Thesis for the Degree for Master of Science

# Development of microsatellite markers to identify geographical origin between Korean and Japanese yellowtail, *Seriola quinqueradiata*

by

Song Yi Baek

Department of Microbiology

The Graduate School of Industry

Pukyong National University

August 2022

# Development of microsatellite markers to identify geographical origin between Korean and Japanese yellowtail, *Seriola quinqueradiata* (한국산과 일본산 방어의 원산지 판별을 위한 microsatellite marker 개발)

Advisor: Prof. Gun Do Kim

by

Song Yi Baek

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Science

in Department of Microbiology, The Graduate School of Industry,

Pukyoung National University

August 2022

Development of microsatellite markers to identify geographical origin between Korean and Japanese yellowtail, *Seriola quinqueradiata*

A dissertation

by

Song Yi Baek

Approved by :

(Chairman) Young Jae Jeon

(Member) Kyoung-Ho Kim

(Member) Gun-Do Kim

August 26, 2022

# Contents

Development of microsatellite markers to identify geographical origin between Korean and Japanese yellowtail, *Seriola quinqueradiata*

Song Yi Baek

Department of Microbiology, The Graduate School of Industry
Pukyong National University

## Abstract

This study is about the development of a microsatellite marker for the identification of the origin of domestic and Japanese yellowtail distributed in Korea and the development of a method for determining the origin using the same. Yellowtail accounts for a high proportion of domestic fish production. According to data from the National Statistical Office, yellowtail imports increased from 298 tons in 2015 to 2,240 tons in 2019, an 8-fold increase, and most of the domestic yellowtail imports depend on imports from Japan. While concerns about Japanese fish and the demand for safe aquatic products are increasing, the number of violations of the labeling of origin of aquatic products has increased. The microsatellite marker used in this study is a repeating sequence scattered throughout the genome and is mainly used for

comparison analysis of mutations between groups. Therefore, this study suggests the possibility of using microsatellite markers in the import and quarantine process to prevent illegal distribution by developing a method for determining the origin of domestic and Japanese yellowtail products. Genomic DNA was extracted from the fins and muscles by securing a total of 180 yellowtail individuals from 90 domestic and 90 Japanese yellowtail specimens each three times. Whole genomic DNA was treated with restriction enzymes to prepare a GBS library and NGS data was obtained. Sixty nine candidate groups were selected based on the locus that was judged to be able to differentiate between domestic and Japanese products. After PCR, 8 markers were selected through primary and secondary screening to confirm size variation. For the 8 selected markers, the size of the allele was calculated through fragment analysis, and the size of the DNA fragment was determined by performing genotyping. Fragment analysis raw data for 8 markers were analyzed with Cervus 3.0.7, Micro-checker 2.2.3, Arlequin 3.5.2.2, FSTAT 2.9.4, and GenAlEx statistical analysis program.

We analyzed the genotype for each group of origin of yellowtail using a combination of newly developed microsatellite markers. It is possible to effectively identify a total of 8 markers and their origin. Among them, based on the FST value, 5 (SqMS32, SqMS33, SqMS36, SqMS43, SqMS45) or 3 (SqMS32, SqMS36, SqMS43) microsatellite marker combinations were used to confirm that efficient origin discrimination was possible. It is expected to be able to effectively discriminate the origin of domestic and Japanese yellowtail by performing PCoA and genotype likelihood matrix analysis through the obtained genotype information.

When 8 microsatellite markers were used, it was observed that a colony capable of distinguishing between domestic and Japanese products was formed. Similarly, when using a combination of five (SqMS32, SqMS33, SqMS36, SqMS43, SqMS45) selected by statistical analysis, it was possible to distinguish between domestic and Japanese products. Additional monitoring of yellowtail will allow us to pinpoint the three most effective combinations. Based on the results of this study, it was possible to confirm the possibility of grafting it into the development of on-site diagnosis.

# 1. Introduction

## 1.1 Background

### 1.1.1 Current status of seafood imports and consumption in Korea

Owing to increasing demand and improving price competitiveness, seafood imports are increasing every year. According to the National Statistical Portal of the National Statistical Office, fishery production in the coastal waters in 2019 was only 91,429 tons, down 98,276 tons from 2018. The production shortage is addressed through imports. A recent analysis of aquatic product import trends by the Korea Customs Service over the past 10 years has revealed that seafood imports have doubled. Skate consumption has also increased since cheap Chilean products appeared 10 years ago: and recently, imports of Argentinean skates are increasing. In the case of cutlassfish, the number of import destinations increased from 18 to 23 countries [1].

As the consumption of overseas seafood has increased, violations of country of origin labeling have also increasing. According to the 2020 State Audit of the Ministry of Oceans and Fisheries, “Status of Violation of Origin Labeling for Seafood,” 3,926 companies violated labeling of origin requirements for aquatic products over the past 5 years (2015-2019), with a scale of 15,840,000,000 won [2].

### 1.1.2 Importance of identifying the origin of the yellowtail (*Seriola quinqueradiata)*

A number of methods have been developed to determine the origin and species of agricultural products, livestock products, and marine products. For

example, in 2015, a study focused on differentiating between species was performed to address the issue of *Cynanchum auriculatum* being sold as a high-quality herbal medicine, *Cynanchum wilfordii* [6]. In the case of livestock products, methods to distinguish among Korean beef, black Korean beef, which cannot be identified by the naked eye based on slaughtered meat, were evaluated [8].

*Seriola quinqueradiata* is a temperate fish belonging to the family *Carangidae* and order *Carangiformes*. It is distributed over the east coast of Korea, Japan, Taiwan, and Hawaii. Yellowtail is an economically important fish consumed as premium sashimi in the winter, and Japan, China, Australia, and Taiwan are the main producers. In particular, Japan produces produces the *S. quinqueradiata*. In Japan, since the 1960s, wild-caught fry have been cultivated, in addition to artificial seedling production.

In Korea, *S. quinqueradiata* has been produced since the 1980s. In 2019, 15,928 tons (69,400,000,000 won) of yellowtail were produced by fisheries, and aquaculture production was found to be 650 tons (7,800,000,000 won) according to the Statistics Office of the Republic of Korea. Yellowtail imports in Korea have increased about tenfold from approximately 298 tons (about 2,183,000,000 won) in 2015 to approximately 2,240 tons (about 20,278,000,000 won) in 2019, and most of the imported yellowtail distributed in Korea is imported from Japan. However, it is difficult to distinguish between Korean and Japanese yellowtails by visual inspection.

After the Fukushima nuclear disaster caused by the Great East Japan Earthquake in March 2011, Korea's imports of Japanese seafood declined sharply. The discharge of radioactively contaminated water in Japane

heightened opposition towards selling Japanese seafood products as domestic products. Imported Japanese yellowtail in Korea is entirely farmed and is not inferior to domestic products in terms of quality. However, fear of the possibility of radioactive contamination of Japanese aquatic products dampened consumption by Korean consumers. Furthermore, the Korean government imposed an import ban on 50 seafood products produced in eight prefectures around Fukushima Prefecture in 2012 and an import ban in September 2013. The expansion of the target to all aquatic products in the eight prefectures accelerated the decline in Japanese seafood imports [4].

The Korean yellowtail spawning season is February-March in the southern part of the East China Sea and in March-April in the South Sea of Korea. The spawning sites are in the area around the southern continental shelf of the East China Sea, southwest of Kyushu, Japan, and south of Jeju Island, Korea. Young fry that hatched in spring are collected and used as seedlings. They are collected and caught in the coastal area of the southern coast from the coastal waters of Jeju Island from May to June [5].

### 1.1.3 Benefits of microsatellite markers

Microsatellite markers, restriction fragment length polymorphisms (RFLP), allozymes, mitochondrial DNA, and random amplified polymorphic DNA are a representative markers used for breed identification or population genetic analyses based on DNA fragments (RAPD), amplified fragment length polymorphism (AFLP), and single nucleotide polymorphisms (SNPs) [3,13].

Microsatellite markers have been widely used for genetic analysis since the 1990s. They are short DNA sequences of 1 to 6 short base pairs repeated

several times, also called simple sequence repeats (SSR) [14].

Because microsatellite markers follow Mendel's laws of inheritance and have high polymorphism, they are used for pedigree analyses, populaion genetic analyses, and paternity confirmation in various organisms, including mammals and plants. It has been used as a genotyping analysis technology for individual identification, and are a useful molecular marker for breed identification and origin identification in animals, such as cattle, pigs, and salmon since [9,10].

### 1.2 Purpose

This study proposes a method for determining the origin of *S. quinqueradiata*, a species with frequent violations of origin labeling among Korean and Japanese seafood.

Biomarkers for identification of origin and species were developed using a bioinformatics approach based on high-throughput genome data for aquatic organisms.

We establish an optimal analysis method marker development for origin discrimination based on genomic data and identify candidate markers for determining the origin of *S. quinqueradiata.*


**Figure 1. Overview the main habitat of *S. quinqueradiata* in coastal areas.**

The species is distributed along the coasts of Korea and Japan. (OBIS, Ocean Biodiversity Information System)

## 2. Materials and methods

### 2.1 Sample collection

Korean and Japanese *S. quinqueradiata* samples were collected from three areas each. In total, 90 Korean yellowtails were purchased from fish markets in Busan Jagalchi, Tongyeong, and Pohang, and 90 Japanese yellowtails were purchased through a seafood importer. DNA samples for the microsatellite marker analysis were extracted from fin or muscle tissues. Tissue samples was stored at −80℃ until DNA extraction. Sample information for Korean and Japanese *S. quinqueradiata* is shown in Table 1.

**Table 1. *S. quinqueradiata* sample information**

| Origin | Group | Number of individuals | Sampling date |
|---|---|---|---|
| Korea | KR_A | 30 | 2020.06.23. |
| | KR_B | 30 | 2020.09.22 |
| | KR_C | 30 | 2020.09.22 |
| Japan | JP_A | 30 | 2020.10.21 |
| | JP_B | 30 | 2020.10.21 |
| | JP_C | 30 | 2020.11.06 |

### 2.2 Genomic DNA extraction

The AccuPrep Genomic DNA Extraction Kit (Bioneer, Daejeon, Korea) was used to extract and purify genomic DNA from the cut fin or muscle tissue (approximately 0.5 cm). Samples were diluted before use and adjusted to a concentration of 10 ng/$\mu$l.

### 2.3 Genotyping-by-sequencing (GBS) analysis

To construct a genotyping-by-sequencing (GBS) library (96 samples/ 1 set), 120 samples of Korean and Japanese *S. quinqueradiata* were obtained (i.e., 60 samples from Korea and 60 samples from Japan). Total genomic DNA was extracted from each individual using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). Total gDNA of 48 Korean and 48 Japanese specimens with distinct major genomic DNA bands were selected for treatment with restriction enzymes for GBS library construction.

To construct GBS libraries, extracted *S. quinqueradiata* gDNA was treated with the restriction enzymes Pst Ⅰ and Msp Ⅰ. After confirming complete digestion, the DNA from each sample was ligated with an adapter sequence suitable for these restriction enzymes. The barcode sequence (4-6 bp) for individual identification followed the terminal end of the adapter sequence. The primer sequences for PCR (usually P5, P7 promoters) were configured so that the barcode sequence was followed by PCR primers. Finally, a library for *S. quinqueradiata* NGS was constructed.

