KFAS 한국수산과학회지 Korean Journal of Fisheries and Aquatic Sciences
한수지 58(4), 280-288, 2025

Original Article
Korean J Fish Aquat Sci 58(4),280-288,2025

차세대 염기서열 분석을 활용한 아홉동가리(*Goniistius zonatus*)의 다형성 Microsatellite 마커 개발 및 특성 분석

동춘매 · 이미난 · 노은수 · 박진우[1] · 김군도[2] · 황인준 · 강정하 · 정효선*

국립수산과학원 생명공학과, [1]국립수산과학원 아열대수산연구소, [2]국립부경대학교 미생물학과

Development and Characterization of Polymorphic Microsatellite Markers in *Goniistius zonatus* Using Next Generation Sequencing

Chun Mae Dong, Mi-Nan Lee, Eun Soo Noh, Jin Woo Park[1], Gun-Do Kim[2], In Joon Hwang, Jung Ha Kang and Hyo Sun Jung*

Biotechnology Research Division, National Institute of Fisheries Science, Busan 46083, Republic of Korea
[1]Subtropical Fisheries Research Institute, National Institute of Fisheries Science, Jeju 63068, Republic of Korea
[2]Department of Microbiology, College of Natural Sciences, Pukyong National University, Busan 48513, Republic of Korea

This study aimed to analyze the genetic diversity of *Goniistius zonatus* by developing microsatellite markers and evaluating their population genetic characteristics. Among a total of 188,476 genome contigs, microsatellite sequences were detected in 39,127 contigs, which contained a total of 48,789 candidate microsatellite markers. After removing redundant simple sequence repeat (SSR) motifs, 34,117 candidate markers were initially selected. Subsequently, polymorphism analysis, including the presence of insertions and deletions (InDels) within the SSR marker regions, resulted in the final selection of 3,648 markers. Among these 3,648 markers, 146 microsatellite markers were initially chosen based on the type of repeat motif, primer size, and amplification product location. PCR amplification and fragment size evaluation further refined the selection to 20 markers. The genetic characteristics of the *Goniistius zonatus* population were analyzed using the 20 selected microsatellite markers. The mean number of effective alleles was 9, ranging from 5 to 15. The observed heterozygosity (HO) and expected heterozygosity (HE) were found to be 0.755 and 0.766, respectively. The polymorphic information content (PIC) values ranged from 0.546 to 0.861, with an average of 0.722, indicating that these markers provide sufficient individual identification power for accurately analyzing the genetic diversity of the *Goniistius zonatus* population.

Keywords: *Goniistius zonatus*, Next generation sequencing, Microsatellite markers, Genetic variability

서 론

기후변화로 인한 해수면 온도 상승은 우리나라 연근해 수산자원에 중대한 영향을 미치고 있다(Han et al., 2023). 국립수산과학원의 2024 수산분야 기후변화 영향 및 연구 보고서에 따르면, 1968년부터 2023년까지 약 56년 동안 우리나라 해역의 표층 수온이 1.44℃ 상승한 것으로 나타났다(NIFS, 2024). 이는 전 세계 평균 수온 상승치(약 0.72℃)의 두 배에 해당하며, 우리나라 해양 생태계가 기후변화에 더욱 민감하게 반응하고 있음을 시사한다. 이러한 기후변화는 해양 생태계의 종 분포 및 군집 구조에도 큰 영향을 미치고 있다. 특히 제주 연안에서는 아열대성 어종의 출현 비율이 지속적으로 증가하고 있으며, 최근 10년(2012–2021) 동안 관찰된 전체 어종 중 약 42%가 아열대성 어종으로 확인되었다(NIFS, 2022). 주요 아열대성 어종으로는 호박돔(*Choerodon azurio*), 아홉동가리(*Goniistius zonatus*), 황놀래기(*Pseudolabrus sieboldi*), 무점황놀래기(*Pseudolabrus eoethinus*), 청줄돔(*Chaetodontoplus septentrionalis*) 등이 있으며, 이들의 출현 빈도는 점진적으로 증가하는 추세를 보이고

***Corresponding author:** Tel: +82. 51. 720. 2452 Fax: +82. 51. 720. 2456
E-mail address: jhs3010@korea.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.5657/KFAS.2025.0280
Korean J Fish Aquat Sci 58(4), 280-288, August 2025
Received 30 May 2025; **Revised** 22 July 2025; **Accepted** 28 July 2025
저자 직위: 동춘매(연구원), 이미난(연구원), 노은수(연구사), 박진우(연구사), 김군도(교수), 황인준(연구관), 강정하(연구관, 생명공학과장), 정효선(연구사)

있다.

그 중 아홉동가리는 기후변화로 인한 생태계 변화를 반영하는 대표적인 어종으로 주목 받고 있다(NIFS, 2022). 아홉동가리(*Goniistius zonatus*)는 농어목(Perciformes) 줄양놀래기과(Latridae)에 속하는 어류로, 일본, 한국 남해 및 제주도, 대만 등 서태평양 연안에 넓게 분포하고 있으며(Nakabo, 2002; Burridge and Smolenski, 2004), 제주 지역에서는 '꽃돔'이라는 명칭으로 불린다.

최근에는 소비 수요가 급격히 증가함에 따라, 상업적 가치 또한 크게 상승하고 있다. 이러한 아홉동가리의 자원을 지속 가능하게 관리하기 위해서는 체계적인 보존 전략이 필수적이다. 그러나 아홉동가리에 대한 기존 연구는 일본에서 수행된 먹이 및 행동 생태에 관한 일부 연구(Matsumoto, 2001; Matsumoto and Kohda, 2002)에 한정되어 있으며, 국내외적으로는 해당 종의 보존을 위한 유전체 수준의 연구가 아직 보고되지 않았다. 이에 따라 아홉동가리 자원의 지속 가능한 이용과 효과적인 보존을 위해 유전체 분석 기반의 과학적이고 체계적인 자원 관리 전략 수립이 시급하다.

