Thesis for the Degree of Master of Engineering

Biomarker Development for Discrimination of Different Geographical Origin on Mud loach (*Misgurnus mizolepis*) Using Metabolites Analysis

by

Hyunsuk Kim

Department of Food Science and Technology

The Graduate School

Pukyong National University

February, 2023

Biomarker Development for Discrimination of Different Geographical Origin on Mud loach (*Misgurnus mizolepis*) Using Metabolites Analysis

(대사체 분석을 이용한 미꾸라지의 원산지 판별 바이오마커의 개발)

Advisor: Prof. Ji-Young Yang

by
Hyunsuk Kim

A thesis submitted in partial fulfillment of the requirements
for the degree of

Master of Engineering

in Department of Food Science and Technology,
The Graduate School, Pukyong National University

February, 2023

Biomarker Development for Discrimination of Different Geographical Origin on Mud loach (*Misgurnus mizolepis*) Using Metabolites Analysis

A dissertation

by

Hyunsuk Kim

Approved by:

(Chairman) Young-Mog Kim. Ph.D.

(Member) Kil-Bo Shim. Ph.D.

(Member) Ji-Young Yang. Ph.D.

February 17, 2023

Contents

대사체 분석을 이용한 미꾸라지의 원산지 판별 바이오마커의 개발

김 현 석

부경대학교 대학원 식품공학과

요 약

국내에서 서식하고 있는 대표적인 미꾸리과(Cobitidae) 어류 중 하나인 미꾸라지(*Misgurnus mizolepis*)는 영양학적으로 우수하여 예로부터 섭취해 온 수산물이다. 자연산 종묘와 전문화된 시설의 부족으로 인하여 국내 미꾸라지 양식이 대폭 줄어듬과 동시에, 값싼 중국 미꾸라지의 수입이 대량으로 증가하는 추세이다. 우리나라 소비자들은 '국내산이 무조건 맛이 좋고 신선하다' 라는 국내선호 사상을 가지고 있기 때문에, 중국산 미꾸라지가 국내산 미꾸라지로 둔갑되는 사례가 자주 적발되고 있다. 따라서 대사체학을 이용하여 지리적으로 기원이 다른 미꾸라지의 원산지 판별을 위한

바이오마커를 개발하고자 하였다. 시료는 계절별로 수집하여 분쇄 및 동결건조 후 LC-QTOF/MS 기기를 통해 분석하였다. KEGG, HMDB, ChEBI Libray를 통하여 대사물질을 확인하였고, 다변량 통계기법을 활용하여 데이터 가공을 실시하였다. 개체군별 판별모델인 PCA, OPLS-DA를 통해 샘플의 전체적인 경향을 파악하였고, VIP score 1이상, *p*-value 0.05미만, fold change를 고려하여 국내산과 중국산 어종에서 유의미한 차이를 나타내는 대사체를 채택하였다. 최종적으로 바이오마커 선정을 위해 ROC curve를 작성하여 0.9이상 AUC value를 가지는 N-acetylhistidine과 anserine을 바이오마커로 선정하였다. 또한 계절별 샘플에 대하여, N-acetylhistidine의 AUC value는 전 계절에서 1을, anserine은 0.92이상 값을 각각 나타내었다. N-acetylhistidine의 정량분석을 위한 HPLC 분석 결과, 국내산은 1.62±0.95 μmol/g, 중국산은 3.27±0.88 μmol/g으로 나타났고, anserine의 정량분석을 위한 LC-ESI-MS/MS 분석 결과, 국내산 123.86±16.05 μg/g, 중국산 156.05±25.05 μg/g으로, 국내산과 중국산 미꾸라지를 유의적으로 구분이 가능함을 확인하였다.

