

Development and Verification of and Single Nucleotide Polymorphism Markers to Determine Country of Origin of Korean and Chinese *Scapharca subcrenata*

Seong Seok Choi^{1†}, Seung Hyun Yoo^{2†}, Yong Bae Seo¹, Jong Oh Kim², Ik Jung Kwon², So Hee Bae² and Gun Do Kim^{2*}

¹Basic Science Research Institute, Pukyong National University, Busan 48513, Korea

²Department of Microbiology, Pukyong National University, Busan 48513, Korea

Received November 7, 2023 /Revised December 4, 2023 /Accepted December 4, 2023

In this study, we analyzed SNPs that appear between Korean and Chinese *Scapharca subcrenata* using the nucleotide sequence data of *S. subcrenata* analyzed by genotyping by sequencing (GBS). To distinguish the country of origin for *S. subcrenata* in Korean and Chinese, we developed a primer set as single nucleotide polymorphism (SNP) markers for quantitative real-time PCR (qPCR) analysis and validated by sequencing SNPs. A total of 180 samples of *S. subcrenata* were analyzed by genotyping by sequencing, and 15 candidate SNPs were selected. SNP marker selection for country of origin were identified through real-time qPCR. Insertion 1 and SNP 21 markers showed the most distinct separation between the sequence types as well as the country of origin through qPCR, with the observed amplification patterns matching the expected outcomes. Additionally, in a blind test conducted by mixing samples of *S. subcrenata* at random, Insertion 1 showed 74% accuracy, 52% sensitivity, and 96% specificity, and SNP 21 showed 86% accuracy, 79% sensitivity, and 93% specificity. Therefore, the two SNP markers developed are expected to be useful in verifying the authenticity of the country of origin of *S. subcrenata* when used independently or in combination.

Key words : Country of origin, genotyping by sequencing, real-time qPCR, *Scapharca subcrenata*, single-nucleotide polymorphism markers

서 론

새꼬막(*Scapharaca subcrenata*)은 돌조개목(*Arcoida*), 돌조개과(*Arcidae*), 피조개속(*Scapharca*)에 속하는 부유식 이매패류로 한국, 중국 및 인근 지역에서 연안의 조간대부터 수심 10 m 사이의 얕은 조하대 사이의 진흙 퇴적물에 서식한다[15]. 새꼬막은 한국, 중국 등과 같은 아시아 국가의 중요한 어업 자원으로[17], 우리나라에서는 연근해어업과 해면양식업을 통해 생산된다. 해양 환경 변화와 남획, 조기채취 및 판매로 인한 새꼬막 자원의 감소 등의 원인으로 새꼬막의 연간 총 생산량은 2019년 10,500톤 이후 점차 감소하여 2020년 7,433톤, 2021년 5,296톤이 생산되었다. 이에 따라 국내 새꼬막 수요를 충족시키기 위해 수입량은 2019년 141톤에서 2021년 304톤으로 두 배 이상

증가하였다.

한국에서는 수산물의 수입량이 지속적으로 증가하고 있으며, 이에 대한 부작용으로 원산지 표시 위반 사례도 증가하고 있다. 2020년 해양수산부에서 제공한 ‘수산물 원산지 표시 위반 현황’에 따르면 2015년부터 2019년까지 적발된 원산지 표시 위반 업체는 모두 3,926개이며, 2015년 769건에서 2019년 915건으로 적발 건수가 꾸준히 증가하였다. 그 중 원산지 국가를 허위로 표기하는 거짓 표시는 1,007건으로 국가별로는 중국산(39.8%), 일본산(15.8%), 원양산(7.2%), 러시아산(6.8%), 국내산(3.4%) 순으로 집계되었다. 이처럼 수산물의 수입과 소비가 많은 우리나라에서 수입 수산물의 품질 보증과 안전성을 위해 원산지에 대한 정보가 명확히 표시되어야 하며 그 진위를 객관적으로 확인할 수 있는 방법이 필요한 실정이다.

국내에서는 수산물을 포함하여 다양한 농산물과 축산물의 원산지를 판별하기 위해 DNA를 활용한 분자생물학적 마커 개발 연구가 진행되고 있다. 당귀의 경우 Random amplified polymorphic DNA (RAPD) 마커를 이용한 중간 구별을 통해 국내산과 수입산을 판별하였다[1]. 동일한 종내의 원산지 판별 사례로는 한우 MS 마커[10], 참돔 단백질 마커 및 MS 마커[13], 바지락 MS 및 SNP 마커[9], 대게 MS 마커 유전자 키트[8] 등이 보고되었다. 새꼬막의 경우

[†]Authors contributed equally.

*Corresponding author

Tel : +82-51-629-5618, Fax : +82-51-629-5619

E-mail : gundokim@pknu.ac.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.

개체군의 구조를 파악하고 유전적 보존 연구를 위하여 SNP 마커가 이용되었지만[17] 원산지를 판별하기 위한 분자생물학적 마커 개발에 대한 연구는 아직 미미한 수준이다. 특히, 새꼬막은 산지에서 가공 공정 단계에서 패각을 제거하여 수입되는 경우도 있기 때문에 육안으로 구분이 쉽지 않으며, 이를 확인하기 위해 분자생물학적 마커 개발이 시급한 실정이다.

Single-nucleotide polymorphism (SNP)은 가장 작은 유전자 변이 단위로 개체 간의 특정 위치에서 이중 대립 유전자 대체 염기(biallelic alternative bases) 또는 하나 내지는 소수의 nucleotides 삽입이나 삭제로 발생하는 변이이며 거의 무제한으로 발생한다[6]. 기존 DNA 마커보다 SNP 마커의 polymorphic information content (PIC)가 높고 공동 우성을 나타내기 때문에 다양한 개체군의 유전적 다양성과 그 구조를 분석하는데 사용한다[3, 14]. SNP를 이용하여 개발한 마커는 분자생물학적 분석 방법인 PCR 또는 real-time PCR법으로 활용 가능하며, 주로 Amplifluor, Invader, Taqman 법 등 형광 염료를 사용한 실시간 또는 정량적으로 측정하는 형광 기반 분석법을 이용하면 매우 빠른 속도와 높은 정확도로 개별 유전자형을 분석할 수 있다[4, 5, 12]. 또한, 급속도로 발전한 next generation sequencing (NGS) 기술에 의해 비모델 생물체에서 대규모 SNP 탐색이 가능하며, 집단간 차이가 나타나는 SNP를 활용한 마커 개발에 이용되고 있다[18].

본 연구에서는 우리나라와 중국 새꼬막의 원산지 판별 방법을 개발하기 위해 NGS 기법 중 하나인 genotyping by sequencing (GBS) 기술을 활용하여 한국과 중국 새꼬막 사이에 나타나는 SNP를 분석하였다. 반응 특이성과 SNP 유전형 분석의 판별력을 높이기 위해, 분석된 SNP들이 3' 말단이 되도록 각각의 프라이머를 디자인하였으며[7], 제작한 프라이머는 quantitative real-time PCR (qPCR)을 통해 원산지 판별 마커로써 효과를 검증하였다.