The GBS library was prepared, and the adapter sequence was removed by trimming. After individual identification and individual sequence separation by demultiplexing, high-quality sequences were isolated by removing individual

barcode sequences and enzyme sequences and checking sequence quality. In addition, by mapping, de novo clustering, and mapping to the standard sequence, microsatellite markers for each individual that differed between Korean and Japanese yellowtails were selected to search for candidate sequences for the development of effective markers to differentiate between fish of different origins. The GBS analysis was performed using TASSEL (Trait analysis by association, evolution and linkage), and the workflow from total DNA extraction to GBS analysis is shown in Figure 2.

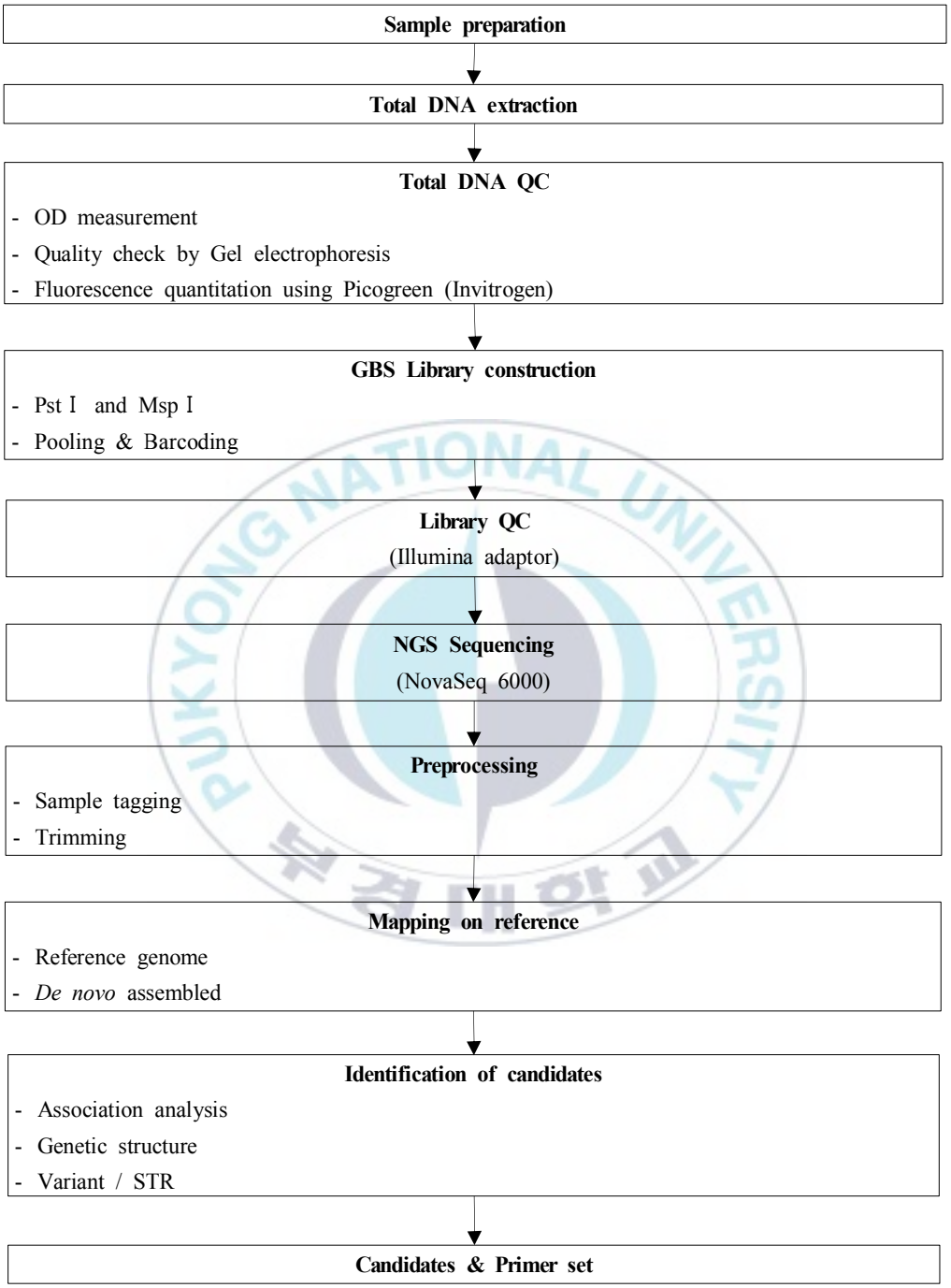


**Figure 2. Workflow from total DNA extraction to GBS analysis [11]**

2.3.1 Obtaining large-scale genomic data by GBS method

After extracting total DNA from the blood samples of 120 Korean and Japanese yellowtail individuals, among them, GBS libraries were constructed from the DNA of 96 samples satisfied with the total DNA quality. Using the raw data produced by performing NGS, a bioinformatic analysis was performed and results were analyzed.

2.3.2 Identification of candidate microsatellite markers by bioinformatic analysis

To select candidate microsatellite markers for *S. quinqueradiata* GBS data, overall GBS analysis were analyzed using TASSEL (Trait Analysis by aSSociation, Evolution and Linkage). A variation search and association analysis were performed. In addition, the genetic structure was confirmed using the structure program, and an in-house script (R-package) was used for clustering and the generation of images.

2.3.3 Development of candidate microsatellite markers for origin identification

To search for microsatellite markers from Korean and Japanese *S. quinqueradiata*, clean reads were obtained from NGS raw data. After searching for each microsatellite marker from the forward and reverse reads, sequences with the same flanking sequence for the microsatellite repeat DNA region (SSR) but differences in the number of repeat sequences were selected.

The Perl script known as MISA (Microsatellite identification tool) was used to detect the SSR loci. A microsatellite DNA region was considered to contain 2-6 nucleotide motifs with a minimum of 4-6 repeat sequences, and mononucleotide repeats sequences were excluded from the analysis.

Microsatellite markers were searched using the NGS reads. Primer sequences were designed for the candidates using and Primer 3.0 with the following settings: primer length to 18-28 bp, PCR product size 100-280 bp, melting temperature corresponding to an optimum annealing temperature of 57-63℃, optimal GC content 50%, and within 40-60%.

**2.4 Primer screening by PCR**

To identify new microsatellite marker candidates by GBS library construction, the GCA_002217815.1 assembly in NCBI was used as a reference. In total, 2826 markers were obtained from the database obtained from GBS reads, and 69 candidate microsatellite markers were selected for discrimination between Korean and Japanese *S. quinqueradiata*. To confirm the suitability of these 69 microsatellite markers, the primer efficiency was confirmed by PCR using the *S. quinqueradiata* genomic DNA from three Korean and three Japanese samples as a template.

PCR was performed to identify effective primers for microsatellite marker candidates with 1 ㎕ of template genomic DNA (10 ng/㎕), 1 ㎕ of forward primer (10 pmol), 1 ㎕ of reverse primer (10 pmol), and 10 ㎕ of GenetBio Prime Taq Premix (2X). The final volume was adjusted to 20 ㎕ by the addition of distilled water. The SimpliAmp™ Thermal Cycler (Applied Biosystems, Waltham, MA, USA) was used. Reaction condition were as follows: 95℃ for 5 minutes, followed by 30 seconds at 94℃, 30 seconds at 60℃, and 1 minute at 72℃, repeated 40 times, and extension reaction at 72℃ for 7 minutes. Then, 2 ㎕ of the PCR product was confirmed by electrophoresis on a 2% agarose gel. Eight microsatellite loci for which the

size difference of PCR products was confirmed by electrophoresis were selected for a fragment analysis. The composition of the PCR reaction solution is shown in Table 2. The reaction conditions are shown in Table 3.

**Table 2. PCR solution composition for microsatellite marker identification**

| Components | Concentration | Volume( $\mu$ L) |
|---|---|---|
| Prime Taq Premix (2X) | - | 10 |
| Forward primer | 10 pmol | 1 |
| Reverse primer | 10 pmol | 1 |
| Template DNA | 1 ~ 10 ng | 1 |
| Distilled water | up to 20 $\mu$ L | 7 |
| Total | | 20 |

**Table 3. PCR conditions for microsatellite marker identification**

| Predenaturation (Temp./TIme) | Denaturation (Temp,/Time) | Annealing (Temp./Time) | Extension (Temp./Time) | Final extension (Temp./Time) |
|---|---|---|---|---|
| 94℃/5min | 94℃/30sec | 60℃/30sec | 72℃/1min | 72℃/7min |
| | 40 Cycles | | | |

For the 45 primer sets, additional PCRs were performed for secondary selection. The initial candidates, which showed clear size variation by electrophoresis, were preferentially identified. Information on the eight primer sets is provided in Table 4.

**Table 4. Information on microsatellite marker candidates**

| No. | Locus | ID | Repeat motif | Sequences (5'→3') | Size (bp) | Primer sequence (5'→3') | |
|---|---|---|---|---|---|---|---|
| 1 | *SqMS17* | seq_35676434 | $(AT)_6$ | TGAACAGACCACTGTTCACGTCTGATATATATATATACACACACACATATATCCTCAGCCG | 129-143 | F | TGGCCCTGCTTCACATTGAT |
| | | | | | | R | ACGGCTGAGGATATATGTGTGT |
| 2 | *SqMS24* | seq_36178528 | $(GT)_8$ | TGTGTTTGAATTGGTTGTTGTAGGAGTGTGTGTGTGTGTGTTTGTATGTGTGGGATGGGCTCGACA | 108-142 | F | ACGTCACAAACAGCTCACCT |
| | | | | | | R | TCGAGCCCATCCCACACAT |
| 3 | *SqMS32* | seq_11310468 | $(GAT)_6$ | ACAGGCCCTGAAAGGTACGGTAACAGATGATGATGATGATGATGTCACGGGGCCTAATCGGGCTTTCT | 111-133 | F | GCAGGAGGCAAAGAGAGGAG |
| | | | | | | R | TAAGATCTGGCTTGCAGCGG |
| 4 | *SqMS33* | seq_48953288 | $(ATG)_8$ | AGCTGTTGCTCCTCTCATAACCGACATGATGATGATGATGATGATGATACTGTCTGATGGATGCAGAGTTT | 110-132 | F | CACCAGTCCAGACAGAACCC |
| | | | | | | R | TGAAACTCTGCATCCATCAGACA |
| 5 | *SqMS35* | seq_19509187 | $(TC)_{11}$ | AAATGTGCTGTCTATCCCTACCTCATCTCTCTCTCTCTCTCTCTCTCTGGGTCTGTAGTTGTAGATTGTGGT | 97-103 | F | TTCTTGCAAAGCTGTGTTGAGT |
| | | | | | | R | CGGATGACCACAATCTACAACT |
| 6 | *SqMS36* | seq_22704166 | $(AC)_9$ | AATTTAAGAGGCAGGACATGGAAATACACACACACACACACACATTCATCTGCTCTCTGGGGGGGTCA | 94-112 | F | ATTTAAGAGGCAGGACATGGAA |
| | | | | | | R | TCCTCCTACTGAGCATCCCA |
| 7 | *SqMS43* | seq_62713477 | $(GT)_{20}$ | AAGGTAAAGGTGCACGGTATCATCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTGTGTATGTGGGGTGCTGAGACTG | 89-139 | F | GCGTGTCTCTGTGGAGCAAA |
| | | | | | | R | TCAGTCTCAGCACCCCACAT |
| 8 | *SqMS45* | seq_67519894 | $(CA)_{17}$ | ATCACCGCCCTCAATCTGAGAGACTCACACACACACACACACACACACACACACACACAAACACATCCCCTTTTTTCACATCAA | 121-150 | F | ACCTCTGCACTGCTAGCTTC |
| | | | | | | R | ACACAGAAAACACAGACACACC |

## 2.5 Fragment analysis

A Fragment analysis was performed confirm the genetic diversity in each group of 90 Korean and 90 Japanese yellowtail using the eight candidate microsatellite markers. A primer labeled with 5'-FAM specific to each marker was synthesized and used for PCR. The PCR product was genotyped using the ABI3730xL Genetic Analyzer (Applied Biosystems) equipped with a capillary array, and GeneMapper was used to determine allele sizes (Applied Biosystems). The amplification products were separated by size using capillary electrophoresis. The genotype of each individual was determined according to the size of the product. The genotype and allele frequencies for eight loci for each group of Korean (KR, n = 90) and Japanese (JP, n = 90) yellowtail are shown.