. Introduction

Mud loach (*Misgurnus mizolepis*) is a species belonging to the family Cobitidae, which has long been a consumed species in Korea due to its nutritional excellence and good flavor. Its habitat is sand or muddy bottoms in the middle and lower reaches of rivers, and it is distributed not only in Korea but also in China and Taiwan (Kim et al., 2022; Song et al., 2017). The domestic aquaculture industry has stalled due to lack of natural seeding and environmental degradation. So, the market size of Chinese imported fish species is showing a trend of increasing (Choi et al., 2020). Recently, there is a case in which Chinese mud loach is disguised into expensive Korean mud loach because of its taste and preference. In addition, mud loach of low quality had a incident about excess ofloxacin and enrofloxacin contents, which are used as animal antibiotics. Therefore, it is important to identify the geographical origins of Korean and Chinese mud loach.

Metabolomics, a field of omics, is comprehensive and quantitative analysis of the metabolites interactions in biological systems (Dettmer et al., 2007). Food metabolomics has been used to discriminate the geographical origin of products, analyze food ingredients, and determine authenticity through identifying and monitoring of low molecular metabolites (Hoffmann et al., 2017; Kim et al., 2011; Mihailova et al., 2021; Wang et al., 2019). Metabolic analysis techniques are based on mass spectrometry, such as liquid chromatography/mass spectrometer (LC/MS), gas chromatography/mass spectrometer (GC/MS), nuclear magnetic resonance (NMR) spectroscopy (Lee et al., 2015; Shin et al., 2021; Yang et al., 2022; Zhao et al., 2020a). In data processing, multivariate data analysis, including principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) extracts information by simultaneously analyzing multiple variables. PCA is a technique for extracting distributed variables by reducing dimensions (Granato et al., 2018). OPLS-DA is used to identify biomarker candidates by excluding independent variables irrelevant to data while simultaneously

considering dependent variables to find relationships between groups (Kang et al., 2022; Zhao et al., 2020b).

In international trade, regulations of origin are becoming stricter and management strategies are being embodied. Therefore, identification of the geographical origin of food is becoming more important as consumers' awareness about the country of origin increases. Since it is possible to identify metabolites according to geographic factors, many studies on establishing a model for determining origin are in progress (Rivera-Pérez et al., 2021). Recently, metabolite profiling of perilla, sesame, goji berry, and tobacco leaves has been reported (Bondia-Pons et al., 2014; Kim et al., 2020; Sun et al., 2018). Due to the specificity of climate and nutrition, studies of agriculture products are progressing in various ways. In the case of livestock products, the origin of beef was discriminated using 1H NMR-based metabolomics (Jung et al., 2010). But, studies on the identification of origin using metabolome are insufficient for seafood. Although a method using a primer targeting the COI (Cytochrome c oxidase subunit I) gene region of mtDNA has been reported to discriminate the origin of mud loach

(Patent No. 10-2012-0077094, 2012), a biomarker for identification using metabolomics has not been developed.

This study used the metabolomic approach to identify the differences between Korean and Chinese mud loach. Untargeted analysis using LC-QTOF/MS and multivariate statistics were used for metabolic profiling to determine the origin. The discriminant model and quantitative analysis aimed to develop a biomarker that can discriminate the geographical origin of mud loach.

. Materials and Methods

1. Chemicals and reagents

N-Acetylhistidine was purchased from Sigma-Aldrich (Saint Louis, MO, USA), anserine was purchased from MedChem Express (Monmouth Junction, NJ, USA) and HPLC grade methanol, isopropyl alcohol, ACN (acetonitrile), formic acid and water were purchased from Honeywell Burdick and Jackson (Muskegon, MI, USA).

2. Sample collection and preparation

Korean (n=20) and Chinese (n=20) mud loach were obtained from a local fish market in Busan and an aquaculture farm for the four quarters of 2022. These included Yangpyeong (Gyeonggi-do), Namwon (Jeollanam-do), Jiangsu, and Jiangxi (Fig. 1). Sample information including total length and body weight is listed in Table 1. Muscle tissue was collected and lyophilized for 48 hours in a freeze dryer (Cooling trap