2000년대 이후 차세대 염기서열 분석(next-generation sequencing, NGS) 기술의 도입으로 유전체 분석 기법이 비약적으로 발전하였다. NGS를 활용하면 대상 종의 전장유전체(whole genome sequencing, WGS) 서열을 대량으로 확보할 수 있어 기존 분자 마커 개발 방법의 한계를 극복할 수 있다(Paetkau, 1999). 특히, 전통적인 microsatellite 마커 개발 방식과 달리, NGS 기반 분석을 통해 반복서열 영역을 직접 탐색하여 높은 다형성(polymorphism)과 재현성(reproducibility)을 갖는 마커를 효율적으로 개발할 수 있다(Mardis, 2008; Gardner et al., 2011). NGS를 이용해 개발된 microsatellite 마커는 높은 유전자 다양성을 보이며, 집단유전학 연구, 유전자 지도 작성, 개체 식별, 친족 관계 분석 등 다양한 유전학적 연구 분야에서 활용되고 있다(Park et al., 2014).

현재 수산자원을 대상으로 한 NGS 기법 연구가 활발히 이루어지고 있으며, 보리새우(*Marsupenaeus japonicus*) (Zhong et al., 2015), 방어(*Seriola quinqueradiata*) (Dong et al., 2020), 홍해삼(*Apostichopus japonicus*) (Lee et al., 2022), 굴(*Crassostrea gigas*) (Dong et al., 2022) 등의 유전학적 특성이 분석되었다. 그러나 아홉동가리에 대한 NGS 기반 유전학적 연구는 아직 보고되지 않고 있다.

따라서 본 연구는 아홉동가리의 지속 가능한 자원 관리 및 종 보존을 목적으로 NGS 기법을 활용하여 새로운 microsatellite 영역을 탐색하고, 이를 기반으로 다형성과 재현성이 높은 microsatellite 마커를 개발하였다. 개발된 microsatellite 마커를 사용하여 아홉동가리 집단의 유전적 다양성과 유연관계를 분석하였으며, 이를 통해 아홉동가리의 유전학적 기초 정보를 확보하고 효율적인 자원 관리 및 보존에 기여하고자 한다.

재료 및 방법

시료 확보 및 genomic DNA 추출

본 연구에 사용된 아홉동가리 시료는 제주도 내 4개 지역(신창리 2024년, n=5; 북촌리 2024년, n=16; 사계리 2024년, n=27; 신흥리 2024년, n=2)에서 정치망으로 포획된 자연산 아홉동가리 50개체를 대상으로 분석에 활용하였다.

Genomic DNA의 순수 분리 및 정제는 DNeasy® 96 Blood & Tissue Kit (Qiagen GmbH, Hilden, Germany)를 사용하여 수행하였다. 각 시료 20 mg을 E-tube에 담고, 제조사의 지침에 따라 ATL buffer (200 μL)와 Proteinase K (20 μL)를 첨가한 후, 56°C에서 12시간 동안 반응시켰다. 이후 AL buffer (200 μL)와 99% ethanol (200 μL)을 첨가하여 혼합한 후, S-Block이 장착된 DNeasy 96 plate로 옮겨 8,000 rpm (6,000 *g*)에서 30초간 원심분리하였다.

그 후, column plates를 교체하고 AW1 buffer (550 μL)를 첨가한 뒤, 동일한 조건에서 30초간 원심분리하였다. 다시 column plates를 교체한 후, AW2 buffer (550 μL)를 첨가하여 20,000 rpm (14,000 *g*)에서 5분간 원심분리하였다. Ethanol을 완전히 제거하기 위해 column을 실온에서 건조한 후, AE buffer (80 μL)를 첨가하여 genomic DNA를 회수하였다.

회수된 genomic DNA는 2.0% (w/v) agarose gel 전기영동(E-Graph Gel Documentation System; ATTO, Daejeon, Korea)을 통해 DNA 밴드의 존재 여부를 확인하였으며, NanoPhotometer N60 Touch (Implen GmbH, München, Germany)를 사용하여 DNA 농도를 측정하였다.

유전체 서열 구축 및 유전자 위치 예측

아홉동가리의 microsatellite 마커 개발을 위해, 2024년 제주도 사계리 집단(n=27) 중 4개체를 대상으로 NGS를 수행하였다.

추출된 DNA는 Agilent 4200 TapeStation을 통해 DNA 품질 관리를 진행하였다. TruSeq Nano DNA (350) 라이브러리 구축을 위해 DNA를 파편화(fragmentation)한 후 양 끝에 어댑터를 부착하여 라이브러리를 구축하고, TapeStation D1000 ScreenTape를 사용하여 라이브러리 품질을 평가하였다. Illumina NovaSeq 6000 플랫폼을 이용하여 paired-end (2 × 151 bp) 방식으로 시퀀싱을 진행하였다.

아홉동가리의 WGS 데이터에서 낮은 시퀀싱 품질의 데이터를 제거하기 위해 BBDuk (v38.96) 프로그램을 활용하여 어댑터 트리밍(adapter trimming) 및 품질 관리(quality control)를 수행하였으며, 이를 통해 고품질 데이터만을 선별하였다. 선별된 1차 고품질 데이터는 Bowtie2 (Langmead and Salzberg, 2012)를 사용하여 박테리아 및 미토콘드리아 데이터베이스에 대해 서열 유사성 기반의 정렬(alignment)을 수행하고, 이를 통해 오염물질을 제거(decontamination)하였다.