## 2.6 Statistical analyses

### 2.6.1 TASSEL

TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) is a Java-based analysis program for genotype analyses using genome and restriction enzyme information, such as GBS data. TASSEL consists of two main pipelines: Discovery and Production [12].

### 2.6.2 Cervus 3.0.7

Cervus was used for the assignment of parents to their offspring based on genetic markers. Cervus uses the likelihood, a well-established statistical method for parentage analysis, to assign parentage. Cervus introduces two key enhancements to this process. Likelihood ratios are calculated allowing for the

possibility that the genotypes of parents and offspring may be mistyped. Simulation is used to determine the level of confidence in the assigned parentage.

2.6.3 FSTAT 2.9.3

FSTAT is an easy tool to evaluate gene diversities and differentiation statistics from codominant genetic markers.

2.6.4 Arlequin 3.5.2.2

Arlequin was used to estimate various population genetics parameters. The statistical tests implemented in Arlequin were chosen to minimize hidden assumptions and maximize power.

# 3. Results

## 3.1 GBS Library

For the GBS analysis of Korean and Japanese *S. quinqueradiata* individuals, the result of gel loading results for one set (96 samples) are summarized in Figure 3 and Table 5.

The quality of the complete GBS library was checked by LightCycler qPCR just before analysis using the NovaSeq 6000 platform. Additional information, such as the concentrations of these libraries, is provided in Figure 4 and Table 6.

NGS (NoveSeq 6000 platform) was performed from one set of the *S. quinqueradiata* GBS library to obtain about 103 Gb of raw data (*S. quinqueradiata* GBS-1). The Q30 values were approximately 95.21% (*S. quinqueradiata* GBS-1). Detailed results are shown in Table 7.

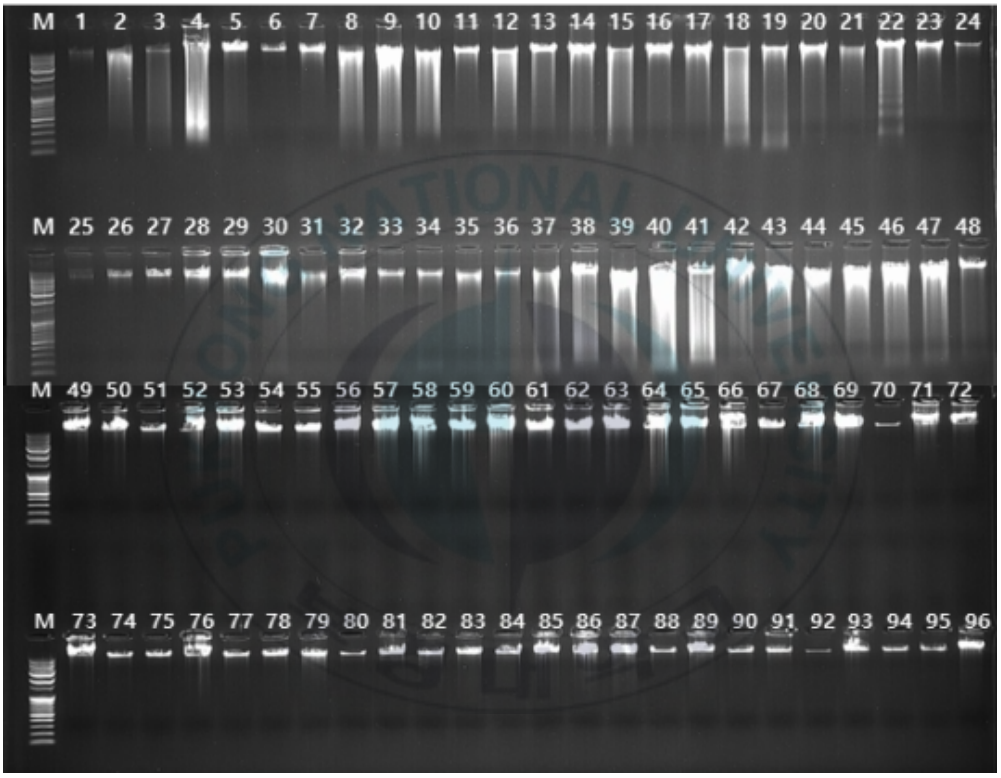


**Figure 3. Electrophoresis results for *S. quinqueradiata* GBS library construction**

**Table 5. Lane information for the electrophoresis results for samples used to construct *S. quinqueradiata* GBS libraries**

| No. | Barcode | Sample | No. | Barcode | Sample | No. | Barcode | Sample | No. | Barcode | Sample |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | CTCC | Korean-31 | 25 | ATTGA | Korean-64 | 49 | CTATTA | Japanese-95 | 73 | GCGGAAT | Japanese-121 |
| 2 | TGCA | Korean-32 | 26 | CATCT | Korean-65 | 50 | GCCAGT | Japanese-96 | 74 | TAGCGGA | Japanese-132 |
| 3 | ACTA | Korean-33 | 27 | CCTAC | Korean-66 | 51 | GGAAGA | Japanese-97 | 75 | TCGAAGA | Japanese-133 |
| 4 | CAGA | Korean-34 | 28 | GAGGA | Korean-67 | 52 | GTACTT | Japanese-98 | 76 | TCTGTGA | Japanese-134 |
| 5 | AACT | Korean-36 | 29 | GGAAC | Korean-68 | 53 | GTTGAA | Japanese-99 | 77 | TGCTGGA | Japanese-137 |
| 6 | GCGT | Korean-37 | 30 | GTCAA | Korean-69 | 54 | TAACGA | Japanese-100 | 78 | ACGACTAC | Japanese-138 |
| 7 | CGAT | Korean-38 | 31 | TAATA | Korean-70 | 55 | TGGCTA | Japanese-101 | 79 | TAGCATGC | Japanese-139 |
| 8 | GTAA | Korean-39 | 32 | TACAT | Korean-71 | 56 | TATTTTT | Japanese-102 | 80 | TAGGCCAT | Japanese-140 |
| 9 | AGGC | Korean-42 | 33 | TCGTT | Korean-72 | 57 | CTTGCTT | Japanese-103 | 81 | TGCAAGGA | Japanese-141 |
| 10 | GATC | Korean-44 | 34 | GGTTGT | Korean-73 | 58 | ATGAAAC | Japanese-104 | 82 | TGGTACGT | Japanese-142 |
| 11 | TCAC | Korean-45 | 35 | CCAGCT | Korean-74 | 59 | AAAAGTT | Japanese-105 | 83 | TCTCAGTC | Japanese-143 |
| 12 | TGCGA | Korean-46 | 36 | TTCAGA | Korean-75 | 60 | GAATTCA | Japanese-106 | 84 | CCGGATAT | Japanese-144 |
| 13 | CGCTT | Korean-47 | 37 | TAGGAA | Korean-76 | 61 | GAACTTC | Japanese-107 | 85 | CGCCTTAT | Japanese-145 |
| 14 | TCACC | Korean-48 | 38 | GCTCTA | Korean-78 | 62 | GGACCTA | Japanese-108 | 86 | AACCGAGA | Japanese-146 |
| 15 | CTAGC | Korean-49 | 39 | CCACCA | Korean-79 | 63 | GTCGATT | Japanese-109 | 87 | ACAGGGAA | Japanese-147 |
| 16 | ACAAA | Korean-52 | 40 | CTTCCA | Korean-80 | 64 | AACGCCT | Japanese-110 | 88 | ACGTGGTA | Japanese-148 |
| 17 | TTCTC | Korean-53 | 41 | GAGATA | Korean-81 | 65 | AATATGC | Japanese-111 | 89 | CCATGGGT | Japanese-149 |
| 18 | AGCCC | Korean-54 | 42 | ATGCCT | Korean-82 | 66 | ACGTGTT | Japanese-113 | 90 | CGCGGAGA | Japanese-150 |
| 19 | GTATT | Korean-55 | 43 | AGTGGA | Korean-83 | 67 | ATTAATT | Japanese-114 | 91 | CGTGTGGT | Japanese-151 |
| 20 | CTGTA | Korean-57 | 44 | ACCTAA | Korean-84 | 68 | ATTGGAT | Japanese-115 | 92 | GCTGTGGA | Japanese-152 |
| 21 | ACCGT | Korean-60 | 45 | ATATGT | Korean-85 | 69 | CATAAGT | Japanese-117 | 93 | GGATTGGT | Japanese-153 |
| 22 | GCTTA | Korean-61 | 46 | ATCGTA | Korean-86 | 70 | CGCTGAT | Japanese-118 | 94 | GTGAGGGT | Japanese-156 |
| 23 | GGTGT | Korean-62 | 47 | CATCGT | Korean-90 | 71 | CGGTAGA | Japanese-119 | 95 | TATCGGGA | Japanese-160 |
| 24 | AGGAT | Korean-63 | 48 | CGCGGT | Korean-91 | 72 | CTACGGA | Japanese-120 | 96 | TTCCTGGA | Japanese-161 |

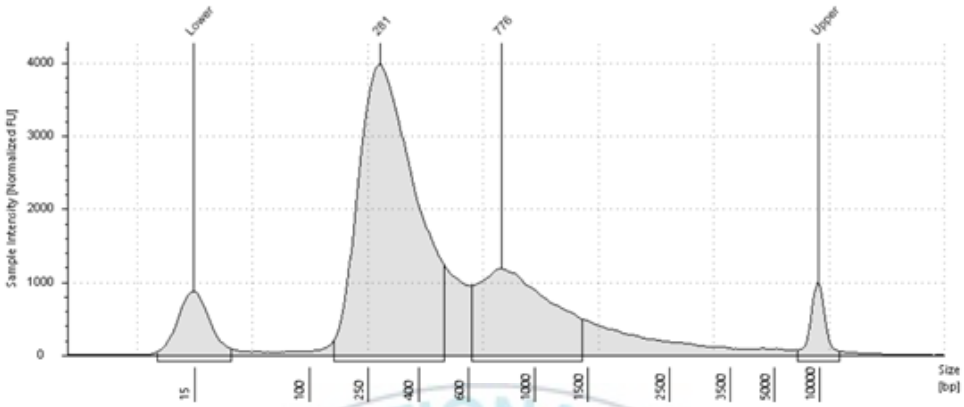


**Figure 4. QC results for each *S. quinqueradiata* GBS library obtained by LightCycler qPCR before obtaining raw GBS data**

The quality of the complete GBS library was checked by LightCycler qPCR just before analysis using the NovaSeq 6000 platform. The Lowest sample intensity is 15 bp, and the uppermost sample intensity is 10000 bp.