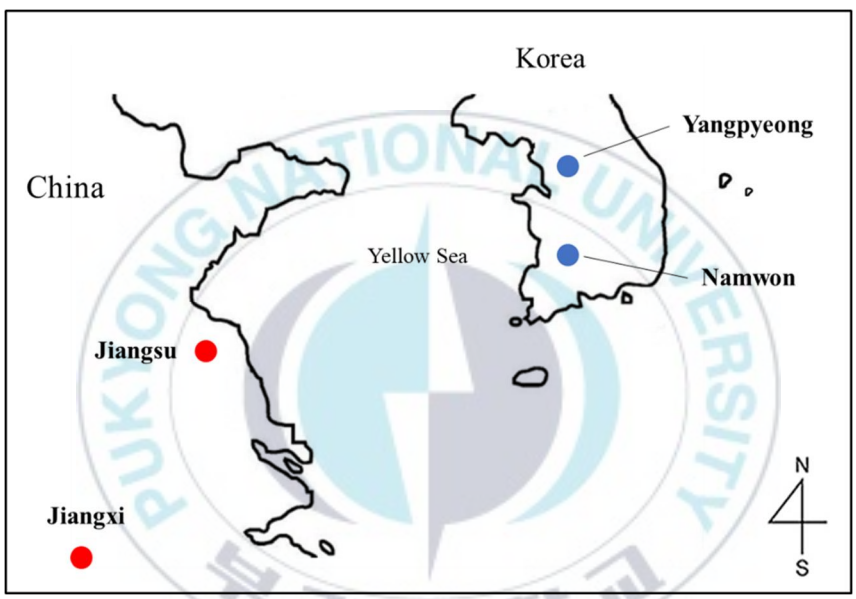


Fig. 1. Geographical origins of mud loach grown in Korea and China.

Table 1. Total lengh and body weight of mud loach by geographical origin

Season	Korea		China	
	Total length (cm)	**Body weight (g)**	**Total length (cm)**	**Body weight (g)**
Spring	18.34±0.34[1)]	32.41±3.97	17.98±0.39	36.70±1.93
Summer	17.44±0.23	26.61±1.26	16.64±0.89	26.14±2.75
Autumn	14.14±0.68	15.43±2.35	15.14±1.00	17.79±0.94
Winter	13.14±2.10	13.19±3.49	14.04±0.71	19.22±4.88

[1)] Mean±S.D (n=5)

HC3110, Bio-medical science Co., Ltd., Seoul, Korea). All samples were vacuum packaged and stored at -80℃ until analysis.

3. Metabolite extraction

Mud loach sample (20 mg) was homogenized by using Mixer Mill MM400 (Retsch GmbH & Co KG., Haan, Germany) at 28 Hz for 1 min and was then mixed with 1.6 mL of methanol:isopropyl alcohol:water (3:3:2, v/v/v) as an extraction solvent. After vortexing, the mixtures were sonicated for 10 min and centrifuged for 15 min (4℃, 13,200 rpm). The supernatant was pipetted into a fresh tube and dried using a speed vacuum concentrator (HyperVac, Gyrozen Co., Ltd, Gimpo, Korea). The dried samples were reconstituted in 100 μL of water. The extracts were filtered using a 0.2 μm PTFE filter (Advantec, Tokyo, Japan).

4. LC-QTOF/MS analyses

Metabolite profiling analysis was carried out using Acquity UPLC I class plus system (Waters, Milford, MA, USA) equipped with quadrupole

time-of-flight mass spectrometer (Synapt XS, Waters, Milford, MA, USA). Chromatographic separation was performed using Waters Acquity UPLC BEH C8 column (Waters, Milford, MA, USA) at a flow rate of 0.3 mL/min. The mobile phase was consisted of 0.1% formic acid in water (phase A) and 0.1% formic acid in ACN (phase B). The gradient elution program was applied as follows: 0-0.1 min, 0.5% B; 0.1-10 min, 80% B; 10-10.1 min, 99.5% B; 10.1-12 min, 99.5% B; 12-12.1 min, 0.5% B; 12.1-15 min, 0.5% B. The injection volume was 1 μL. QTOF mass analysis was operated in electrospray ionization (ESI) source in positive and negative ion modes with a mass range from *m/z* 50 to 1200. Mass spectrometric parameters were as follows: source capillary, 3 kV; sampling cone, 60; source temperature, 120℃; desolvation temperature, 350℃; cone gas, 20 L/hr; desolvation gas, 650 L/hr; MSe collision energy, 20-45 V. To estimate the stability of metabolites analysis, quality control (QC) samples containing the same amount of all samples were inserted. The operating conditions of LC-QTOF/MS for analysis were presented in Table 2.