재료 및 방법

새꼬막 시료 채집 및 genomic DNA 추출

국내산 및 중국산 새꼬막 시료는 국가별로 세 지역에서 30개체씩 채집하여 90개체의 국내산 새꼬막과 90개체의

중국산 새꼬막을 실험에 사용하였다. 국내산 및 중국산 새꼬막에 대한 정보는 Table 1에 나타났다. GBS 분석을 위한 genomic DNA는 새꼬막의 근육 조직으로부터 AccuPrep® Genomic DNA Extraction Kit (Bioneer, Daejeon, Korea)를 이용하여 추출하였다. 추출한 genomic DNA는 NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific™, MA, USA)를 사용하여 260 nm와 280 nm 흡광도에서 순도를 측정하고 정량하였다.

Genotyping-By-Sequencing 분석

GBS library를 제작하기 위하여 추출한 새꼬막 gDNA는 *Pst*I과 *Msp*I 제한효소로 처리한 후 잘린 각 DNA 조각에 PCR primer 서열을 포함하는 adaptor 서열과 개체 식별을 위한 4-6 bp의 barcode 서열을 순서대로 연결하였다. 구축된 새꼬막 GBS library는 NovaSeq 6000 system (Illumina, CA, USA)으로 NGS를 수행하였다. NGS 데이터에서 trimming으로 adaptor 서열을 제거한 다음 개체를 식별하고 barcode 서열과 효소 서열 제거하는 demultiplexing 과정을 진행하여 고품질의 서열을 분류하였다. 분리된 서열은 De novo clustering과 standard 서열에 대한 mapping을 통해 국내산과 중국산 새꼬막 개체에 존재하는 SNP 서열을 식별하였다.

SNP 마커 후보의 선별은 표현형과 유전자형의 연관성을 통계학적으로 분석하는 Trait Analysis by Association, Evolution and Linkage (TASSEL) [2] 프로그램을 이용하여 한국과 중국 새꼬막 사이에 존재하는 SNP 서열에 대한 특성을 계산하고, GBS-SNP-CROP (The GBS SNP Calling Reference Optional Pipeline) 소프트웨어를 이용하여 분석한 GBS 데이터에서 한국산과 중국산 사이에 차이를 보이는 SNP 중 신뢰도가 높은 SNP를 선별하였다. 이 결과를 바탕으로 한국산과 중국산 그룹에 대한 유의적인 차이를 확인하기 위해 두 그룹간 통계 분석에서 최상위의 *p*-value를 가지면서 동일한 flanking 영역 내에서 서로 다른 SNP type을 가지고 PCR 증폭 결과 product size가 큰 순으로 후보 SNP 마커를 선별하였다.

Conventional PCR 및 genotype 분석

선별된 후보 SNP 마커를 스크리닝하기 위해 PCR을 이

Table 1. The information of *S. subcrenata* samples

Country	Group	Region	Collection location	Collection date	Number of individuals
Korea (KR)	KR_A (01-30)	Gunsan,, Jeonbuk	35°27'49.9"N 126°20'30.4"E	2021.01.28	30
	KR_B (31-60)	Beolgyo, Jeonnam	34°46'00.1"N 127°28'44.5"E	2021.02.10	30
	KR_C (61-90)	Hampyeong, Jeonnam	35°06'34.7"N 126°22'20.5"E	2021.03.19	30
China (CN)	CN_D (01-30)	Nantong, Jiangsu	32°30'15.1"N 121°45'25.1"E	2021.03.17	30
	CN_E (31-60)	Tangshan, Hebei	39°04'57.0"N 119°16'12.4"E	2021.04.07	30
	CN_F (61-90)	Dongguan, Guangdong	22°15'01.9"N 114°30'42.2"E	2021.04.07	30

용하여 각 SNP 서열을 증폭시켰다. 본 연구에 사용된 모든 primer는 Primer3Plus 프로그램[16]으로 디자인하였으며, conventional PCR에 사용된 primer 서열은 Table 2에 나타냈다. 반응 혼합물은 10 µl의 AccuPower® PCR PreMix (Bioneer, Daejeon, Korea)에 1 µl의 genomic DNA (10 ng/µl), 1 µl의 forward primer (10 pmol), 1 µl의 reverse primer (10 pmol), 그리고 최종 부피가 20 µl가 되도록 증류수를 첨가하여 제작하였다. PCR은 AllInOneCycler™ PCR system (Bioneer, Daejeon, Korea)을 사용하였으며, denaturation은 94°C에서 30초, annealing 60°C에서 30초, extension은 72°C에서 30초로 35 cycle진행하였다. 증폭된 PCR products는 1.5% agarose gel에서 150V로 30분간 전기영동하여 확인하였다. 증폭이 확인된 PCR products는 AccuPrep® PCR/Gel Purification Kit (Bioneer, Daejeon, Korea)을 이용하여 분리 및 정제하였다. 정제된 PCR product는 NanoDrop™

2000 Spectrophotometer (Thermo Fisher Scientific™, MA, USA)로 260 nm, 280 nm에서 흡광도를 측정하여 순도와 농도를 확인하였다. 분리한 PCR products는 유전자형 분석을 위해 ABI3730XL DNA Analyzer (Applied Biosystems, MA, USA)으로 sanger sequencing을 진행하였다. 제공받은 분석 결과는 Molecular Evolutionary Genetics Analysis (MEGA) [11] 프로그램으로 유전자형을 분석하여 SNP 영역을 확인하였다.

SYBR green dye 염색을 통한 quantitative real-time PCR

선별된 SNP 마커는 원산지 간 개체의 분리 여부를 확인하기 위해 qPCR을 진행하였다. Training step에서는 sanger sequencing에 의해 SNP type이 확인된 genomic DNA sample을 사용하여 primer의 특이도를 확인하였다. Validation

Table 2. The information of primers used in conventional PCR for SNP markers screening