**Table 6. QC results for each *S. quinqueradiata* GBS library obtained by LightCycler qPCR before obtaining raw GBS data**

| Library name | Conc. (ng/ul) | Conc. (nM) | size (bp) |
|---|---|---|---|
| GBS-1 | 13.72 | 42.83 | 493 |

**Table 7. Statistical summary of raw *S. quinqueradiata* NGS data**

| Sample | Composition of the samples | Read -pairs # | Average read # | Total bases (bp) | Average bp per individual | Q30 (%) | clean read # | bases of clean reads (bp) |
|---|---|---|---|---|---|---|---|---|
| S.quin GBS-1 | Korean 48 + Japanese 48 | 725,367,956 | 7,555,916 | 109,530,561,356 | 1,140,943,347 | 93.18 | 658,115,494 | 89,657,806,187 |

### 3.2 Identification of candidate microsatellite DNA sequences using data generated by GBS

The microsatellite DNA sequences were searched using MISA from each data generated by the GBS-1 set, as shown in Table 8. To design primer sets using these results, Primer 3 was used.

Microsatellite DNA markers were searched against each sample using MISA. Two-nucleotide repeats were most frequent, three or more nucleotide repeats. Similar trends were detected in all of the investigated samples. Plots are shown in Figure 5.

**Table 8. Microsatellite DNA marker analysis results based on *S. quinqueradiata* GBS analysis data**

| Name | *S. quinqueradiata* GBS (Reads) |
|---|---|
| Total number of sequences examined | 115,253,450 |
| Total size of examined sequences (bp) | 15,230,032,123 |
| Total number of identified SSRs | 16,063,461 |
| Number of SSR containing sequences | 12,541,071 |
| Number of sequences containing more than 1 SSR | 2,889,870 |
| Number of SSRs present in compound formation | 3,519,628 |

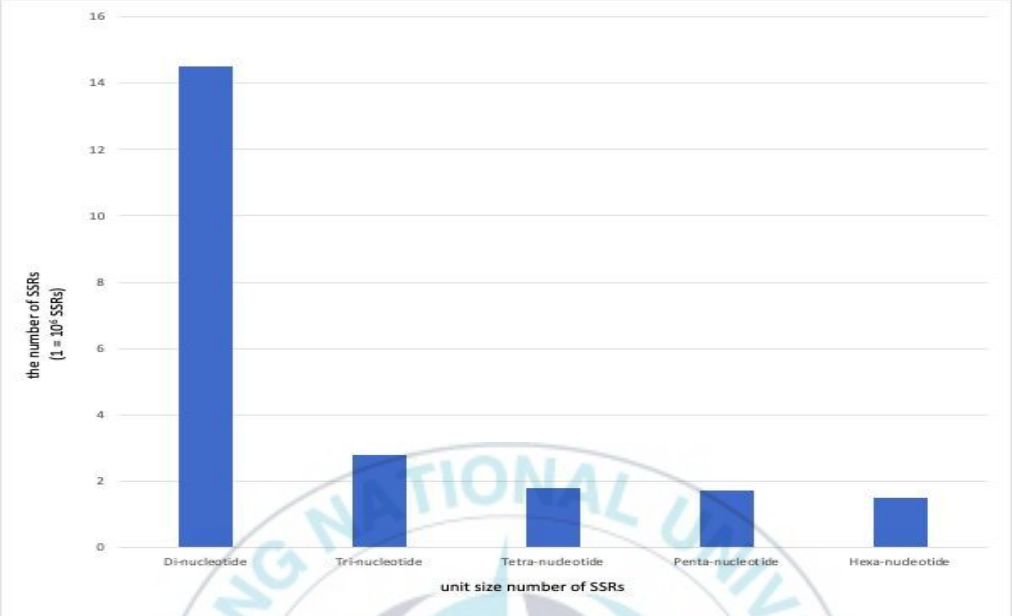


**Figure 5. Distribution of repeat types**

Microsatellite DNA markers were searched against each sample using MISA. Two-nucleotide repeats were most frequent, three or more nucleotide repeats. Similar trends were detected in all of the investigated samples.

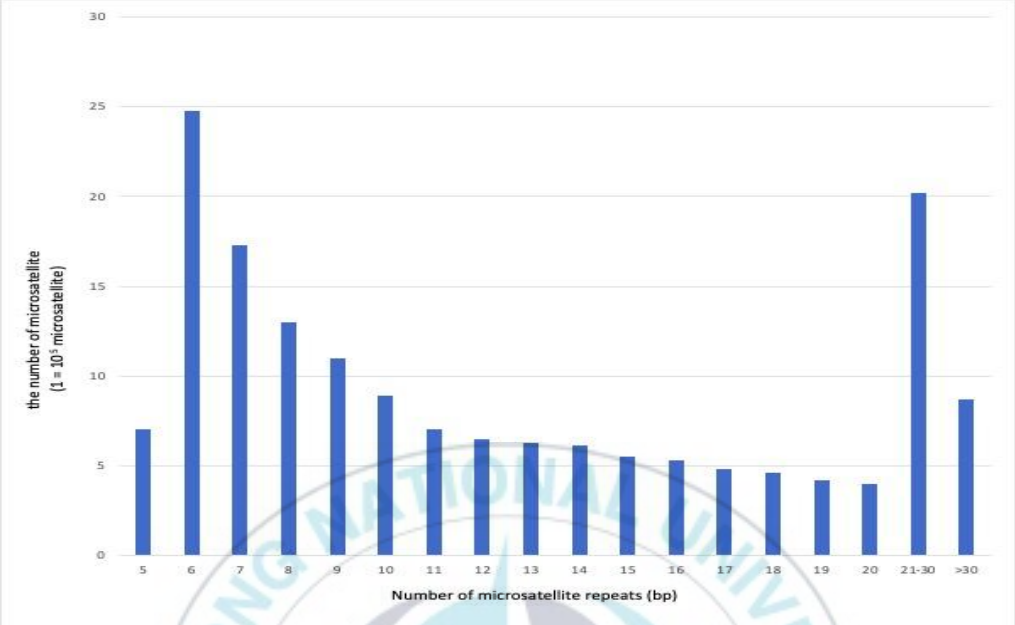


**Figure 6. Frequency of identified SSR motifs**

Based on the microsatellite DNA markers for each sample, the number of repetitions of the microsatellite marker unit were determined. The most frequent number of repeats was 6 repetitions (14.71%), followed by 7 repetitions (10.47%), 8 repetitions (7.88%), 9 repetitions (6.52%), 10 repetitions (5.64%), and 5 repetitions (4.01%). For repeat numbers exceeding 10, the frequency decreased substantially.

Classification of SSR motif sequences identified from reads for each sample. From the data for *S. quinqueradiata* GBS sets obtained by NGS sequencing, the degree of repetition of each microsatellite DNA marker was confirmed. The results for the top 20 repeat sequences are shown in Figure 7.


**Figure 7. Frequency distribution of repeat types**

The sequence with the most repeats was AC/GT (74.08%), followed by AG/CT (13.04%), with sharp decreases in the frequencies of other sequences.

To select candidate microsatellite DNA markers among motif sequences, the flanking sequence of each SSR sequence was first investigated. Assuming that these flanking sequences are identical to the SSR sequence at the same locus, they were collected and selected by SSR type. In addition, the distributions of these SSRs sequences in Korean and Japanese was investigated. Significant sequences with a p-value of less than 0.05 were selected. A total of 2,628 sequences were selected as primary candidates. Using these sequences, primer sets were designed using Primer 3 (Table 9).