Table 2. Conditions of LC-QTOF/MS analysis for untargeted profiling

		Conditions
Apparatus		Waters aquicity I class plus
Column		Waters BEH C8 (2.1 x 100 mm, 1.7 μm)
Oven temp.		40℃
Sample temp.		15℃
Injection volume		1 μL
Flow rate		0.3 mL/min
Mobile phase	A	0.1% Formic acid in Water
	B	0.1% Formic acid in ACN
Mass spectrometry		Synapt XS
	Ionization mode	ESI POS/NEG
	Mass arrange	50-1200 *m/z*
	Sampling cone	60
	Source capillary	3 kV
	Desolvation temp.	350℃
	Source temp.	120℃
	Cone gas	20 L/hr
	Desolvation gas	650 L/hr

5. Data processing

Peak alignment and picking were performed by the Progenesis QI (v.2.4, Waters, Milford, MA, USA) software from MS data. The metabolite matrix is obtained according to retention time (RT) and mass-to-charge ratio (*m/z*). To discriminate geographic origins between two groups of mud loach, MetaboAnalyst 5.0 (www.metaboanalyst.ca) was employed for multivariate statistical analysis. PCA was performed for clustering between groups, and OPLS-DA was carried out to find potential biomarker in the detected metabolites. Each ion data was normalized by Pareto scaling. The discriminating metabolites were selected based on variable importance in the projection (VIP) score > 1 in the OPLS-DA model and p-value < 0.05 in the t-test. The receiver operating characteristic (ROC) approach was used to identify biomarker. After that, metabolite structure identification was performed in databases (Kyoto Encyclopedia of Genes and Genomes, Chemical Entities of Biological Interest, The Human Metabolome Database) by matching 5 ppm of precursor tolerance and 5 ppm of fragment tolerance.

6. Quantitative analysis of N-acetylhistidine

Identification of N-acetylhistidine was performed by the reversed-phase HPLC method (O'Dowd et al., 1990; Yamada et al., 2009). The sample (0.5 g) was homogenized 10 times the 80% ethanol and centrifuged at 2,000 *g* for 20 min. The supernatants were dried in a speed vacuum concentrator, dissolved in phosphate buffer (pH 2), and filtered through 0.45 μm PTFE filter (Advantec, Tokyo, Japan). An isocratic reverse phase HPLC was performed, using the Acclaim C18 column (Thermo Scientific INC., Waltham, MA, USA). 0.1 M phosphate buffer (pH 2) was used as eluting solvent, with a flow rate of 0.6 mL/min. The injection volume was 10 μL and detected by UV absorbance at 210 nm (Table 3).

7. Quantitative analysis of anserine

Identification of anserine was performed as described in a previous study (Wang et al., 2021). The sample (0.5 g) was homogenized in 10% trichloroacetic acid and centrifuged at 10,000 *g* for 15 min at 4℃. The extracts were filtered using a 0.2 μm PVDF filter (Whatman, Maidstone,

Table 3. Conditions of HPLC analysis for N-acetylhistidine

	Conditions
Apparatus	Dionex ultimate 3000 UHPLC
Column	Dionex Acclaim C18 (4.6 x 250 mm, 5 μm)
Sample temp.	4℃
Injection volume	10 μL
Flow rate	0.6 mL/min
Mobile phase	0.1 M Phosphate buffer (pH 2)
UV detector	210 nm

UK) before using Thermo TSQ Endura equipped Ultimate 3000 UHPLC system (Thermo Scientific Inc., Waltham, MA, USA). Chromatographic separation was Xbridge BEH Amide column (Waters, Milford, MA, USA) at a flow rate of 0.15 mL/min. The mobile phase consisted of 0.3% ACN in water (phase A) and 0.8% ACN in water (phase B). And 0.1% ammonium hydroxide was added to the mobile phases. The gradient elution program was applied as follows: 0-10 min, 100% B; 10-11 min, 0% B; 11-14 min, 0% B; 14-15 min, 100% B; 15-25 min, 100% B. The injection volume was 2 μL. Mass analysis was operated in ESI source in positive mode and mass spectrometric parameters were as follows: sheath gas, 45 Arb; aux gas, 10 Arb; sweep gas, 1 Arb; ion transfer tube temperature, 400℃; vaporizer temperature, 300℃ (Table 4).