SNP marker ID	Marker name	Primer name	Sequence (5'→3')	Tm (°C)	Product Size (bp)
MOCKREFGENOME_28429426	Insertion 1	Shell_SNP_I_001_F	TGCGTGAATGCATTGGGATG	59.543	217
		Shell_SNP_I_001_R	CGGATATGGAGAACGAGGCT	59.038	
MOCKREFGENOME_32022891	Insertion 2	Shell_SNP_I_002_F	ACAGCCCTTCGGAACATTTCT	59.926	113
		Shell_SNP_I_002_R	AAGACGAACTTACAGGGCGC	60.670	
MOCKREFGENOME_15784924	Insertion 3	Shell_SNP_I_003_F	TTCACAATGGCATGCTGTCA	58.382	252
		Shell_SNP_I_003_R	CACTTACAACGGCAAGGCG	59.762	
MOCKREFGENOME_12882585	Insertion 4	Shell_SNP_I_004_F	ACTGTCAACGCTCACGATGT	59.968	172
		Shell_SNP_I_004_R	GAGCTCGATTAAGCCCGTGA	59.898	
MOCKREFGENOME_28644452	Insertion 5	Shell_SNP_I_005_F	TGCCATATGTTATGATGTGCAGC	59.746	184
		Shell_SNP_I_005_R	ACTTCGTACAACACACGGCA	60.179	
MOCKREFGENOME_20589633	Insertion 6	Shell_SNP_I_006_F	ACCCCTCCCCAGTCCATTAT	59.649	171
		Shell_SNP_I_006_R	CTTCAAATTAGTGGCCGCGG	59.901	
MOCKREFGENOME_35590589	Deletion 12	Shell_SNP_D_012_F	CGCAGCAGAACTCATTGTGG	59.833	197
		Shell_SNP_D_012_R	CGGATTGTGCCACGTGTATG	59.627	
MOCKREFGENOME_24193176	Deletion 13	Shell_SNP_D_013_F	TTGTACAACGAAGCAGGCTG	58.774	119
		Shell_SNP_D_013_R	TCGACAATACTGTGATGACGGT	59.509	
MOCKREFGENOME_26032670	Deletion 14	Shell_SNP_D_014_F	TGTGGAGTAACTGGAATGCCA	59.893	120
		Shell_SNP_D_014_R	GCCATGAGAACTCCCTGCA	60.323	
MOCKREFGENOME_37425568	Deletion 15	Shell_SNP_D_015_F	TGGCTGAAAACTGGCATGC	59.966	88
		Shell_SNP_D_015_R	CATGCATGCACACTCGCTAT	59.058	
MOCKREFGENOME_42188563	SNP 21	Shell_SNP_021_F	CAACGCTTGTGAGCCTTGTC	60.041	190
		Shell_SNP_021_R	ACGTAACGAGAACATGGCGT	59.999	
MOCKREFGENOME_28437068	SNP 22	Shell_SNP_022_F	TGCTTGAACGATTTGCAACGA	59.668	184
		Shell_SNP_022_R	AAATCCCTGGGTACAGCGAC	59.748	
MOCKREFGENOME_28437137	SNP 23	Shell_SNP_023_F	CGAGTTCGTGACGGCAAAAC	60.383	174
		Shell_SNP_023_R	CGGGATGGTAACCCTCTCTT	58.503	
MOCKREFGENOME_33531338	SNP 24	Shell_SNP_024_F	AATCGTTCCTCAGCAGCTC	60.108	251
		Shell_SNP_024_R	CCTGGCTTGATTGGATGTTTGG	60.094	
MOCKREFGENOME_35920482	SNP 25	Shell_SNP_025_F	TGGTTGGGGTGTCAATCCTG	59.888	108
		Shell_SNP_025_R	ATCTCCAAACACGAAGCGA	59.680	

step에서는 각 primer 별로 SNP type이 판별된 genomic DNA sample을 사용하였고 validation test 결과를 바탕으로 원산지 판별에 필요한 기준 Ct value를 선정하였다. 마지막으로 training과 validation step에 사용하지 않은 개체들을 무작위로 섞어 원산지 판별 모델의 정확성을 평가하기 위한 blind test를 진행하였다.

반응 혼합물은 AccuPower® 2xGreenStar™ qPCR Master Mix (Bioneer, Daejeon, Korea) 10 µl, forward primer (10 pmol) 1 µl, reverse primer (10 pmol) 1 µl, genomic DNA (10 ng/µl) 1 µl, 그리고 최종 부피가 20 µl가 되도록 증류수를 첨가하여 제작하였으며, 반응에 사용한 primer는 Table 3에 나타내었다. qPCR은 QuantStudio™ 6 Flex System (Thermo Fisher Scientific™, MA, USA)를 이용하였으며, 반응 조건은 95°C에서 10분간 반응 후, 95°C 15초, 62°C에서 30초, 72°C에서 30초로 40 cycle 진행하였다. Melt curve analysis는 qPCR reaction mixture를 95°C에서 15초, 60°C에서 1분간 반응시키고, 60°C에서 95°C까지 0.05°C/s (0.05°C every seconds) 속도로 올리면서 측정하였다. 민감도 analysis를 위해 primer 별로 서로 다른 SNP type을 가지는 4개의 genomic DNA (10 ng/µl)를 각각 10~0.0001 ng/µl로 희석하여 qPCR을 진행하였다.

Quantitative real-time PCR에 사용된 primer의 검출한계 및 효율 검증

SYBR green qPCR assay를 위해 개발된 primer들의 증폭 효율성과 limit of detection (LOD)를 평가하기 위하여 원산지 및 SNP type 별로 1개체씩 총 4개의 *S. subcrenata* genomic DNA를 tenfold into six series (10~0.0001 ng/µl)로 희석하여 qPCR을 진행하였다. Standard curve는 각각의 primer가 표적으로 하는 SNP sequence의 homo-type인 *S. subcrenata* 샘플의 DNA 10~0.01 ng/µl 범위 내에서 얻었다. Correlation coefficient (R^2)은 linear regression method를 사용하여 결정되었다. 증폭 효율은 standard curve의 기울기를 기반으로 다음 방정식을 사용하여 계산하였다:

$$E (\%) = (10^{(-1/\text{slope})} - 1) \times 100$$

LOD는 95% 확률로 검출할 수 있는(위음성 결과가 5%

미만인) 분석 물질의 최소량으로 정의된다.

결 과

Genotyping-By-Sequencing 분석을 통한 SNP 마커 스크리닝 및 선별

한국산과 중국산 *S. subcrenata*의 180개의 근육 조직으로부터 genomic DNA를 추출한 후 GBS 분석을 진행한 결과에서 9,753개의 SNP 염기서열을 확보하였으며, SNP data의 수는 6,367개, Deletion/Insertion Polymorphisms (DIPs) data의 수는 3,386개였다. *S. subcrenata* 원산지 판별을 위한 후보 SNP 마커는 *p*-value가 낮고, 한국산과 중국산의 SNP type이 다르며, 후보군 primer 기준 product length가 긴 data 순으로 선정하였다. 그 결과 후보 SNP 마커는 polymorphisms type 별로 각각 insertion 6개, deletion 4개, SNP 5개로 총 15개를 선별하였다(Table 4). 후보 SNP 마커의 초기 스크리닝을 위하여 SNP 영역을 확인할 수 있는 primer들을 디자인하고 *S. subcrenata*의 genomic DNA에 대한 PCR 및 sanger sequencing을 진행하였다.

SNP 마커 결정 및 primer 제작

제공받은 염기서열은 multiple alignment를 통해 GBS data에서 확보한 SNP 영역에서의 variation을 확인하였다. Sequence alignment 결과와 GBS data를 비교하여 유의미한 원산지 구분이 가능할 것이라 판단되는 후보 SNP 마커로 Insertion 1, Deletion 12, SNP 21을 선정하였다.