**Table 9. Primary microsatellite DNA marker candidates**

| Marker | ID | P-value | SSR | Tm(℃) | Product size(bp) | Marker | ID | P-value | SSR | Tm(℃) | Product size(bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SqMS1 | seq_30332860 | 3.80E-14 | (TGC)6 | 59.605 | 121 | SqMS36 | seq_22704166 | 0.00116 | (AC)7 | 60.107 | 113 |
| SqMS2 | seq_5856219 | 3.80E-14 | (AC)10 | 60.057 | 104 | SqMS37 | seq_3467931 | 0.00116 | (TG)6 | 59.579 | 100 |
| SqMS3 | seq_89416432 | 1.52E-13 | (CACT)9 | 58.143 | 111 | SqMS38 | seq_28241752 | 0.00116 | (AAT)7 | 57.381 | 102 |
| SqMS4 | seq_32191481 | 1.52E-13 | (GT)7 | 59.894 | 129 | SqMS39 | seq_46720083 | 0.00116 | (TG)9 | 59.965 | 134 |
| SqMS5 | seq_75167185 | 1.52E-13 | (CCT)5(TCC)5 | 59.675 | 146 | SqMS40 | seq_33208710 | 0.00116 | (CA)9 | 59.924 | 122 |
| SqMS6 | seq_92025120 | 7.51E-12 | (ATC)6 | 57.506 | 139 | SqMS41 | seq_53175531 | 0.002587 | (CCTTCT)5 | 57.032 | 145 |
| SqMS7 | seq_83788636 | 2.55E-11 | (CA)8 | 57.268 | 107 | SqMS42 | seq_96252529 | 0.002587 | (AAT)9 | 58.715 | 100 |
| SqMS8 | seq_110229194 | 1.55E-07 | (CTC)7 | 62.135 | 121 | SqMS43 | seq_62713477 | 0.002587 | (CAG)7 | 59.891 | 123 |
| SqMS9 | seq_24194260 | 1.07E-06 | (TG)6 | 59.545 | 108 | SqMS44 | seq_3065461 | 0.002587 | (CA)9 | 60.321 | 123 |
| SqMS10 | seq_51666652 | 2.72E-06 | (CCT)5ccg(CCT)5 | 59.963 | 103 | SqMS45 | seq_67519894 | 0.002587 | (AC)7 | 59.678 | 101 |
| SqMS11 | seq_53340553 | 4.03E-05 | (CA)24 | 59.547 | 100 | SqMS46 | seq_25815671 | 0.002587 | (TG)7 | 59.68 | 107 |
| SqMS12 | seq_58803377 | 9.56E-05 | (GACA)6 | 59.51 | 118 | SqMS47 | seq_10840758 | 0.002587 | (TG)6 | 60.037 | 113 |
| SqMS13 | seq_24467822 | 0.000512 | (GTGGAC)6 | 59.11 | 114 | SqMS48 | seq_70136083 | 0.002587 | (GT)6gcgtgtgtgtctgtctc(TG)7 | 59.895 | 108 |
| SqMS14 | seq_19995732 | 0.000512 | (TCC)8 | 60.619 | 100 | SqMS49 | seq_13914864 | 0.002587 | (TG)8 | 59.385 | 114 |
| SqMS15 | seq_2117788 | 0.00116 | (GCA)8 | 59.824 | 101 | SqMS50 | seq_10770479 | 0.002587 | (TCA)11 | 60.364 | 110 |
| SqMS16 | seq_50121102 | 0.00116 | (AC)11 | 57.152 | 106 | SqMS51 | seq_2596890 | 0.002587 | (TG)11 | 60.81 | 111 |
| SqMS17 | seq_35676434 | 0.002587 | (CTG)8 | 60.319 | 101 | SqMS52 | seq_28827718 | 0.002587 | (CTC)7 | 60.252 | 117 |
| SqMS18 | seq_41659523 | 0.002587 | (CT)6g(TC)8 | 58.772 | 120 | SqMS53 | seq_32276145 | 0.002587 | (GA)10 | 59.757 | 100 |
| SqMS19 | seq_113252850 | 0.002587 | (CA)8 | 60.762 | 132 | SqMS54 | seq_71808836 | 0.002587 | (CAA)7 | 59.068 | 129 |
| SqMS20 | seq_94101440 | 0.002587 | (CA)8 | 59.32 | 137 | SqMS55 | seq_67961165 | 0.002587 | (AACAC)14 | 59.868 | 131 |
| SqMS21 | seq_1092540952 | 0.002587 | (AC)6 | 59.399 | 104 | SqMS56 | seq_18169444 | 0.002587 | (AC)13 | 59.602 | 104 |
| SqMS22 | seq_17831053 | 0.002587 | (GCTGGA)7 | 60.077 | 126 | SqMS57 | seq_89061399 | 0.026474 | (AC)14 | 59.536 | 107 |
| SqMS23 | seq_28803135 | 0.002587 | (TG)6 | 60.39 | 139 | SqMS58 | seq_54908141 | 0.000223 | (GAG)6 | 58.489 | 114 |
| SqMS24 | seq_36178528 | 0.002587 | (AG)8 | 57.055 | 124 | SqMS59 | seq_1863411 | 0.012355 | (TG)7 | 60.688 | 118 |
| SqMS25 | seq_58473513 | 6.81E-06 | (GT)8 | 59.1 | 133 | SqMS60 | seq_15216677 | 0.026474 | (TTG)6 | 59.96 | 138 |
| SqMS26 | seq_89362985 | 1.67E-05 | (CA)10 | 59.391 | 117 | SqMS61 | seq_25199294 | 0.026474 | (CTG)7 | 58.412 | 110 |
| SqMS27 | seq_43464773 | 9.56E-05 | (GT)8 | 59.497 | 136 | SqMS62 | seq_12886800 | 6.31E-20 | (CAA)7 | 59.809 | 134 |
| SqMS28 | seq_11178693 | 9.56E-05 | (AC)20 | 57.364 | 125 | SqMS63 | seq_31324325 | 2.67E-10 | (TTA)10 | 59.56 | 109 |
| SqMS29 | seq_42577439 | 0.000223 | (TG)13 | 60.022 | 100 | SqMS64 | seq_81791988 | 5.70E-08 | (CA)7 | 59.704 | 106 |
| SqMS30 | seq_46674313 | 0.000223 | (TCC)8 | 60.036 | 115 | SqMS65 | seq_5765429 | 4.11E-07 | (CA)10 | 59.746 | 105 |
| SqMS31 | seq_35518038 | 0.00512 | (TTA)8 | 59.822 | 105 | SqMS66 | seq_24648897 | 4.11E-07 | (GAG)7 | 60.396 | 103 |
| SqMS32 | seq_11310468 | 0.00512 | (AC)6tcacgcgt(AC)6 | 61.332 | 107 | SqMS67 | seq_41615583 | 4.11E-07 | (TC)10 | 59.363 | 116 |
| SqMS33 | seq_48954388 | 0.00512 | (GT)8 | 60.675 | 136 | SqMS68 | seq_20925314 | 4.11E-07 | (CCT)6 | 59.66 | 101 |
| SqMS34 | seq_19509187 | 0.00512 | (GT)6 | 59.855 | 101 | SqMS69 | seq_35099026 | 1.07E-06 | (TGTT)7 | 57.798 | 109 |
| SqMS35 | seq_22704166 | 0.00116 | (GT)10 | 60.109 | 101 | | | | | | |

### 3.3 Microsatellite marker genotyping

To confirm the suitability of the 69 microsatellite markers identified from GBS data, PCR was performed using genomic DNA from each of the three Korean and three Japanese *S. quinqueradiata* as a templates. The product were confirmed by agarose gel electrophoresis, as shown in Figure 8.

For 46 primer sets, the number of individuals was increased and additional PCRs were performed for secondary selection. As a result, eight candidates whose size variation was confirmed by electrophoresis were selected.

A fragment analysis was performed to confirm the genetic diversity of each group of 90 Korean and 90 Japanese *S. quinqueradiata* using the eight microsatellite markers. A primer labeled with 5'-FAM specific to each marker was synthesized and used for PCR. The PCR product was genotyped using the ABI3730xL Genetic Analyzer (Applied Biosystems) equipped with a capillary array and allele sizes were calculated using GeneMapper 5 (Applied Biosystems) (Figure 9). The amplified product was separated by capillary electrophoresis to determine the size of the DNA fragment and to determine individual genotypes. The allele frequencies for eight loci for each group of Korean (KR, n = 90) and Japanese (JP, n = 90) fish are shown in Figure 10.