이후 각 개체의 고품질 서열을 사용하여 k-mer 분석을 통해 아홉동가리 유전체 크기를 예측하였다. 아홉동가리 reference genome sequence 구축을 위해 CLC Assembly Cell (v5.2.1; QIAGEN Bioinformatics, Aarhus, Denmark)의 CLC Assembler를 사용하여 *de novo* assembly를 진행하였다. 유전체 내 유전자 구조 예측을 위해 ab initio gene prediction 방법을 적용하였으며, 최종 유전자 모델의 단백질 서열에 대해 homology-based annotation을 수행하였다. 이 과정에서 UniProt (The UniProt Consortium, 2019) 데이터베이스와 BLAST (Altschul et al., 1990)를 활용하였다.

SSR 마커 후보 선발 및 primer 디자인

아홉동가리 유전체 내 SSR (simple sequence repeat) 마커 탐색은 SSRIT (simple sequence repeat identification tool)을 사용하였다. SSRIT는 설정된 기준에 따라 유전체 서열 내의 microsatellite 반복서열을 식별하고 예측하였다.

예측된 SSR은 motif length와 반복 수를 기준으로 정리되었으며, 마커 개발을 위해 유전체 내 SSR의 위치 정보를 annotation하였다. 또한, SSR이 위치한 영역을 3'-UTR, 5'-UTR, intron, CDS (coding sequence), intergenic region으로 구분하여 정리하였다.

SSR genotyping을 위한 PCR primer는 Primer3 (v2.3.6) 소프트웨어(Untergasser et al., 2012)를 사용하여 설계하였으며, SSR을 중심으로 양쪽 인접 서열의 보존 영역(flanking sequences)을 대상으로 primer를 디자인하여 안정적인 서열 증폭이 이루어질 수 있도록 하였다. 각각의 SSR을 대상으로 amplicon size를 100, 200, 300, 400 bp로 설정하여 multiplex PCR이 가능하도록 primer를 디자인하였다. 디자인된 primer 서열은 아홉동가리 유전체 서열에 매핑하여 의도한 단일 영역이 증폭되는지 확인하였다.

시퀀싱을 통해 확보한 아홉동가리 개체별 유전체 데이터를 이용하여 SSR 영역 내 다형성을 분석하였다. *De novo* assembly로 구축된 아홉동가리 대표 유전체 contig에 각 개체별 시퀀싱 서열을 매핑한 후, SSR 영역에서 InDel을 포함한 다형성 여부를 확인하였다. 이를 위해 BWA (v0.7.12; Broad Institute, Cambridge, MA, USA) (Li and Durbin, 2009)의 MEM 알고리즘을 사용하여 매핑을 진행하였고, SAMtools (v1.10; Broad Institute)과 BCFtools (v1.9; Wellcome Sanger Institute, Hinxto, UK)를 사용해 다형성을 분석하였다(Danecek et al., 2021). 최종적으로 SSR 영역에서 개체별 다형성이 확인되었으며, 안정성이 검증된 primer 서열을 SSR 마커 후보로 선별하였다.

Microsatellite 마커 선별

NGS를 통해 탐색된 microsatellite 영역의 증폭 여부를 확인하기 위해 2024년 제주도 사계리 집단(n=27) 중 16개체를 대상으로 PCR 분석을 실시하였다. PCR 조성은 10X Ex Taq buffer (Mg^{2+} plus, 20 mM, 1 mL) 1 μL, dNTP Mixture (2.5 mM each, 800 μL) 0.8 μL, TaKaRa Ex Taq (5 U/μL, 250 U) 0.1 μL, forward primer와 reverse primer 각각 0.3 μL (10 mM), template DNA 0.5 μL를 포함하여 최종적으로 10 μL의 반응 용액을 준비하였다. 그 후, Veriti™ 96-Well Fast Thermal Cycler에서 PCR을 진행하였다. PCR 조건은 95°C에서 7분간 사전 변성(preincubation) 후, 95°C에서 45초 동안 변성(denaturation)을, 60°C에서 45초 동안 primer 결합(annealing)을, 72°C에서 45초 동안 DNA 합성(extension)을 진행하였다. 이 과정은 총 35회 반복되었고, 마지막으로 72°C에서 5분간 최종 DNA 합성(full extension)을 실시하였다. PCR이 완료된 후, 2.0% agarose gel을 사용한 전기영동을 통해 증폭된 DNA 밴드의 유무를 확인하였다.

이후, 증폭된 후보 microsatellite 마커의 forward primer 정방향 서열에 형광물질(TAMRA 및 6-FAM, HEX)을 결합하여 PCR을 다시 수행하였다. PCR 증폭 산물에 GeneScan 400HD ROX size standard와 Hi-Di Formamide를 혼합한 후, 95°C에서 3분간 변성시켰다. 그 후, ABI PRISM 3730XL Genetic Analyzer를 사용하여 유전자 단편 크기(fragment size)를 분석하였다. 이 과정을 통해 아홉동가리의 microsatellite 마커가 안정적으로 증폭되었는지 확인하고, 형광 표지를 통해 증폭 산물의 크기를 정확히 측정하였다.

Microsatellite 마커의 효율성 검증

최종 선발된 microsatellite 마커의 효율성을 검증하기 위해 제주도 내 4개 지역(신창리 2024년, n=5; 북촌리 2024년, n=16; 사계리 2024년, n=27; 신흥리 2024년, n=2)에서 정치망으로 포획된 자연산 아홉동가리 50개체를 대상으로 분석을 수행하였다.