Insertion 1은 6개의 개체에서 GBS data 결과의 insertion 영역 외에 두 개의 SNP 후보 영역을 확인하였다(Fig. 1A). 첫 번째 위치는 염기서열이 A 또는 C로 존재하는 영역이며, 두 번째 위치는 TT/AT insertion 또는 insertion이 일어나지 않은 영역으로 나타났다. Insertion 1은 국내산 37개, 중국산 37개의 개체의 염기서열을 추가로 분석한 결과 80개 모두 두 SNP 영역이 동일한 패턴을 가지고 있음을 확인하였으며, 그 중 첫 번째 SNP 위치에 대하여 개체를 구분하면 국내산의 경우 AA의 homo-type이 18개, AC hetero-type이 18개, CC homo-type이 4개인 반면, 중국산의 경우 모든 개체가 AA homo-type인 것으로 나타났다(Table

Table 3. The information of primers for qPCR with SYBR green dye staining

Marker name	Primer name	Sequence (5'→3')	Tm (°C)
Insertion 1	KRS001_F	GCAGAATGCTTACAGAAAGAC	54.3
	Shell_SNP_I_001_R	CGGATATGGAGAACGAGGCT	59.038
Deletion 12	KRS012_F	GATGGACTGACGGATGTACT	55.5
	Shell_SNP_D_012_R	CGGATTGTGCCACGTGTATG	59.627
SNP 21	KRS021_F	AGAACTGCTTCAGAAATTAGTTTG	54.9
	Shell_SNP_021_R	ACGTAACGAGAACATGGCGT	59.999

Underlined nucleotides in primer sequences indicate SNP sequence.

Table 4. 15 candidate SNP markers based on GBS analysis

SNP marker ID	Marker name	Sequence (5'→3')	T _m (°C)	p-value	SNP type		Product size (bp)
					Korean allele/	Chinese allele	
MOCKREFGENOME_28429426	Insertion 1	F: TGC GTGAATGCATTGGGATG R: CGGATATGGAGAACGAGGCT	59.543 59.038	1.66E-26	AATAAT/AA		217
MOCKREFGENOME_32022891	Insertion 2	F: ACAGCCCTTCGGAACATTTCT R: AAGACGAACTTACAGGGCCG	59.926 60.670	5.01E-06	CCACAGT/CC		113
MOCKREFGENOME_15784924	Insertion 3	F: TTCACAAATGGCATGCTGTCA R: CACTTACAACCTGGCAAGGGC	58.382 59.762	8.03E-05	CTCT/CTC		252
MOCKREFGENOME_12882585	Insertion 4	F: ACTGTCAAACGCTCACGATGT R: GAGCTCGATTAAGCCCGTGA	59.968 59.898	1.91E-04	TTA/TT		172
MOCKREFGENOME_28644452	Insertion 5	F: TGCCATATGTTATGATGTGCAGC R: ACTTCGTACAACACACGGCA	59.746 60.179	6.42E-04	TTA/TT		184
MOCKREFGENOME_20589633	Insertion 6	F: ACCCTCCTCCAGTCCATTAT R: CTTCAAATTAGTGCCCGCGG	59.649 59.901	9.96E-04	AA/AAAAG		171
MOCKREFGENOME_35590589	Deletion 12	F: CGCAGCAGAACTCATTGTGG R: CGGATTGTGCCACGTGTATG	59.833 59.627	2.10E-04	AGAG/AGG		197
MOCKREFGENOME_24193176	Deletion 13	F: TTGTACAACGAAAGCAGGCTG R: TCGACAATACTGTGATGACGGT	58.774 59.509	2.99E-04	TT/GATATTGATATT		119
MOCKREFGENOME_26032670	Deletion 14	F: TGTGGAGTAAACTGGAATGCCA R: GCCATGAGAAACTCCCTGCA	59.893 60.323	0.02937	AA/AAAAAA		120
MOCKREFGENOME_37425568	Deletion 15	F: TGGCTGAAAAACTGGCATGC R: CATGCATGCACACTCGCTAT	59.966 59.058	0.048	TCC/TCTC		88
MOCKREFGENOME_42188563	SNP 21	F: CAACGCTTGTACGCCTTGTC R: ACGTAACTGAGAACATGGCGT	60.041 59.999	1.75E-31	CT/CC		190
MOCKREFGENOME_28437068	SNP 22	F: TGGTTGAACGATTTGCAACGA R: AAATCCCTGGGTACAGCGAC	59.668 59.748	7.45E-25	GA/GG		184
MOCKREFGENOME_28437137	SNP 23	F: CGAGTTCGTGACGGCAAAAAC R: CGGGATGGTAACCCCTCTCTT	60.383 58.503	3.36E-23	CT/CC		174
MOCKREFGENOME_33531338	SNP 24	F: AATCGGTTCCCTCAGCAGCTC R: CCTGGCTTGATTTGGATGTTGG	60.108 60.094	4.77E-23	AC/AA		251
MOCKREFGENOME_35920482	SNP 25	F: TGGTTGGGGTGTCAATCCTG R: ATCTCCAAAACACGAAAGCGA	59.888 59.680	7.32E-23	GA/GG		108