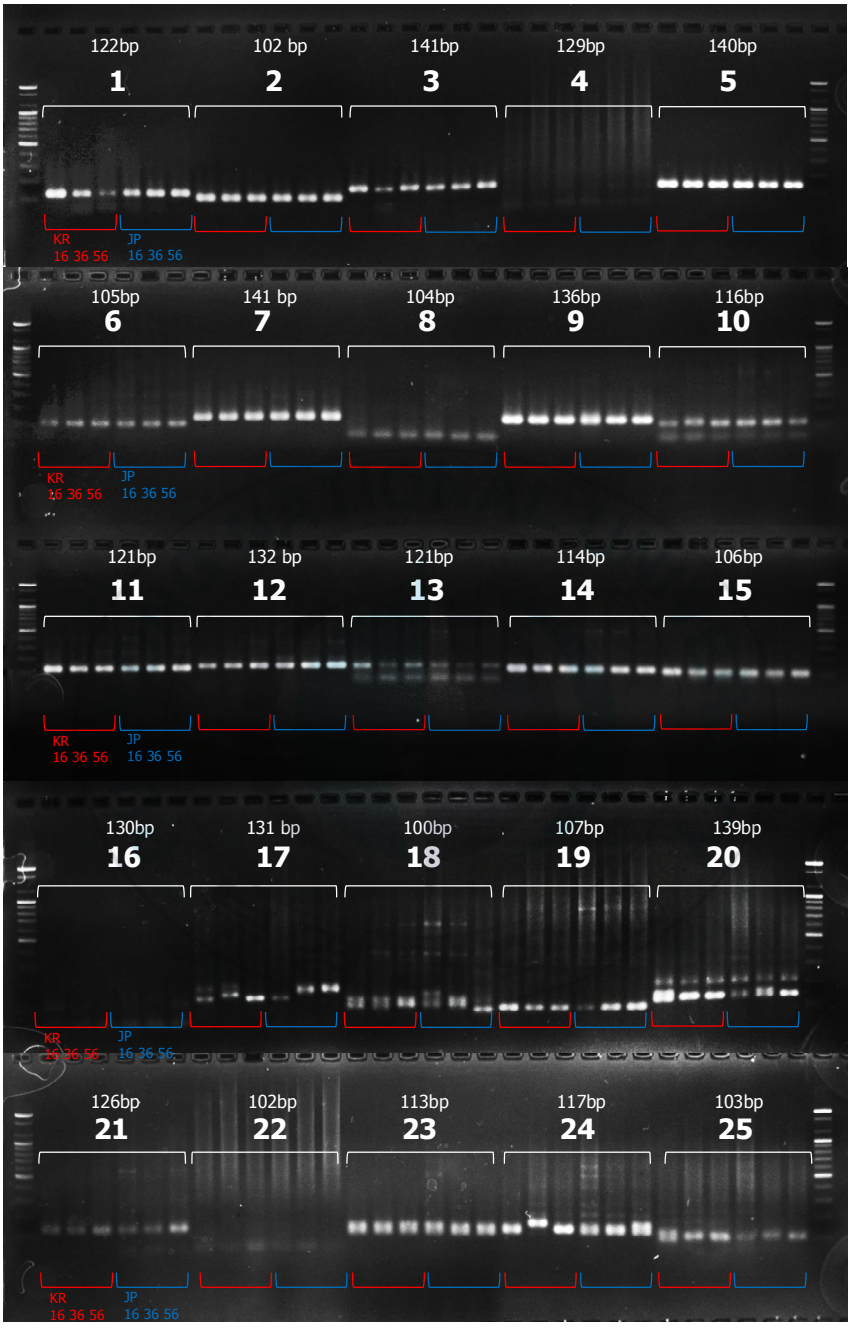


**Figure 8. Primary microsatellite marker selection using three Korean and three Japanese individuals**

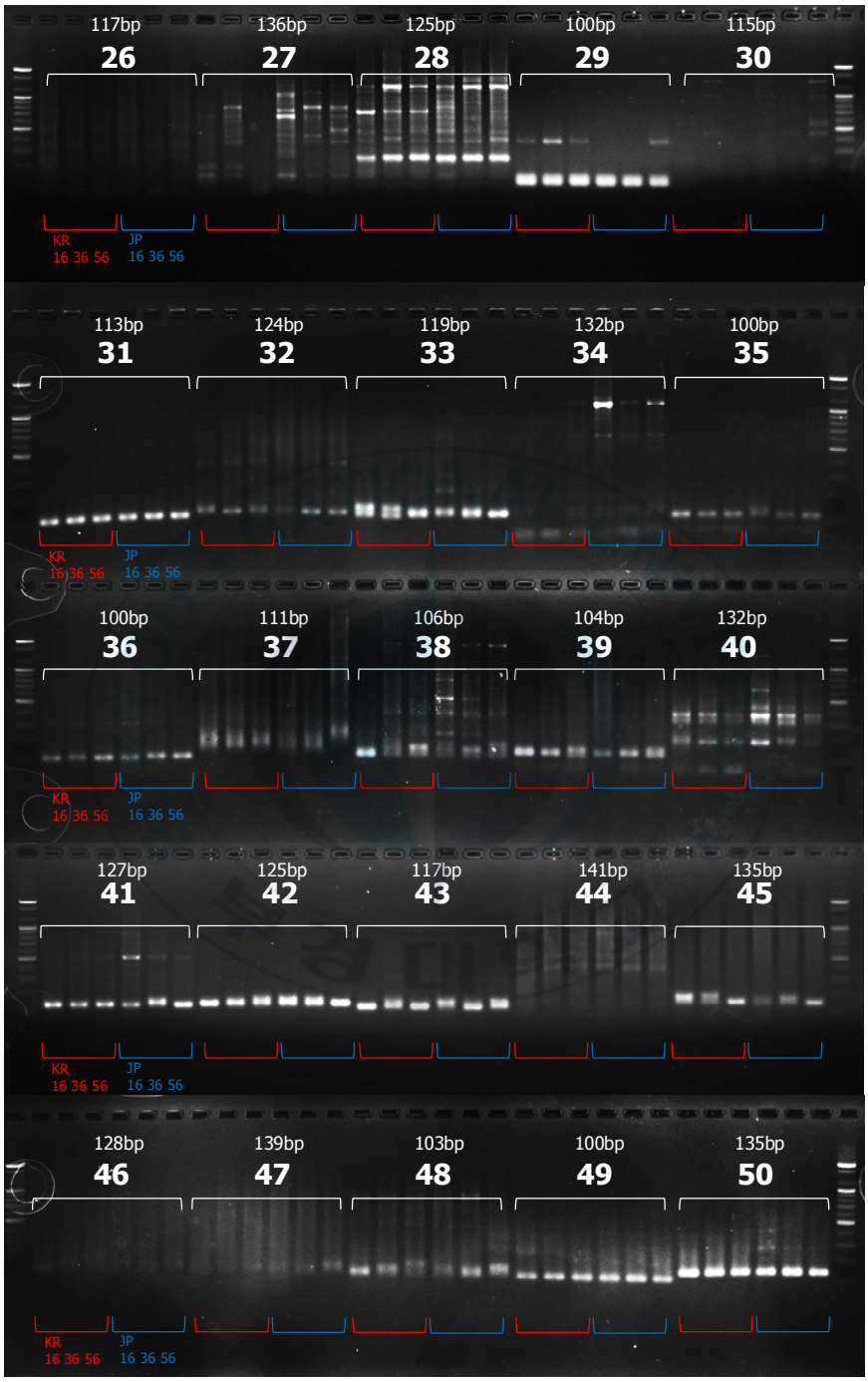


**Figure 8. (continued)**

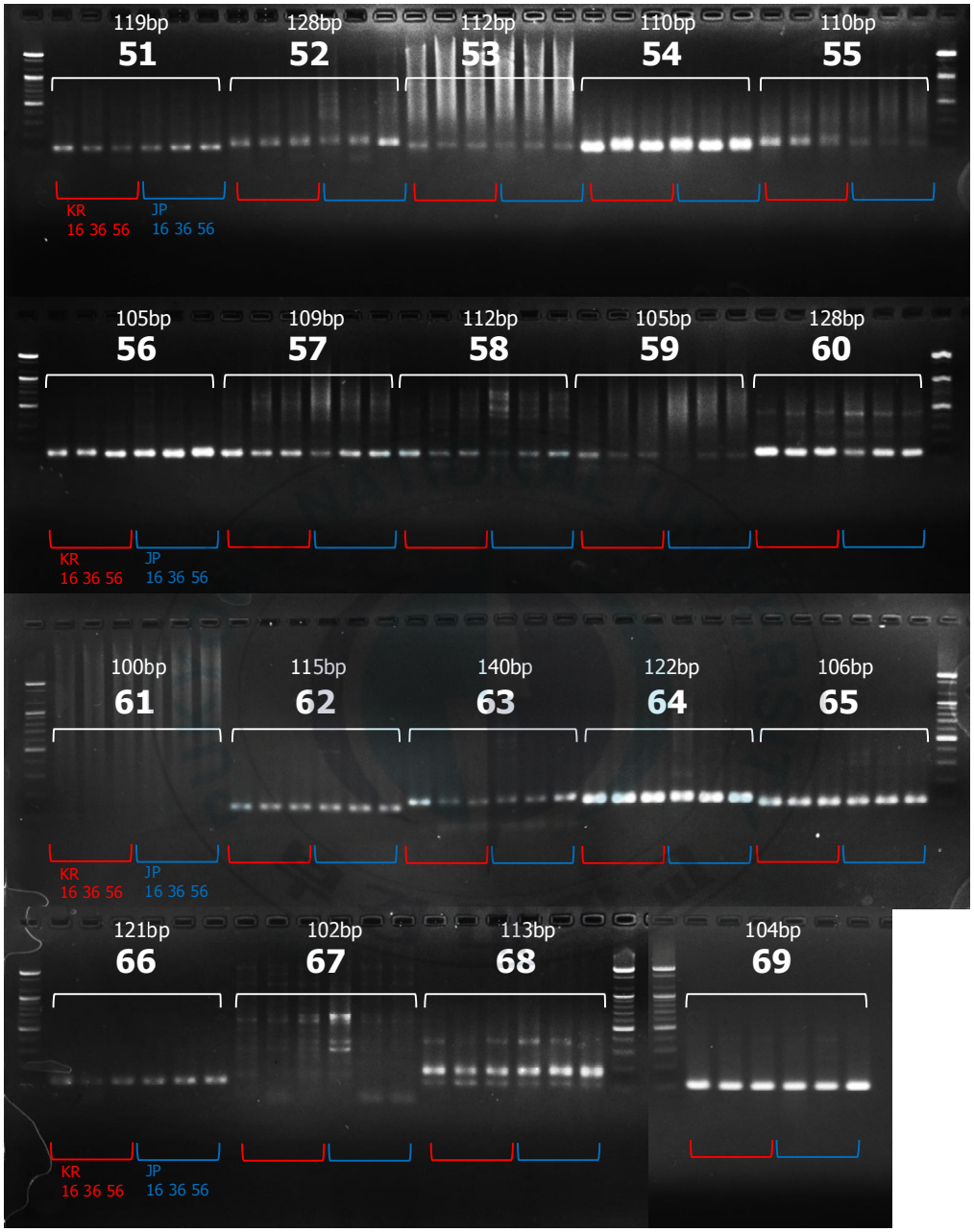


**Figure 8. (continued)**

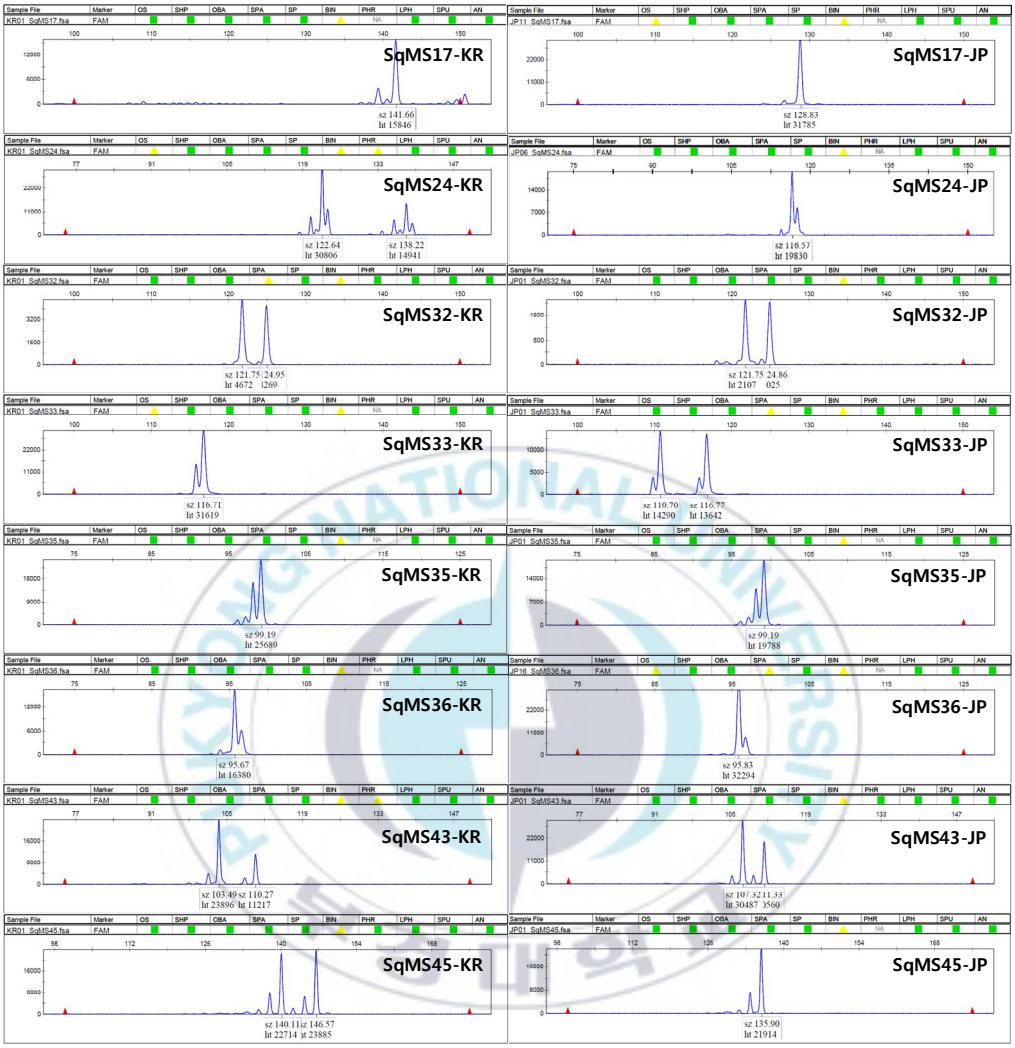


**Figure 9. DNA fragment size analysis of Korean and Japanese *S. quinqueradiata* using eight microsatellite markers**


**Figure 10. Allele frequencies of Korean and Japanese *S. quinqueradiata* against eight microsatellite marker loci using DNA fragment capillary electrophoresis**

### 3.4 Analysis of genetic polymorphism by the origin of *S. quinqueradiata*

The number of alleles per locus (*N*a), observed heterozygosity (*H*o), expected heterozygosity (*H*e), and polymorphic information content (PIC) were calculated using Cervus version 3.0.7 (Marshall et al., 1998). Allelic richness (*Ar*) and the inbreeding coefficient ($F_{IS}$) were calculated using the FSTAT 2.9.4 (Goudet 1995). The p-value (pHWD) for the Hardy-Weinberg equilibrium test was calculated using Arlequin 3.5.2.2 (Exocoffier et al., 2005).

When genetic variation was analyzed by setting 90 Korean (KR, n = 90) and Japanese (JP, n = 90) specimens as one population using eight microsatellite markers, 103 alleles were detected. The overall mean number of alleles per locus ranged from 4 to 29, with an average of 12.9. The fewest were observed in SqMS35 (i.e., 4) and the highest number (i.e., 29) was observed in SqMS43. The overall mean of allelic richness (*Ar*) was 11.607;and SqMS43 (Ar: 25.155) showed the highest value and SqMS35 (Ar: 3.664) showed the lowest value.

*H*e for each group represents the expected degree of heterozygosity assuming the population is in equilibrium, and *H*o represents the observed degree of heterozygosity. The overall mean *H*e for eight loci and 180 subjects was 0.633, and the overall mean *H*o was 0.497. *H*o was higher than *H*e at all loci except for SqMS45. When viewed by locus, *H*o (0.926) was highest in SqMS43.