Arlequin version 3.1 소프트웨어(Excoffier et al., 2005)와 GENEPOP version 4.0 컴퓨터 패키지(Rousset, 2008)를 사용하여 대립유전자 수(number of alleles, NA)와 유전자 좌위별 이형접합도(heterozygosity)를 계산하였다. 이형접합도는 관찰된 이형접합률(observed heterozygosity, HO)과 Hardy-Weinberg 균형에서 기대되는 이형접합률(expected heterozygosity, HE)로 계산되었다.

또한, microsatellite의 다형성 정보 지수(polymorphic information content, PIC)는 대립인자 빈도 분석법을 적용하여 Cervus ver. 3.0.7 프로그램(Kalinowski et al., 2007)을 사용하여 산출하였다. 대립유전자 크기가 결정된 후에는 Micro-Checker 프로그램(Oosterhout et al., 2004)을 사용하여 null allele에 의한 유전자형(genotyping) 오류가 발생했는지를 검증하였다.

결과 및 고찰

유전체 서열 분석 및 유전자 예측

아홉동가리의 microsatellite 마커 개발을 위해 2024년 제주

도 사계리 집단(n=27) 중 4개체를 대상으로 NGS를 수행하였다. 각 개체의 게놈 DNA를 분석하기 위해 총 DNA 양과 DNA 무결성 지수(DNA integrity number, DIN)를 측정한 결과, 총 DNA 양이 0.100 μg 이상이며 DIN 값이 7 이상으로 확인되었다. 이에 따라 모든 개체가 NGS 분석에 적합한 품질 기준을 충족하는 것으로 판단되었다. DNA를 효소 처리하여 파편화한 후, 어댑터 라이게이션(adapter ligation)을 수행하여 TruSeq Nano DNA (350) 라이브러리를 제작하였다. 제작된 라이브러리는 평균 637 bp 크기로 확인되었으며, 모든 개체가 품질 검사를 통과하였다. 이후 WGS 시퀀싱은 Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) 플랫폼을 이용하여 paired-end (2×151 bp) 방식으로 수행되었으며, 그 결과 개체별 평균 11.39 Gb의 유전체 데이터가 확보되었다. 시퀀싱 데이터 전처리 과정에서는 BBDuk (v38.96) 프로그램을 사용하여 어댑터 트리밍 및 품질 관리를 수행한 결과, 낮은 품질의 시퀀싱 데이터 약 7.50%가 제거되었다. 이후, Bowtie2 (v2.2.28)를 이용한 서열 유사성 분석을 통해 박테리아 서열이 평균 6.63%, 미토콘드리아 서열이 0.05% 포함되어 있음이 확인되었다(Table 1).

개체 간 시퀀싱 데이터 양에는 큰 차이가 없었으나, CHEZO2 개체가 가장 많은 69,131,959개의 서열을 확보하였다. 확보된 고품질 데이터를 활용하여 *k-mer* 분석(*k*=17)을 수행한 결과, 단일 피크(single peak)가 확인되었으며, 이를 바탕으로 평균 유전체 크기는 553,135,905 bp로 예측되었다(Table 1).

가장 많은 유전체 서열을 확보한 CHEZO2 개체의 데이터를 활용하여 CLC Assembler를 사용해 아홉동가리의 *de novo* assembly를 진행하였다. assembly 후, clean paired-end 데이터를 사용하여 error correction을 수행한 결과, 188,476개의 contig로 구성된 536,816,551 bp 크기의 유전체를 완성하였다(Table 2).

아홉동가리의 유전자 모델을 예측하기 위해 Augustus (v3.1) 프로그램(Stanke and Morgenstern, 2005; Stanke et al., 2006)을 사용하였으며, Zebrafish의 training HMM을 기반으로 분석을 진행하였다. 이 분석을 통해 총 72,802개의 유전자가 예측되었고, 이들의 평균 길이는 3,051.92 bp로 확인되었다(Table 3). 또한, 예측된 유전자 중 14,173개는 UniProt 데이터베이스를 활용하여 기능이 예측되었으며, 이들 유전자는 E-value 1e-5 이하 및 similarity percentage 70% 이상의 기준을 충족하였다.

SSR 후보 마커 대량 탐색

아홉동가리의 유전체 서열을 바탕으로 microsatellite 마커 후보를 대량으로 탐색하기 위해 motif 길이에 따른 반복 횟수 조건을 설정하였다. SSR 마커로 정의되기 위해 각 motif의 반복 기준을 다르게 적용하였으며, di-nucleotide (2개 염기) 반복 서열은 9회 이상, tri-nucleotide (3개 염기) 반복 서열은 6회 이상, tetra-nucleotide (4개 염기) 반복 서열은 5회 이상, penta-nucleotide (5개 염기) 반복 서열은 4회 이상, hexa-nucleotide (6개 염기) 및 hepta-nucleotide (7개 염기) 반복 서열은 각각 3회 이상 반복될 경우 SSR 마커로 정의하였다.

SSR 마커 후보를 선별한 결과, 전체 188,476개의 genome contig 중 39,127개에서 microsatellite가 감지되었고, 이들 contig에는 총 48,789개의 microsatellite 마커 후보가 포함되어 있었다(Table 4). 또한, 두 개 이상의 microsatellite가 감지된 genome contig는 9,713개였으며, 각 microsatellite motif type의 분포는 di-nucleotide가 29,467건, tri-nucleotide가 10,098건, tetra-nucleotide가 3,652건, penta-nucleotide가 2,799건, hexa-nucleotide가 5,126건, hepta-nucleotide가 1,249건으로 확인되었다(Table 5). 이 중 di-nucleotide의 비율이 가장 높았으며, hepta-nucleotide가 가장 낮은 비율을 나타냈다. 이는 NGS

Table 1. Whole genome sequencing data quality and genome size estimation statistics

