP markers in *Ophiocephalus argus* Cantor base on high throughput sequencing. *Conservation Genetics Resources*, 10(**1**), 35–38. https://doi.org/10.1007/s1268 6-017-0757-6

Xue, D., Zhang, T. and Liu, J., 2014. Microsatellite evidence for high frequency of multiple paternity in the marine gastropod *Rapana venosa*. *PloS ONE*, 9, e86508.

Yu, L., Bai, J., Cao, T. *et al* 2014. Genetic variability and relationships among six grass carp *Ctenopharyngodon idella* populations in China estimated using EST-SNP markers. *Fisheries Science*, 80, 475–481. DOI:10.1007/s1256 2-014-0709-y

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R. and Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany*, 99:193–208

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

Zareian, H., Esmaeili, H.R. and Freyhof, J., 2016. *Capoeta anamisensis*, a new species from the Minab and Hasan Langhi River drainages in Iran (Teleostei: Cyprinidae). *Zootaxa,* 4083, 126-142.

Zhou, J., Zhou, B., Li, Q., Zhang L., Dum, J., Ye, H. 2017. Isolation and characterization of 33 EST-SNP markers in *Schizothorax prenanti*. *Conservation Genetics Resources*, 10, 205–207. https ://doi.org/10.1007/s1268 6-017-0799-9.