The average PIC was 0.612 for 180 subjects, and SqMS43 showed the highest value. Deviations from Hardy-Weinberg equilibrium were observed in both groups in SqMS17 and SqMS24 ($p < 0.0001$). In addition, deviation from the Hardy-Weinberg equilibrium was observed in Korean SqMS35 and

Japanese SqMS43.

$F_{IS}$ was, on average, 0.258 in Korean and 0.179 in Japanese samples, indicating that heterozygosity was low due to a high degree of inbreeding. Overall, *S. quinqueradiata* did not show a significant differentiation between populations of the two origins based on the eight newly established markers.

Pairwise $F_{ST}$ values decreased in the order of SqMS32, SqMS17, SqMS43, SqMS36, SqMS45, SqMS33, SqMS35, and SqMS24; only SqMS32 and SqMS43 showed significant differentiation (i.e.,$p < 0.00001$ and $p < 0.05$). In the case of SqMS17, deviations from Hardy-Weinberg equilibrium were observed in both the Korean and Japanese groups. In a comparison between groups (i.e., with respect to origin) based on five markers with high $F_{ST}$ values (SqMS32, SqMS33, SqMS36, SqMS43 and SqMS45), $F_{ST}$ values were 0.11347 ($p < 0.00001$) and a comparion based on the top three markers (SqMS32, SqMS36, and SqMS43) showed higher $F_{ST}$ values of 0.18779 ($p < 0.00001$).

A PCoA for each marker revealed that the proportion of total variance explained by SqMS32, which had the highest $F_{ST}$ value, was 84.66% (48.80%, 19.18% and 16.68% for each component).

Korean and Japanese populations were clearly distinguished on the two-dimensional graph, and SqMS32 was included in all microsatellite marker combinations for *S. quinqueradiata* origin classification (Tables 10, 11).

**Table 10. Analysis of the genetic diversity indexes for domestic (n = 90) and Japanese (n = 90) *S. quinqueradiata* populations based on 8 microsatellite markers for origin identification**

| Populations | | SqMS17 | SqMS24 | SqMS32 | SqMS33 | SqMS35 | SqMS36 | SqMS43 | SqMS45 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|
| KR (n=90) | Na | 5 | 15 | 3 | 9 | 4 | 11 | 27 | 13 | 10.9 |
| | Ar | 5 | 12.989 | 2.433 | 7.897 | 3.328 | 9.042 | 20.810 | 10.696 | 9.024 |
| | $H_o$ | 0.011 | 0.506 | 0.539 | 0.700 | 0.128 | 0.450 | 0.817 | 0.828 | 0.497 |
| | $H_e$ | 0.280 | 0.840 | 0.411 | 0.728 | 0.147 | 0.536 | 0.923 | 0.829 | 0.587 |
| | PIC | 0.265 | 0.818 | 0.330 | 0.681 | 0.141 | 0.512 | 0.912 | 0.804 | 0.558 |
| | HWE | 0 | 0 | 0.493 | 0.241 | 0 | 0.001 | 0.084 | 0.552 | 0.172 |
| | $F_{IS}$ | 1 | 0.399 | 0.109 | 0.054 | 0.247 | 0.171 | 0.085 | -0.005 | 0.258 |
| | Anull | 0.216 | 0.179 | 0.030 | 0.020 | 0.031 | 0.058 | 0.038 | -0.005 | 0.071 |
| JP (n=90) | Na | 6 | 12 | 8 | 8 | 3 | 11 | 23 | 14 | 10.6 |
| | Ar | 6 | 11.485 | 7.512 | 7.512 | 2.533 | 8.242 | 20.383 | 12.116 | 9.473 |
| | $H_o$ | 0.021 | 0.506 | 0.711 | 0.711 | 0.144 | 0.456 | 0.788 | 0.822 | 0.520 |
| | $H_e$ | 0.380 | 0.838 | 0.708 | 0.708 | 0.136 | 0.461 | 0.925 | 0.823 | 0.622 |
| | PIC | 0.360 | 0.814 | 0.658 | 0.658 | 0.127 | 0.423 | 0.913 | 0.799 | 0.594 |
| | HWE | 0 | 0 | 0.387 | 0.383 | 1.000 | 0.127 | 0 | 0.202 | 0.262 |
| | $F_{IS}$ | 0.946 | 0.397 | -0.005 | -0.005 | -0.066 | 0.013 | 0.148 | 0.001 | 0.179 |
| | Anull | 0.258 | 0.178 | -0.004 | -0.004 | -0.009 | 0.002 | 0.068 | -0.002 | 0.061 |
| All population (n=180) | Na | 7 | 16 | 11 | 9 | 4 | 12 | 29 | 15 | 12.9 |
| | Ar | 7 | 14.174 | 9.862 | 8.657 | 3.664 | 10.764 | 25.155 | 13.579 | 11.607 |
| | $H_o$ | 0.011 | 0.506 | 0.539 | 0.700 | 0.128 | 0.450 | 0.817 | 0.828 | 0.497 |
| | $H_e$ | 0.338 | 0.838 | 0.780 | 0.717 | 0.141 | 0.500 | 0.926 | 0.826 | 0.633 |
| | PIC | 0.327 | 0.818 | 0.749 | 0.671 | 0.135 | 0.472 | 0.919 | 0.804 | 0.612 |

**Table 11. $F_{ST}$ values for comparison between Korean and Japanese populations**

| Locus | Fst | p-value |
|---|---|---|
| SqMS17 | 0.01695 | 0.25225 |
| SqMS24 | -0.00327 | 0.96396 |
| SqMS32 | 0.44044 | p < 0.00001 |
| SqMS33 | -0.00199 | 0.56757 |
| SqMS35 | -0.00273 | 0.67568 |
| SqMS36 | 0.00439 | 0.22523 |
| SqMS43 | 0.0059 | 0.03604 |
| SqMS45 | -0.00036 | 0.42342 |
| Total 8 markers | 0.11078 | p < 0.00001 |
| 5 markers (SqMS32, SqMS33, SqMS36, SqMS43, SqMS45) | 0.11347 | p < 0.00001 |
| 3 markers (SqMS32, SqMS36, SqMS43) | 0.18779 | p < 0.00001 |

### 3.5 Comparison of genetic affinity by the origin of *S. quinqueradiata*

An AMOVA (analysis of molecular variance) for both domestic and Japanese groups and principal coordinates analysis (PCoA) were performed to two-demensionally represent relationships between groups of different origin based on genetic variation using GenAlEx (Peakall and Smouse 2006). In addition, pairwise $F_{ST}$ values and genotype likelihood matrix values were calculated using Arlequin 3.5.2.2 as an intergroup genetic relationship metric using eight microsatellite loci, and the results were plotted using OriginPro 8.5.

### 3.6 Determination of the origin of Korean and Japanese *S. quinqueradiata*

Based on the genotype analysis results using the eight microsatellite markers, the genotype likelihood values obtained using Arlequin were plotted in two dimensions. However, in the PCoA, Korean and Japanese individuals were mixed and clear grouping by origin was not observed.

Using a combination of five microsatellite markers (SqMS32, SqMS33, SqMS36, SqMS43, and SqMS45) based on the $F_{ST}$ values, the genotype likelihood and PCoA anlaysis results revealed distinct clusters corresponding to Korean and Japanese *S. quinqueradiata*. In addition, even when a combination of three microsatellite markers (SqMS32, SqMS36, and SqMS43) was used, origin discrimination was possible in the genotype likelihood analysis, and the proportion of total variance explained was 46.95% in the PCoA. Differentiation between domestic and Japanese *S. quinqueradiata* was effectively reflected by the five-marker and three-marker combinations.

These results indicated that the eight newly developed markers sufficiently reflect genetic diversity within and between Korean and Japanese *S. quinqueradiata*. Combinations of five (SqMS32, SqMS33, SqMS36, SqMS43, and SqMS45) or three (SqMS32, SqMS36, and SqMS43) markers were most effective. We expect these marker combinations to be applied for the discrimination between Korean or Japanese *S. quinqueradiata* by PCoA and genotype likelihood matrix analyses.

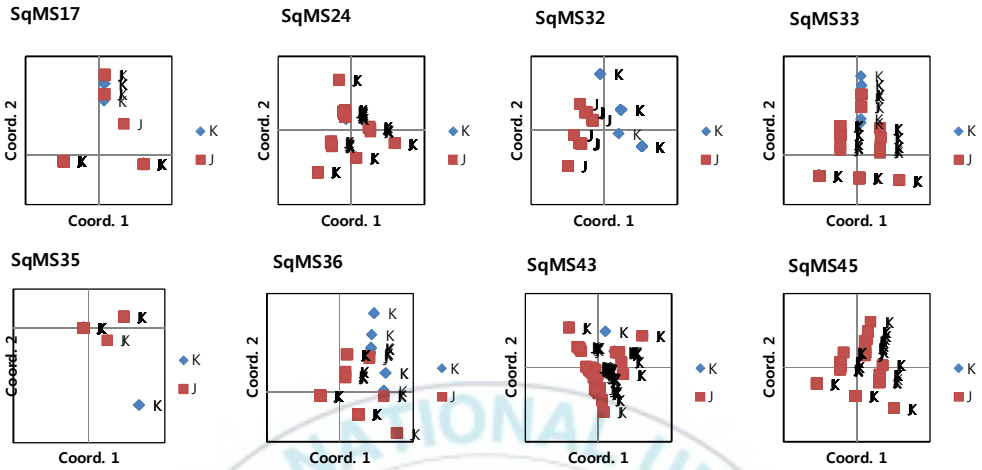


**Figure 11. Distribution of genetic variation in Korean and Japanese *S. quinqueradiata* for each of eight microsatellite markers using PCoA**

**Table 12. AMOVA of Korean and Japanese *S. quinqueradiata* for eight microsatellite markers**

| Source | Degrees of freedom | Sum of squares | Mean of squares | Estimated variation | Percentage variation |
|---|---|---|---|---|---|
| Among population | 1 | 23267.286 | 23267.286 | 62.558 | 1% |
| Among individual | 178 | 2137225.483 | 12006.885 | 5948.288 | 97% |
| Within individuals | 180 | 19855.500 | 110.308 | 110.308 | 2% |
| Total | 359 | 2180348.269 | | 6121.154 | 100% |

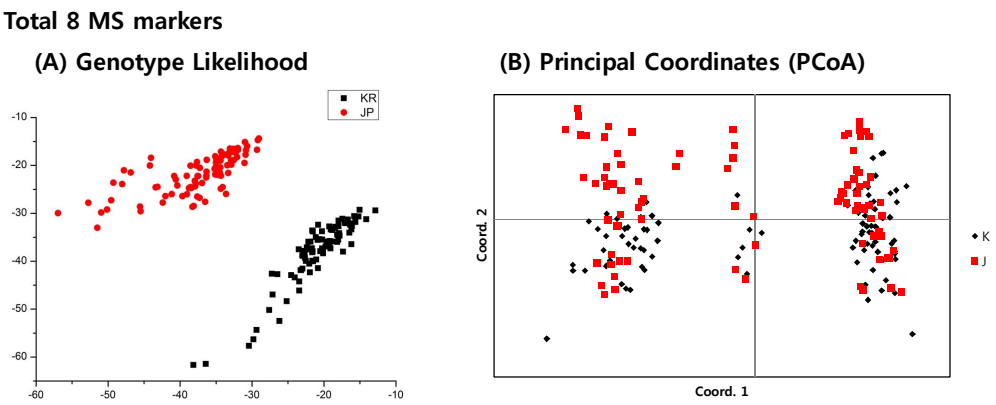


**Figure 12. Mapping results for the genetic distance between Korean and Japanese *S. quinqueradiata* individual based on the eight newly identified microsatellite markers determined by the genotype likelihood (A) and PCoA (B)**