	Total		Trimmed	Contamination		Cleaned					
	Sequences	Reads	Low quality (%)	Bacteria (%)	Mitochondria (%)	Sequences	Reads	Sequences (%)	Reads (%)	Hash length (*k-mer*)*	Estimated genome size (bp)
CHEZO1	11,571,700,780	76,633,780	7.0	6.74	0.5	9,893,024,255	66,153,393	85.49	86.32	17	552,360,521
CHEZO2	12,032,222,392	79,683,592	7.0	6.45	0.49	10,333,164,944	69,131,959	85.88	86.76	17	542,562,549
CHEZO3	10,873,015,928	72,006,728	9.0	6.77	0.51	9,271,007,215	62,027,597	85.27	86.14	17	551,697,521
CHEZO4	11,083,010,722	73,397,422	7.0	6.54	0.5	9,502,438,921	63,561,063	85.74	86.60	17	565,923,029
Average	11,389,987,456	75,430,381	7.5	6.63	0.05	9,749,908,834	65,218,503	85.60	86.46	17	553,135,905

*K-mer: Assembly k-mer length.

Table 2. Genome contig statistics and sequence quality assessment

	No. Contigs	Residues (bp)	Avg. length (bp)	Min. length (bp)	Max. length (bp)	N50 (bp)[1]	N (%)	GC (%)[2]
Genome	188,476	536,816,551	2,848.20	195	101,290	8,188	1.71	42.46

[1]N50, Contig and scaffold length corresponding to 50% of the total number of total nucleotide sequences produced when accumulating in descending order for the generated contigs and scaffolds. [2]GC (%), GC content.

Table 3. Consensus gene models were predicted using the Zebrafish training HMM

	Count (#)
Gene	72,802
CDS (Exon)	277,523
Intron	204,721
CDS/gene	3.81
	Average length (bp)
Gene	3,051.92
CDS	186.27
Intron	832.8
Max gene	67,116
Min gene	201
	Coverage (%)
Gene	41.39
CDS	9.63
Intron	31.76

HMM, Hidden markov model; CDS, Coding sequence.

Table 4. Statistics of microsatellite markers identified in genome contigs

Searching items	# Counts
No. genome contig	188,476 (100%)
No. identified SSRs	48,789
No. genome contig containing SSRs	39,127 (5.7%)
No. genome contig containing more than 2 SSR	9,713
No. Di-nucleotide SSRs	29,467
No. Tri-nucleotide SSRs	10,098
No. Tetra-nucleotide SSRs	3,652
No. Penta-nucleotide SSRs	2,799
No. Hexa-nucleotide SSRs	5,126
No. Hepta-nucleotide SSRs	1,249

SSRs, Simple sequence repeats.

방법을 이용한 벤자리 연구(Dong et al., 2023) 결과와 비교해 보았을 때, 전체적인 분포 양상은 유사하지만 아홉동가리에서 탐색된 SSR 마커의 개수와 프라이머 쌍의 비율이 더 높은 것으로 나타났다.

Microsatellite 마커의 특성 분석 및 선별

유전자 microsatellite 마커의 특성 분석 및 선별은 두 단계로 나누어 진행되었다. 첫 번째 단계에서는 마커의 유전체 내 위치 정보를 기준으로 마커를 선별하였다. 유전자 영역 내에서 안정적으로 유지되는 마커를 확보하기 위해, 3'-UTR 영역을 중심으로 genic 영역을 우선적으로 선발하였다. 분석 결과, CDS 영역

Table 5. Distribution of microsatellite motif lengths by type

Motif length	Di	Tri	Tetra	Penta	Hexa	Hepta
Total	29,467	10,098	3,652	2,799	5,126	1,249
3	-	-	-	-	4,982	1,233
4	-	-	-	2,679	138	15
5	-	-	3,347	117	5	1
6	-	7,211	277	3	-	-
7	-	2,548	21	-	-	-
8	-	272	5	-	-	-
9	15,282	41	1	-	-	-
10	9,692	16	1	-	-	-
11	3,806	6	-	-	1	-
12	444	2	-	-	-	-
13	122	1	-	-	-	-
14	65	-	-	-	-	-
15	35	-	-	-	-	-
16	9	1	-	-	-	-
17	10	-	-	-	-	-
18	1	-	-	-	-	-
19	-	-	-	-	-	-
20	-	-	-	-	-	-
>=21	1	-	-	-	-	-

에는 3,017개의 microsatellite 마커가 포함되었고, 3'-UTR에는 4,727개, 5'-UTR에는 4,790개, intergenic region에는 37,125개가 분포하였다(Table 6).

두 번째 단계에서는 PCR primer를 디자인하고, 이를 기반으로 후보 마커를 선별하였다. 첫 번째 단계에서 선별된 48,798개의 microsatellite 마커 후보군을 대상으로 서로 중복되는 SSR motif를 배제하였으며, 추가적으로 반복 수 길이가 10 bp 이상, 증폭산물 크기가 100–400 bp, primer 길이가 20–24 bp, 적정 온도은 48–62°C, GC 함량이 40–60%인 조건을 만족하는 microsatellite 마커 후보군 34,117개를 선별하였다.

선별된 34,117개 마커 중 primer 서열 내 low complexity 서열(단순 반복 서열 포함)의 존재 여부를 확인하여 1차로 29,861개 마커를 선별하였다. 이후, PCR 반응 효율성을 검토하기 위해 primer 서열의 유전체 내 유일성을 평가하였다. 이를 위해 아홉동가리 대표 유전체 서열에 디자인된 primer 서열을 reference 매핑하여, 각 microsatellite 마커 primer가 정확하게 목표 위치에 매핑 되는지를 확인하였다. 그 결과, 24,480개 마커를 2차로 선별하였다.