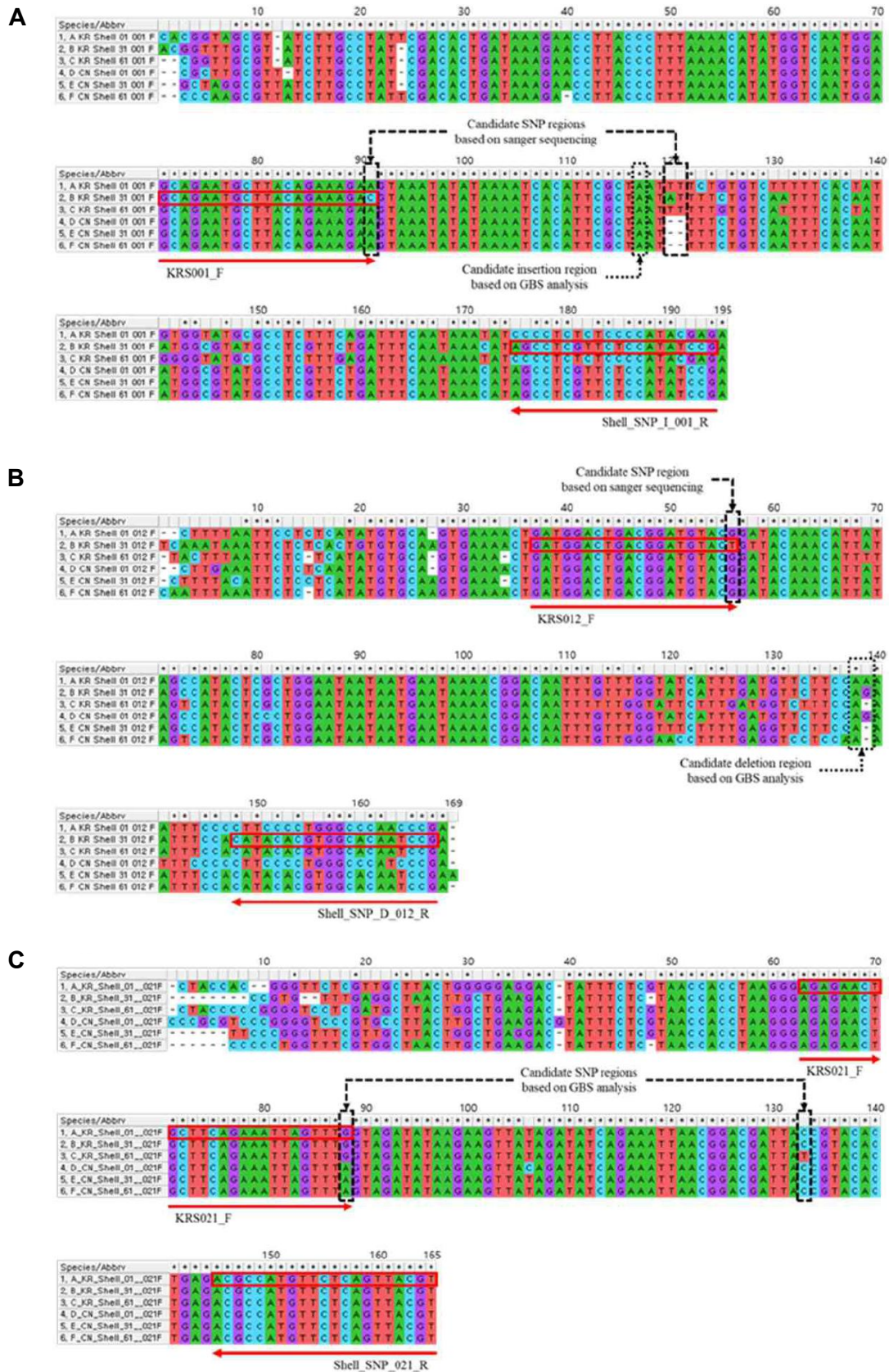


Fig. 1. Multi alignments of candidate SNP marker regions of *S. subcrenata* in Korean and Chinese and primers designed for qPCR. (A) Insertion 1; (B) Deletion 12; (C) SNP 21.

5). 원산지 판별 정확도를 높이기 위하여 첫 번째 SNP 영역을 분석 대상으로 선택하였고, KRS001_F primer는 중국산이 다수를 차지하는 AA homo-type의 증폭이 이루어지지 않게 하기 위하여 primer의 마지막 서열이 SNP type 중 하나인 cytosine이 되도록 디자인하였다.

Deletion 12는 6개의 개체에서 GBS analysis 결과에서 확인한 deletion 영역 외에 하나의 후보 SNP 영역을 확인하였다(Fig. 1B). Deletion 영역은 AA 또는 AG 또는 A type이 존재하며, 국내산 36개, 중국산 36개의 개체를 추가로 분석한 결과 국내산과 중국산 모두 SNP type에 대해 비슷한 분포를 보였다(data not shown). SNP 영역은 G 또는 T로 존재하는 영역으로, 추가 염기서열 분석에서 78개의 국내산/중국산 새꼬막은 GG homo-type이 12개/36개, GT hetero-type이 23개/3개, TT homo-type이 4개/0개로 나타났다(Table 5). 이 결과를 바탕으로 Deletion 12는 새꼬막 원산지 구분을 위하여 G/T SNP sequence를 분석 대상으로 선택하고 대다수의 중국산이 포함되는 GG homo-type이 증폭되지 않도록 KRS012_F primer 끝에 thymine을 붙여 설계하였다.

SNP 21은 GBS data에서 확보한 두 개의 SNP 영역이 sequencing 결과에서도 동일하게 나타났다(Fig. 1C). 첫 번째 SNP 영역은 G 또는 A로 존재하며 두 번째 위치는 T 또는 C로 존재하는 것으로 나타났다. 이 중 첫 번째 SNP

서열에 대해 국내산 21개, 중국산 16개를 추가로 분석한 결과 24개의 국내산 새꼬막은 GG homo-type, GA hetero-type, AA homo-type이 각각 9개, 12개, 3개로 나타났다. 반면 19개의 중국산 새꼬막은 GG homo-type은 존재하지 않았으며 GA hetero-type 이 1개, 그리고 나머지 18개는 AA homo-type인 것으로 나타났다(Table 5). 따라서 SNP 21의 G/A SNP 영역에서 중국산 대다수가 포함되는 AA homo-type의 증폭을 막기 위하여 KRS021_F primer 서열 끝에 guanine이 포함되도록 제작하였다.

SNP 마커 training step 결과

qPCR 실험을 위해 디자인한 primer가 특이성을 가지고 효과적으로 원산지를 구분할 수 있는지 확인하기 위하여 sanger sequencing으로 SNP type이 확인된 한국산 *S. subcrenata* 3개(KR01, 31, 61)와 중국산 *S. subcrenata* 3개(CN01, 31, 61)를 training set로 선정하고 Table 3에 표기한 primer set를 사용하여 qPCR을 진행하였다. Table 6에서 볼 수 있듯이 KRS001_F와 KRS021_F primer를 사용한 Insertion 1과 SNP 21의 경우 한국산 *S. subcrenata*은 Ct value가 모두 25 이하로 나타났고 중국산은 모두 30 이상의 값을 보여 두 primer 모두 sanger sequencing으로 확인하였을 때 예상한 qPCR 증폭 여부와 실제 증폭 여부가 일치하였다. 하지만 KRS012_F primer를 사용한 Deletion 12는

Table 5. Classification of *S. subcrenata* by SNP types based on sanger sequencing results

Origin		AA	AC	CC	Total
Insertion 1	KR	02, 32, 33, 35, 37-44, 64-66, 68, 71, 72	01, 04, 05, 06, 07, 08, 09, 10, 11, 34, 36, 45, 61, 62, 63, 67, 69, 70	03, 12, 31, 46	40 samples
	CN	18 samples 01-13, 31-43, 61-74 40 samples	18 samples - 0 sample	4 samples - 0 sample	40 samples
Origin		GG	GT	TT	Total
Deletion12	KR	03, 09, 38, 40, 42, 43, 61, 62, 66, 67, 71, 72	01, 02, 04-08, 10, 11, 12, 32-34, 37, 39, 41, 45, 46, 63, 65, 68-70	31, 35, 36, 44	39 samples
	CN	12 samples 01-07, 09-13, 31-43, 61-63, 65, 68-74 36 samples	23 samples 08, 64, 66 3 samples	4 samples - 0 sample	39 samples
Origin		GG	GA	AA	Total
SNP 21	KR	03, 04, 32, 41, 42, 43, 61, 65, 71	01, 02, 05, 09, 31, 33, 35, 37, 38, 39, 40, 64	44, 66, 68	24 samples
	CN	9 samples - 0 sample	12 samples 64 1 sample	3 samples 01, 05, 06, 07, 09, 10, 31, 32, 33, 34, 35, 36, 61, 62, 63, 65, 66, 67 18 samples	19 samples

Table 6. Ct mean from training set of SNP markers

Marker name		KR01	KR31	KR61	CN01	CN31	CN61
Insertion 1	SNP type	AC	CC	AC	AA	AA	AA
	Ct mean ± SD	24.146±0.221	21.857±0.128	23.260±0.230	31.935±0.808	30.492±0.120	31.813±0.381
Deletion 12	SNP type	GT	TT	GG	GG	GG	GG
	Ct mean ± SD	23.559±0.066	23.426±0.065	24.355±0.015	24.352±0.022	24.603±0.051	24.439±0.036
SNP 21	SNP type	GA	GA	GG	AA	AA	AA
	Ct mean ± SD	24.499±0.137	22.672±1.306	22.80±0.282	30.515±0.279	30.714±0.194	31.366±1.325

SNP types in bold should be amplified.