5 MS markers (SqMS32, SqMS33, SqMS36, SqMS43, SqMS45)

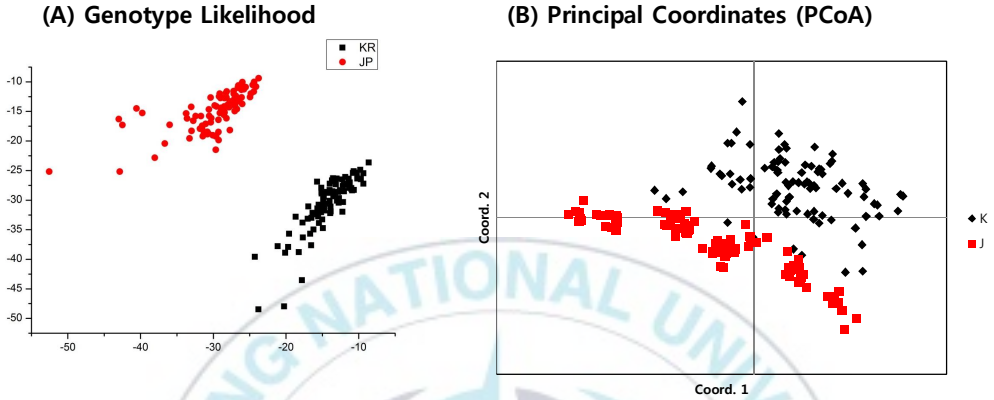


**Figure 13. Mapping results for the genetic distance between each Korean and Japanese *S. quinqueradiata* individual using the combination of five microsatellite markers based on the genotype likelihood (A) and PCoA (B)**

3 MS markers (SqMS32, SqMS36, SqMS43)

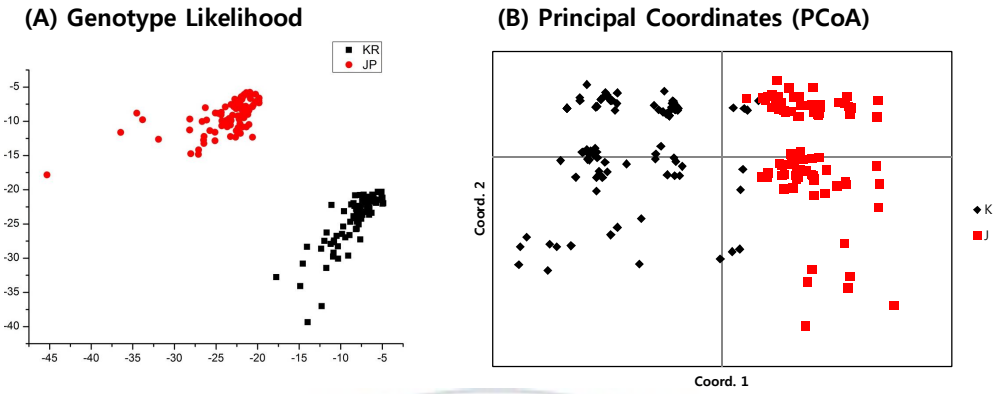


**Figure 14. Mapping results for the genetic distance between each Korean and Japanese *S. quinqueradiata* individual using the set of three microsatellite markers based on the genotype likelihood (A) and PCoA (B)**

## 4. Discussion

Based on a GBS analysis, microsatellite markers were used to distinguish between Korean and Japanese *S. quinqueradiata*. Among eight newly identified microsatellite markers, one marker showed a PIC of 0.6 and three markers showed high PIC values of 0.8. Compared with previously reported study, even considering the difference in sample size. Therefore, the markers in this study are appropriate for genetic analyses. When three microsatellite markers (SqMS32, SqMS36, and SqMS43) were anayzed together, Korean and Japanese *S. quinqueradiata* were effectively separated. These markers showed different allele frequency distributions between Korean and Japanese *S. quinqueradiata*.

The GBS analysis in this study is a cost-effective method because it does not require a reference genome for sequence tag mapping. Compared to alternative methods, such as RAD-seq, a small fregment size of 150-300 bp from a small amount of DNA (100 ng) is used as a target. Therefore, it is robust to a certain level of DNA degradation.

There is evidence that SNP markers, which are frequently used as molecular diagnostic markers, are not suitable for distinguishing between Korean and Japanese *S. quinqueradiata*. The results of this study indicate that microsatellite markers provide a more effective discrimination method.

However, The expression and functions of newly developed markers were not evaluated. This is a limitation of the library constructed by linking small fragments using restriction enzyme. Therefore, further studies of the relationship between genotypes, genes, and phenotypes are needed.

In order to more reliably identify newly developed microsatellite markers, it is possible to monitor whether additional *S. quinqueradiata* can be identified.

## 5. 국문요약

본 연구는 국내에서 유통되는 국내산과 일본산 방어의 원산지 판별을 위한 microsatellite marker의 개발과 이를 이용한 원산지 판별법 개발에 대한 것이다. 방어는 국내 어류 생산량에서 높은 비중을 차지한다. 통계청 자료에 따르면 2015년 298톤이었던 방어 수입량이 2019년 2,240톤으로 8배 가량 증가하였고, 국내 방어 수입량 대부분을 일본으로부터의 수입에 의존하고 있는 실정이다. 일본산 어류에 대한 우려와 안전한 수산물 먹거리에 대한 수요가 증가하는 반면 수산물 원산지 표시 위반 현황의 증가로 이를 방지하기 위해 보다 정확한 원산지 판별법을 개발할 필요성이 대두되었다. 본 연구에 사용한 microsatellite marker는 genome 전체에 산재되어 있는 반복 서열로 집단 간 변이 비교 분석을 위해 주로 사용되는 marker이다. 따라서 본 연구는 microsatellite marker를 이용하여 국내산과 일본산 방어의 원산지 판별법 개발로 부정 유통을 방지하기 위한 수입과 검역 과정에서의 활용 가능성을 제시한다.

국내산 방어 90개체와 일본산 방어 90개체를 각 3번에 걸쳐 총 180개의 방어 개체를 확보하여 지느러미와 근육에서 genomic DNA를 추출하였다. Whole genomic DNA에 제한효소를 처리하여 GBS library를 제작하고 NGS data를 확보하였다. 국내산과 일본산을 구분지을 수 있을 것이라고 판단되는 locus를 기준으로 69개의 후보군을 선정하였다. PCR 후 size variation을 확인하는 1,2차 스크리닝을 통해 8개의 marker를 선별하였다. 선별된 8개 marker에 대해 fragment analysis를 통해 대립유전자 크기를 산출하고 유전자형 분석을 수행하여 DNA fragment의 크기를 결정하였다. 8개 marker에 대한 fragment analysis raw data를 Cervus 3.0.7, Micro-checker 2.2.3, Arlequin 3.5.2.2, FSTAT 2.9.4, GenAlEx의 statistical analysis program으로 분석하였다.

새롭게 개발된 microsatellite marker의 조합을 이용하여 방어의 원산지 집단별 유전형을 분석하였다. 총 8개의 marker와 이를 이용한 효과적인 원산지 판별이 가능하다. 이 중 $F_{ST}$ 값을 기준으로 하여 5개(SqMS32, SqMS33, SqMS36, SqMS43, SqMS45) 또는 3개 (SqMS32, SqMS36, SqMS43) 의 microsatellite marker 조합을 통해 효율적인 원산지 판별이 가능함을 확인하였다.

해당 marker 조합으로 얻은 genotype 정보를 통해 PCoA, genotype likelihood matrix 분석을 수행함으로서 국내산과 일본산 방어의 원산지를 효과적으로 판별할 수 있을 것으로 예상된다.

8개의 microsatellite marker 사용 시 국내산과 일본산의 구분이 가능한 군집을 형성하는 것을 관찰할 수 있었다. 마찬가지로 통계 분석으로 추려낸 5개(SqMS32, SqMS33, SqMS36, SqMS43, SqMS45)의 조합 사용 시에도 국내산과 일본산의 구분이 가능하였다. 추가적인 방어 개체 모니터링을 통해 가장 효과적인 3개 조합을 추려낼 수 있을 것이다. 본 연구의 결과를 바탕으로 현장 진단 개발에 접목할 수 있을 것이다.

## 6. References


[1] Fuji, K. *et al.* Construction of a high-coverage bacterial artificial chromosome library and comprehensive genetic linkage map of yellowtail Seriola quinqueradiata. *BMC Res. Notes***7**, (2014).

[2] 통계청 2020년 어류양식동향조사결과

[3] Grover, A. & Sharma, P. C. Development and use of molecular markers: Past and present. *Critical Reviews in Biotechnology* vol. 36 (2016).

[4] 국립수산물품질관리원 수산물 원산지 표시 위반 현황 (2015~2019)

[5] 국립수산과학원 주요양식 기술 어류 방어

[6] Park, H. S. *et al.* Mitochondrial plastid DNA can cause DNA barcoding paradox in plants. *Sci. Rep.***10**, (2020).

[7] Baek, J.-H., Kim, D.-W., Kim, J. & Lee, T.-H. Improvement of SNPs detection efficient by reuse of sequences in Genotyping By Sequencing technology. *J. Korea Inst. Inf. Commun. Eng.***19**, (2015).

[8] Kim, C.-S., Ko, J.-M., Cha, H.-C., Park, J. K. & Jeong, J. PCR Technique for Determining Jeju Black Cattle, Hanwoo and Imported Beef. *J. Life Sci.***24**, (2014).

[9] Smouse, P. E. How many SNPs are enough? *Molecular Ecology* vol. 19 (2010).

[10] Yoo, D. A., Kim, K., Kim, D. K. & Cho, S. Effects of DNA Degradation on the Genotyping-by-Sequencing (GBS) Data for Genetic Marker Discovery of Hanwoo Population. *Ann. Anim. Resour. Sci.***28**, (2017).

[11] Xu, S. Y., Song, N., Xiao, S. J. & Gao, T. X. Whole genome survey analysis and microsatellite motif identification of Sebastiscus marmoratus. *Bioscience Reports* vol. 40 (2020).

[12] Glaubitz, J. C. *et al.* TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One***9**, (2014).

[13] Costa-Silva, J., Domingues, D. & Lopes, F. M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* vol. 12 (2017).

[14] Yun, Y. E., Yu, J. N., Lee, B. Y. & Kwak, M. H. An Introduction to Microsatellite Development and Analysis. *Korean J. Plant Taxon.***41**, (2011).