다형성 정보를 확보하기 위해 BWA (v0.7.12; Broad Institute) (Li and Durbin, 2009)의 MEM 알고리즘을 사용하여 개체별 시퀀싱 데이터를 매핑하고, SAMtools (v1.10; Broad Institute) 및 BCFtools (v1.9; Wellcome Sanger Institute)를 활용하여 변

이 분석을 수행하였다. 이를 통해 SSR 마커 영역 내 InDel을 포함한 다형성 여부를 확인하였으며, 최종적으로 3,648개 마커를 선발하였다. 이들 마커는 유전체 내 위치에 따라 3'-UTR 344개, 5'-UTR 298개, CDS 2,021개, intergenic region 2,817개로 분류되었다(Table 6).

이 중에서는 특히 3'-UTR 및 5'-UTR 영역과 같은 비암호화(non-coding) 유전자 영역에서 프라이머를 설계하였다. 이러한 UTR 부위는 단백질로 번역되지 않으나, mRNA의 안정성, 번역 효율, 그리고 유전자 발현 조절에 중요한 기능을 하는 영역으로 알려져 있어, 분자 마커로서의 활용 가능성이 높다고 판단하였다.

최종 선별된 642개 마커 중 반복 염기의 종류, primer 크기, 증폭산물의 위치 등을 고려하여 1차로 146개의 microsatellite 마커를 선별하였다. 이후, 1차로 선별된 146개 후보 서열을 대상으로 PCR 증폭 여부 및 증폭산물의 크기를 확인하여 2차 선별을 진행하였다. 2차로 선별된 56개의 microsatellite 마커에 형광물질(dye)을 합성한 후, 유전자형(genotyping) 분석을 수행하여 NA와 크기 등을 평가하였다. 이를 통해 최종적으로 20개의 microsatellite 마커를 선별하였다(Table 7).

Microsatellite 마커의 효율성 평가

최종 선발된 20개의 microsatellite 마커의 효율성을 검증하기 위해 제주도 내 4개 지역(신창리 2024년, n=5; 북촌리 2024년, n=16; 사계리 2024년, n=27; 신흥리 2024년, n=2)에서 정치망으로 포획된 자연산 아홉동가리 50개체를 대상으로 유전학적 특성 분석을 수행하였다.

Table 6. Microsatellite locus distribution and InDel polymorphism across genomic regions

Locus	Di	Tri	Tetra	Penta	Hexa	Hepta	Total	InDel polymorphism
3'-UTR	2,652	931	333	242	461	108	4,727	344
5'-UTR	2,650	927	337	275	462	139	4,790	298
CDS	105	1,981	125	71	691	44	3,017	2,021
Intergenic	22,137	6,197	2,609	2,014	3,315	853	37,125	2,817

Table 7. Characteristics of 20 microsatellite loci developed in *Goniistius zonatus*

No.	Name	Sequence (5'→3')		Repeat	Product size	Temperature (°C)	SSR type	Dye
1	9DGR-6	F-AGTGTCCAGTTGTCATATTAGTGCT	R-ACCTTTTTCCCCCATTTGTCCT	(TAT)6	119	60.09		
2	9DGR-8	F-GGCTGAGAACCAACACTGAGAT	R-AGCTCCACTGTTACACCTGTTT	(CA)9	111	59.83		
3	9DGR-17	F-GCTGACATCACTTTTTCAGCCA	R-CAACTGACGCATGACTACACAA	(CA)10	214	59.20		
4	9DGR-20	F-GTGAGTCAAAGCCCAAACTGTG	R-CTGGTGTATTGGCTGTGCTTTC	(GT)10	217	60.10		
5	9DGR-32	F-GCATAAAGGCGTTTTCTCCCTC	R-GGAGCTGATAAGTGATGCAGGT	(CA)10	308	60.16	3'-UTR	
6	9DGR-40	F-CCAGATACGCCCACTTACTCTC	R-CACCGGAGCAATCTATCGTCTA	(TG)9	284	59.45		
7	9DGR-41	F-ATCAAAGTTCACAGACGAGCAC	R-GCAGCCCTTAACAGCAATAACA	(TTG)6	319	59.51		
8	9DGR-59	F-ACCTCATGTTCTTTGCCCTTCT	R-GGGCAAAGTGGACTGAGAAAAC	(AC)9	395	59.97		
9	9DGR-60	F-CACATCCAAGAGACAGAGGCTT	R-CACTACAAGAAAGCCCCCATCT	(TG)10	398	60.03		
10	9DGR-65	F-GTTTGCTTGTGAAGACGTGCTA	R-GTGCAAAATAACCAAACTGACAGTG	(TG)9	406	59.77		6-FAM
11	9DGR-84	F-ACTCAGACTCTTACACCGGGTA	R-CTTGTTAGGGCAGCCACTTTTG	(AC)10	119	60.29		
12	9DGR-86	F-CACGTAGGCACCATCTCACAC	R-TCAAAGAGCACATTCAGCAAGG	(AG)9	119	59.44		
13	9DGR-96	F-TCACCTCCATCCTGTCCATTTG	R-AACTCTCCCCTACTTCCCTACC	(GT)9	217	60.02		
14	9DGR-98	F-GCTCAACTCCAACAGAACCATC	R-AATCAGCTGCAGGAAAGTTTGG	(AG)10	220	59.70		
15	9DGR-107	F-GGGGTCTTGGAATGAGGTGAAT	R-ACGGGGTCTCACAGTGGATATA	(CTT)6	217	60.09	5'-UTR	
16	9DGR-110	F-AGGGAAAGGCAAACAGATCGAT	R-GCGACAGAGTTTATTGCAGGTG	(TG)10	308	60.16		
17	9DGR-117	F-CTTGGCAGTCCTTTCATCCTCT	R-CTGTCTCTTGTCCTCTCCCTCT	(AC)9	290	60.29		
18	9DGR-127	F-CGAGTCTGCGTTTGAATTCCTG	R-AACCAGTAATCCAGGCCACAAA	(GT)10	408	60.16		
19	9DGR-132	F-CGACCTCTTTTATTGCGCTTGT	R-GCTCTGTACTCTGTGATGTGCA	(GCT)6	405	60.35		
20	9DGR-134	F-TAATTAGCCGCACCTGTAGCTC	R-TCTCATAATTGCCCTGTCTCGG	(CA)10	403	59.90		