Ct mean이 모두 25 이하로 나타나 원산지 판별이 불가능하였다.

Primer의 검출 민감성 및 증폭 효율 검증

설계한 primer들의 검출 민감성을 평가하기 위하여 training step에서 원산지 구분이 불가능하다고 판단된 Deletion 12를 제외하고 Insertion 1과 SNP 21에 대하여 원산지 및 SNP type별로 1 개씩 4개의 genomic DNA를 10~0.0001 ng/μl 농도로 희석하여 qPCR을 수행하였다(Fig 2). KRS001_F primer가 표적으로 한 CC homo-type (KR03)과 KRS021_F primer가 표적으로 한 GG homo-type (KR03)의 10~0.01 ng/μl DNA 농도 범위 내에서 얻은 standard curve 들 모두 R2 값이 0.99 이상으로 높은 상관관계를 나타냈다. 두 Standard curve의 slope는 KRS001_F의 경우 -3.731, KRS021_F는 -3.318이며, 이를 바탕으로 계산한 증폭 효율은 각각 85.36%와 100.14%이다. LOD는 0.660 ng/μl (KRS001_F), 0.431 ng/μl (KRS021_F)이며 이 DNA 농도에서의 예측 Ct 값은 29.181(KRS001_F), 29.180(KRS021_F) 이었다. 표적 대립 유전자를 하나씩 포함하는 AC hetero-type (KR01)과 GA hetero-type (KR01) 또한 각각의 homo-type과 비슷한 증폭 양상을 보였다. 하지만 두 primer 모두 표적 대립 유전자가 없는 TT homo-type (CN01)의 경우 모든 DNA 농도에서 30 이상의 Ct 값을 보였다.

SNP 마커의 validation 및 blind test 결과

S. subcrenata 원산지 감별을 위해 개발된 KRS001_F와 KRS021_F primer들의 특이성을 검증하기 위해 sanger sequencing으로 SNP type이 판별된 *S. subcrenata* genomic DNA를 사용하여 validation test를 진행하였다(Fig. 3). Table 5를 참고하여 Insertion 1의 경우 한국산 40개, 중국산 40개의 *S. subcrenata*를 사용하였으며 SNP 21에 대해서는 한국산 24개, 중국산 19개의 *S. subcrenata*를 사용하였다. Fig. 3과 Table 5를 비교하면 Insertion 1의 경우 증폭을 예측한 AC hetero-type과 CC homo-type의 한국산 22개 샘플은 Ct 값이 26 이하로 나타났고, 증폭을 예상하지 않은 AA homo-type의 한국산 18개 샘플과 중국산 40개 샘플은 모두 29 이상의 Ct 값을 나타냈다. SNP 21 또한 증폭을 예상한 GG homo-type, GA hetero-type의 한국산 21개와 중국산 1개 샘플은 26 이하의 Ct 값을 나타내었고, 비증폭을 예상한 AA homo-type의 한국산 3개와 중국산 18개 샘플은 모두 Ct 값이 30 이상이었다. 두 마커의 primer 모두 validation test에서 사용된 모든 *S. subcrenata* 샘플의 표적 SNP sequence 존재 유무에 따른 qPCR 증폭 양상이 예상과 실제가 정확히 일치하였다. 또한 training step과 validation test, 검출 한계 분석 결과를 종합하여 *S. subcrenata*의 DNA 농도가 10 ng/μl일 때 효과적으로 원산지를 판별할 수 있도록 Insertion 1과 SNP 21의 기준 Ct 값을 28로 설정하였다.

앞선 실험으로 검증된 두 primer의 원산지 판별 예측

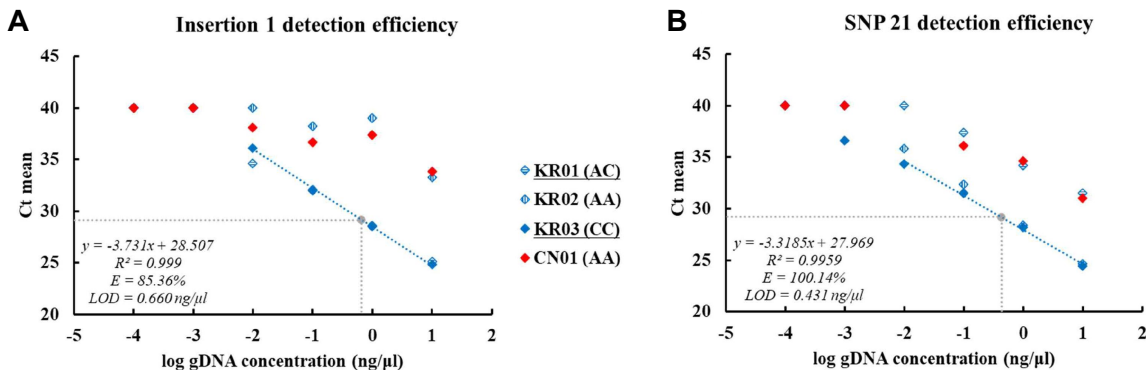


Fig. 2. Detection efficiency of each primer set according to sample concentration in qPCR. (A) Insertion 1; (B) SNP 21