Table 8. Genetic variabilities at 20 microsatellite markers in the population of *Goniistius zonatus*

	Locus	N_A	A_R	R	H_O	H_E	PIC	F_{IS}	HW
Jeju (50)	9DGR-6	9	9	117-141	0.760	0.816	0.785	0.069	0.194
	9DGR-8	14	14	102-130	0.920	0.854	0.827	-0.078	0.133
	9DGR-17	11	11	246-268	0.820	0.801	0.770	-0.023	0.858
	9DGR-20	7	7	216-228	0.780	0.785	0.746	0.006	0.009
	9DGR-32	13	13	282-320	0.820	0.856	0.831	0.043	0.573
	9DGR-40	6	6	266-286	0.620	0.627	0.546	0.011	0.291
	9DGR-41	9	9	303-333	0.660	0.749	0.703	0.120	0.036
	9DGR-59	10	10	395-415	0.840	0.821	0.789	-0.023	0.836
	9DGR-60	5	5	384-402	0.700	0.715	0.653	0.021	0.776
	9DGR-65	12	12	404-428	0.840	0.875	0.852	0.040	0.479
	9DGR-84	7	7	107-119	0.720	0.725	0.672	0.007	0.683
	9DGR-86	11	11	115-137	0.760	0.836	0.811	0.092	0.152
	9DGR-96	8	8	213-233	0.640	0.713	0.658	0.104	0.388
	9DGR-98	7	7	206-224	0.780	0.699	0.640	-0.117	0.254
	9DGR-107	5	5	213-228	0.640	0.682	0.607	0.062	0.054
	9DGR-110	6	6	308-318	0.700	0.704	0.651	0.006	0.616
	9DGR-117	6	6	284-294	0.720	0.762	0.715	0.055	0.905
	9DGR-127	15	15	405-439	0.940	0.882	0.861	-0.067	0.320
	9DGR-132	6	6	398-413	0.660	0.700	0.650	0.058	0.149
	9DGR-134	9	9	397-419	0.780	0.725	0.680	-0.076	0.522
	Mean all loci	9	9	-	0.755	0.766	0.722	0.015	0.411

N_A, Number of alleles per locus; A_R, Allelic richness; H_O, Observed heterozygosity; H_E, Expected heterozygosity; PIC, Polymorphic information content; F_{IS}, Inbreeding coefficient; HW, Hardy-Weinberg. *Not in conformity with Hardy-Weinberg Equilibrium ($P<0.003$, Bonferroni-corrected value).

본 연구에서 사용된 microsatellite 마커들의 다형성 정보량을 나타내는 PIC 값은 0.546에서 0.861 범위로 나타났으며, 평균 0.722로 계산되었다(Table 8). Botstein et al. (1980)에 따르면, PIC 값이 0.5 이상일 경우 개체 식별력이 충분한 효율적인 마커로 평가되므로, 본 연구에서 개발된 마커들은 아홉동가리 집단의 유전적 다양성을 분석하는데 적합한 개체 식별력을 제공하는 것으로 판단된다. 또한, 마커의 신뢰성을 평가하기 위해 null 대립유전자 분석을 수행한 결과, null allele이나 유전자형 판독 오류(scoring errors), 대형 대립유전자 소실(large allele dropout)이 발견되지 않았다. 이러한 결과는 본 연구에서 사용한 마커들의 유전자형 빈도가 안정적으로 유지되고 있음을 의미하며, 아홉동가리 집단의 유전적 다양성을 분석하는데 있어 신뢰할 수 있는 마커임을 확인할 수 있었다.

아홉동가리 집단 내에서 분석된 NA는 평균 9개로 나타났으며, 각 집단에서 HO는 평균 0.755, HE는 평균 0.766으로 확인되었다(Table 8). 이러한 결과는 해산 어류의 평균 HE 값인 0.79 (DeWoody and Avise, 2000)와 유사한 수준을 보이며, 본 연구에 사용된 아홉동가리 집단의 유전적 다양성이 비교적 안정적으로 유지되고 있음을 시사한다. 또한, 집단 내 무작위 교배 여부를 평가하기 위해 내교배 계수를 분석한 결과, 평균 0.015 (-0.117–0.120)로 나타났다(Table 8). 이는 분석에 사용된 아홉동가리 집단 내에서 무작위 교배가 이루어지고 있으며, 근친교배로 인한 유전적 다양성 감소의 위험이 낮다는 것을 의미한다 (Hedrick, 2000).

본 연구는 NGS 기법을 활용하여 아홉동가리의 새로운 유전체 정보를 확보하고, 이를 바탕으로 다형성과 재현성이 높은 20개의 microsatellite 마커를 개발하였다. 새로 개발된 microsatellite 마커는 아홉동가리 집단 간 유전학적 분석 및 유전자 다양성 평가에 신뢰할 수 있는 도구로 활용될 수 있으며, 이를 통해 아홉동가리 집단의 지속 가능한 관리 및 종 보존을 위한 중요한 기초 자료로 활용될 것으로 기대된다.

사 사

논문은 2025년도 국립수산과학원 수산과학연구사업(R2025016, 해수 수산생물 종 보존 연구)의 지원으로 수행되었습니다.

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Botstein D, White RL, Skolnick M and Davis RW. 1980. Construction of a genetic linkage map in manusing restriction fragment length polymorphisms. Am JHum Genet 32, 314-331.

Burridge CP and Smolenski AJ. 2004. Molecular phylogeny of the Cheilodactylidae and Latridae (Perciformes: Cirrhitoidea) with notes on taxonomy and biogeography. Mol Phylogenet Evol 30, 118-127. https://doi.org/10.1016/s1055-7903(03)00157-x.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM and Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008.

DeWoody JA and Avise JC. 2000. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. J Fish Biol 56, 461-473. https://doi.org/10.1111/j.1095-8649.2000.tb00748.x.

Dong CM, Lee HM, Lee MN, Noh ES, Nam BH, Kim YO and Kim EM. 2022. Development of new microsatellite DNA markers using next-generation sequencing in pacific oyster *Crassostrea gigas*. Korean J Malacol 38, 129-138. https://doi.org/10.9710/kjm.2022.38.3.129.

Dong CM, Lee MN, Kim EM, Park JY, Kim GD and Noh JK. 2020. Development and genetic diversity analysis of microsatellite markers using next-generation sequencing in *Seriola quinqueradiata*. J Life Sci 30, 291-297. https:// doi.org/10.5352/JLS.2020.30.3.291.

Dong CM, Lee MN, Noh JK, Park JW, Kim YO and Kim EM. 2023. Identification and characterization of polymorphic microsatellite DNA markers using next-generation sequencing in *Parapristipoma trilineatum*. J Life Sci 33, 623-631. https://doi.org/10.5352/JLS.2023.33.8.623.

Excoffier L, Laval G and Schneider S. 2005. Arlequin (ver.3.0): An integrated software package for population genetics data analysis. Evol Bioinform 1, 47-50. https://doi.org/10.1177/117693430500100003.

Gardner MG, Fitch AJ, Bertozzi T and Lowe AJ. 2011. Rise of the machines-recommendations for ecologists when using next generation sequencing for microsatellite development. Mol Ecol Resour 11, 1093-1101. https://doi.org/10.1111/j.1755-0998.2011.03037.x.

Han IS, Lee JS, Kim CS and Yang JY. 2023. Impacts, projections and assessments related to climate change in ocean and fisheries. J Climate Change Res 14, 965-972. https://doi.org/10.15531/KSCCR.2023.14.6.965.

Hedrick PW. 2000. Genetics of Population. Jones and Bartlett Publishers, Burlington, MA, U.S.A., 34-47.

Hubisz MJ, Falush D, Stephens M and Pritchard JK. 2009. Inference weak population structure with the assistance of sample group information. Mol Ecol Resour 9, 1322-1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x.

Kalinowski ST, Taper ML and Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16, 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x.

Langmead B and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359. https://doi.org/10.1038/nmeth.1923.

Lee TW, Kim SW, Kim JS, Chi WJ, Bang WY, Kim JH, Yang CW, Bang KH and Gal SW. 2022. Analysis of microsatellite patterns in the genome of red sea cucumber. J Life Sci 32, 690-697. https://doi.org/10.5352/JLS.2022.32.9.690.

Li H and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324.

Mardis ER. 2008. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9, 387-402. https://doi.org/10.1146/annurev.genom.9.081307.164359.

Matsumoto K. 2001. Overlapping territory of a benthophagous fish, *Goniistius zonatus* (Teleostei: Cheilodactylidae). Ecol Res 16, 715-726. https://doi.org/10.1046/j.1440-1703.2001.00428.x.

Matsumoto K and Kohda M. 2002. The effect of feeding habitats on dietary shifts during growth in a benthophagous suction-feeding fish. Zool Sci 19, 709-714. https://doi.org/10.2108/zsj.19.709.

Nakabo T. 2002. Fishes of Japan with pictorial keys to the species. 2nd ed. Tokai University Press, Tokyo, Japan, 1749.

NIFS (National Institute of Fisheries Science). 2022. Annual Report for Climate Change Trends in Fisheries, 2022. NIFS, Busan, Korea, 1-46.

NIFS (National Institute of Fisheries Science). 2024. Annual Report for Climate Change Trends in Fisheries, 2024. NIFS, Busan, Korea, 1-87.

Oosterhout CV, Hutchinson WF, Wills DPM and Shipley P. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4, 535-538. https://doi.org/10.1111/j.1471-8286.2004.00684.x.

Paetkau D. 1999. Microsatellites obtained using strand extension: An enrichment protocol. BioTechniques 26, 690-697.

https://doi.org/10.2144/99264st05.
Park CJ, Nam WS, Lee MS, Kang JY and Kim KK. 2014. Microsatellite multiplex PCR method for selective breeding studies in Pacific abalone (*Haliotis discus hannai*). Korean J Malacol 30, 383-390. https://doi.org/10.9710/kjm.2014.30.4.383.
Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8, 103-106. https://doi.org/10.1111/j.1471-8286.2007.01931.x.
Stanke M and Morgenstern B. 2005. Augustus: A web server for gene finding in eukaryotes. Nucleic Acids Res 33, W465-W467. https://doi.org/10.1093/nar/gki458.
Stanke M, Keller O, Gunduz I, Hayes A, Waack S and Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res 34, W435-W439. https://doi.org/10.1093/nar/gkl200.
The UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res 47, D506-D515. https://doi.org/10.1093/nar/gky1049.
Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res 40, e115. https://doi.org/10.1093/nar/gks596.
Zhong S, Su Y, Mao Y, Liu M and Wang J. 2015. Development of 20 microsatellite markers for kuruma shrimp (*Marsupenaeus japonicus*) using illumina sequencing. Conserv Genet Resour 7, 717-719. https://doi.org/10.1007/s12686-015-0442-6.