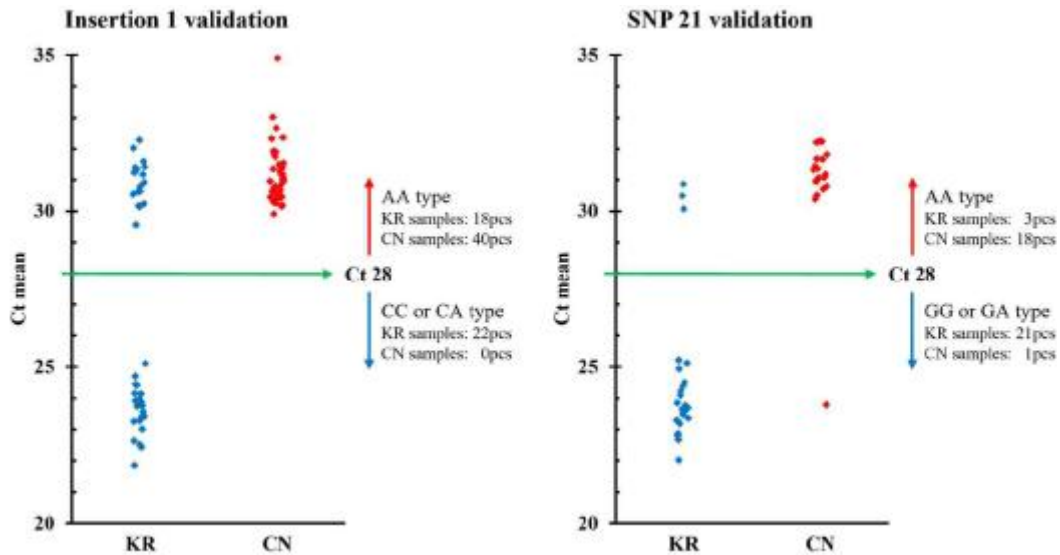


Fig. 3. Validation test results of Insertion 1 and SNP 21 markers with *S. subcrenata* samples from Korean and Chinese.

정확도를 확인하기 위한 방법으로 blind test를 진행하였다. Training과 validation set에 포함되지 않고 정확한 SNP type을 알 수 없는 한국산 및 중국산 *S. subcrenata*를 모두 무작위로 섞어서 qPCR을 수행하였다. Insertion 1에는 한국산 50개, 중국산 50개의 *S. subcrenata* gDNA가 포함되었으며 SNP 21에는 한국산 66개, 중국산 71개의 *S. subcrenata* gDNA가 포함되었다. Fig. 4를 보면 Insertion 1은 Ct 값 28을 기준으로 총 100개의 *S. subcrenata* 중에서 한국산으로 판별된 것은 28개, 중국산으로 판별된 것은 72개였다. Blind test 결과와 실제 원산지를 비교해보면 Insertion 1에 의해 한국산을 중국산으로 잘못 판별된 것이 24개, 중국산을 한국산으로 잘못 분류된 것은 2개였다. SNP 21 결과 또한 Ct mean 28을 기준으로 총 138개의 *S. subcrenata* 중에서 한국산으로 판별된 것은 57개, 중국산으로

판별된 것은 80개로, 이 중 한국산을 중국산으로 잘못 판별한 것이 14개, 중국산을 한국산으로 잘못 분리한 것이 5개로 나타났다.

고 찰

국내에서 새꼬막은 경제적으로 중요한 수산 자원이지만, 생산량 감소와 소비량 증가에 따라 수산물의 수입이 급속도로 증가하고 있다. 이에 따라 일부 중국산을 한국산으로 속여 유통하는 사례가 발생하였으며, 한국 소비자들의 안전한 수산물 먹거리에 대한 관심이 증가하였다. 하지만 새꼬막의 경우 육안으로 원산지 식별이 어렵고, 패각을 제거하는 가공품으로 수입되는 경우, 육안으로 판별이 거의 불가능하다. 따라서 이 문제를 해결하기 위해

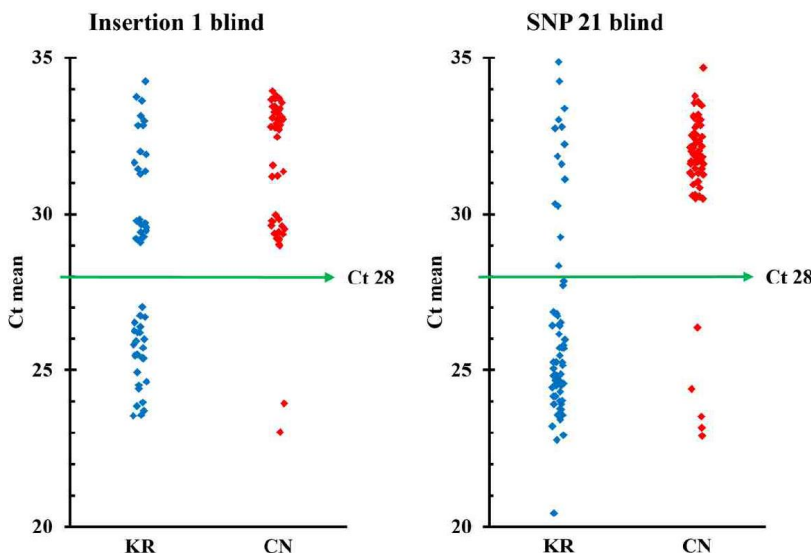


Fig. 4. Blind test results of Insertion 1 and SNP 21 markers with *Scapharca subcrenata* samples of Korean and Chinese.

서는 수산물 수입 과정에서 객관적인 새꼬막의 원산지 판별 기법이 필요하다.

본 연구에서는 국내산과 중국산 새꼬막의 원산지 판별에 활용 가능한 분자마커를 개발하기 위한 방법으로 GBS 분석에 기초하여 15개의 후보 SNP 마커를 선별하였다. SNP 마커는 PCR과 SYBR green-based qPCR 분석을 활용하여 국내산과 중국산 새꼬막 간의 증폭 형태를 비교함으로써 최종적으로 두 개의 SNP 마커를 개발하고 객관적인 원산지 판별 가능성을 검증하였다.

새꼬막에 대한 GBS 분석 및 sanger sequencing으로 1차 스크리닝 된 3개의 SNP 마커는 총 7개의 DIP 및 SNP 영역을 포함하였다. 그 중 SNP type에 의해 원산지가 명확하게 구분되고 training step에서 qPCR의 증폭 양상이 차이나는 SNP 영역은 Insertion 1의 C/A SNP 영역(KRS001_F)과 SNP 21의 G/A SNP 영역(KRS021_F)으로 나타났다. 이를 바탕으로 validation test를 위한 qPCR 분석을 진행한 결과 두 마커 모두 SNP type으로 예측한 qPCR 증폭 여부와 실제 Ct 값이 100% 일치하였다. 즉, 원산지 판별을 위해 새롭게 설계한 두 SNP primer는 SNP type에 따라 정확하게 새꼬막을 분리함을 시사한다. Insertion 1에서 국내산 새꼬막은 Ct 값이 22~26 또는 29~33 사이로 나타났으며 중국산 새꼬막은 29~35 사이의 Ct 값을 가지는 것으로 확인되었다. SNP 21에서는 국내산 새꼬막의 Ct 값이 22~26 또는 29~31 사이로 나타났으며 중국산 새꼬막의 Ct 값은 23~24와 30~33 사이로 나타났다. Ct 값 분석 결과를 토대로 qPCR에서 새꼬막의 원산지 판별 기준 Ct 값은 두 마커 모두 28로 설정하였다. 추가적으로 새꼬막을 원산지 구분 없이 무작위로 섞어서 진행한 blind test에서 Insertion 1은 새꼬막 100개에 대하여 정확도 74%, 민감도 52%, 특이도 96%로 국내산을 국내산으로 판별할 확률은 다소 낮았지만 중국산을 중국산으로 판별할 확률은 매우 높았다. SNP 21 또한 새꼬막 137개에 대하여 정확도 86%, 민감도 79%, 특이도 93%로 정확도와 민감도는 Insertion 1보다 높았고 특이도는 Insertion 1보다는 낮지만 세 지표 모두 상당히 높은 수준이었다.

SNP는 보존 유전학이나 집단 유전적 다양성 분석, 품종 판별 연구를 수행하기 위하여 자주 사용되는 유전자형 분석 방법이다. 새꼬막에서 SNP 마커는 높은 정확성으로 두 그룹을 구분할 수 있는 것으로 나타났다. 새롭게 개발된 두 개의 SNP 마커를 단독 또는 함께 활용하여 qPCR 방법으로 새꼬막의 SNP를 분석할 경우 국내산 및 중국산 새꼬막을 효과적으로 판별할 수 있을 것으로 예상된다. 그러나 원산지 판별을 위해서는 필수적으로 genomic DNA의 농도를 10 ng/μl 로 조절해야 한다는 점은 보완해야 할 사항이며, 원산지 판별의 정확성을 더 높이기 위한 추가적인 마커 개발 또는 새로운 분석 방법의 개발이 요구된다.

감사의 글

이 논문은 2023년도 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임(20200425, 수산생물 원산지 판별기술 및 현장단속 키트개발)

The Conflict of Interest Statement

The authors declare that they have no conflicts of interest with the contents of this article.

References

- Bang, K. H., Yu, H. S., Koo, D. H., Cho, J. H., Park, H. W., Seong, N. S., Park, S. I. and Kim, H. S. 2002. Selection of RAPD marker to discriminate the bolting-resistant varieties and commercial dried medicinal materials of *Angelica* species. *Kor. J. Medicinal Crop Sci.* **10**, 46-50.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.
- Guo, C., Wang, Y., Hu, W., Mei, J. and Guo, W. 2020. Development of SNP markers and validation 24 SNPs in darkbarbel catfish (*Pelteobagrus vachelli*). *Conserv Genet Resour.* **12**, 413-416.
- Gut, I. G. 2001. Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat.* **17**, 475-492.
- Gut, I. G. 2004. An Overview of Genotyping and Single Nucleotide Polymorphisms (SNP). *J. Wiley.*
- Johnson, G. C. L., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., di Genova, G., Ueda, H., Cordell, H. J., Eaves, I. A., Dudbridge, F., Twells, R. C. J., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S. C. L., Clayton, D. G. and Todd, J. A. 2001. Haplotype tagging for the identification of common disease genes. *Nat Genet.* **29**, 233-237.
- Kalendar, R., Shustov, A. V., Akhmetolayev, I. and Kairov, U. 2022. Designing allele-specific competitive-extension PCR-based assays for high-throughput genotyping and gene characterization. *Front Mol. Biosci.* **9**, 773956.
- Kang, J. H., Park, J. Y., Kim, E. M. and Ko, H. S. 2014. Population genetic analysis and origin discrimination of snow crab (*Chionoecetes opilio*) using microsatellite markers. *Mol. Biol Rep.* **40**, 5563-5571.
- Kim, E. M. 2014. Genetic diversity and population structure of manila clam, *Ruditapes philippinarum* based on SNP and MS DNA markers. Ph. D. dissertation, Pukyong National University, Busan. Korea.
- Kim, S., Cho, C. Y., Roh, H. J., Yeon, S. H. and Choi, S. B. 2017. Development of microsatellite markers for discriminating native Korean and imported cattle breeds. *J.*

Life Sci. **27**, 464-470.

11. Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549.
12. Kwok, P. Y. 2001. Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet.* **2**, 235-258.
13. Lee, Y. H., Kim, K. E., Jung, D. K., Kang, M. H., Kim, Y. A., Lee, Y. J. and Lee, Y. C. 2010. Development of origin and species identification techniques for fishery products. pp. 134-146. Korea Ocean Research and development Institute. Korea.
14. Nazari, S. and Pourkazmi, M. 2021. Isolation and characterization of SNP markers of rainbow trout (*Onchorhynchus mykiss* Walbaum, 1792) from transcriptomic sequences. *Mol Biol Rep.* **48**, 989-995.
15. Park, H. J., Lee, W. C., Choy, E. J., Choi, K. S. and Kang, C. K. 2011. Reproductive cycle and gross biochemical composition of the ark shell *Scapharca subcrenata* (Lischke, 1869) reared on subtidal mudflats in a temperate bay of Korea. *Aquaculture* **322**, 149-157.
16. Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for General Users and for Biologist Programmers. *Bioinformatics Methods and Protocols.* pp. 365-386.
17. Sun, X., Li, L., Wang, X. 2017. Development and characterization of 37 SNP markers in the ark shell *Scapharca subcrenata* using RAD sequencing and high resolution melting analysis. *Conserv Genet Resour.* **9**, 365-368.
18. Yu, A., Shi, Y. and Yan, Y. 2020. Development and characterization of 50 SNP markers in *Coilia ectenes*. *Conserv Genet Resour.* **12**, 177-181.

초록 : 한국산과 중국산 새꼬막(*Scapharca subcrenata*)의 원산지 판별을 위한 SNP 마커의 개발 및 검증

최성석^{1*} · 유승현^{2*} · 서용배¹ · 김종오² · 권익정² · 배소희² · 김군도^{2*}
 (¹부경대학교 기초과학연구원, ²부경대학교 미생물학과)

본 연구에서는 한국산과 중국산 새꼬막(*Scapharca subcrenata*) 사이의 원산지 판별을 위하여 quantitative real-time PCR (qPCR) 분석을 기반으로 하는 Single-nucleotide polymorphism (SNP) 마커의 primer set를 개발 및 검증하였다. 총 180개의 새꼬막 sample을 genotyping by sequencing으로 분석하여 원산지 판별에 유용할 것이라 판단되는 7개의 후보 MS 마커와 15개의 후보 SNP 마커를 선정하였다. 후보 SNP 마커는 PCR과 sanger sequencing, SYBR green-based qPCR을 통해 원산지별 분리 여부를 확인하였다. 이 중 Insertion 1, SNP 21 마커가 qPCR 증폭 양상에서 집단이 확연히 분리되었으며 예상과 실제 증폭 형태가 일치하였다. 추가적으로 새꼬막을 무작위로 섞어서 진행한 blind test에서 Insertion 1은 새꼬막 100개에 대하여 74%의 정확도, 52%의 민감도, 96%의 특이도를 보였고, SNP 21은 새꼬막 137개에 대하여 86%의 정확도, 79%의 민감도, 93%의 특이도를 보였다. 따라서 개발된 두 개의 SNP 마커는 독립적 또는 복합적으로 사용하면 새꼬막 원산지 판별의 진위 여부를 검증하는 데 유용할 것으로 기대된다.