8. Statistical analysis

Statistical processing was performed using Minitab R19 software and significant differences were confirmed at the 95% confidence interval through one-way ANOVA Tukey analysis.

Table 4. Conditions of LC-ESI-MS/MS analysis of anserine

	Conditions	
Apparatus	Dionex ultimate 3000 UHPLC	
Column	Xbridge BEH Amide (2.1 x 150 mm, 2.5 μm)	
Oven temp.	35℃	
Injection volume	2 μL	
Flow rate	0.15 mL/min	
Mobile phase	A	0.3% ACN in Water
	B	0.8% ACN in Water
Mass spectrometry	TSQ Endura	
	Ionization mode	ESI POS
	Sheath gas	45 Arb
	Aux gas	10 Arb
	Sweep gas	1 Arb
	Ion transfer tube temp.	400℃
	Vaporizer temp.	300℃

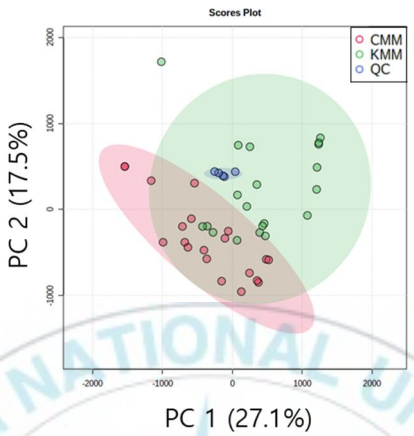
. Results and Discussion

1. Metabolite profiling of mud loach

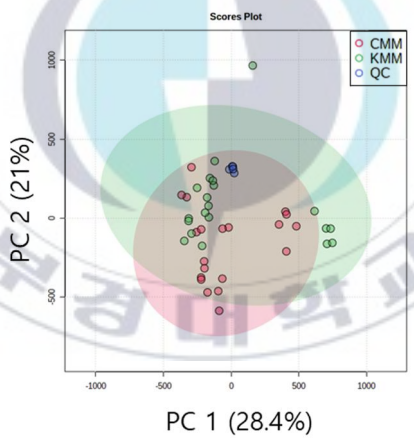
To discriminate the geographic origin of the mud loach, metabolite profiling was carried out using LC-QTOF/MS. Putative metabolites' identification was based on Chem Spider (KEGG, ChEBI) and HMDB, referring to RT and *m/z*. 21,772 metabolite ion features were extracted from samples of four seasons in positive ion and negative ion modes. QC samples were presented in PCA, a multivariate data analysis. As shown in Fig. 2, QC samples of mud loach are clustered, indicating that the data set is valid and stable.

2. PCA and OPLS-DA for geographic discrimination of mud loach

PCA is a statistical technique that extracts information, expresses it as a new orthogonal variable, and simplifies the size of the data set (Abdi and



(A)



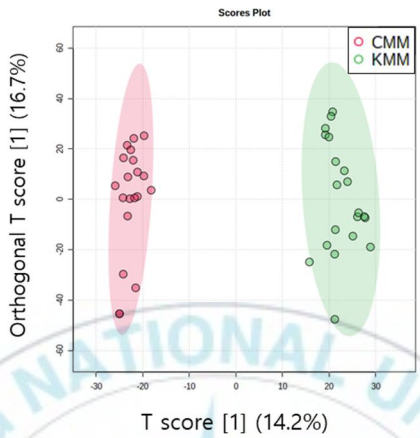
(B)

Fig. 2. PCA score plots of QC samples and mud loach samples in positive ion mode (A) and negative ion mode (B) derived from LC-MS. Abbreviation: KMM, Korean *Misgurnus mizolepis*; CMM, Chinese *Misgurnus mizolepis*; QC, Quality control.

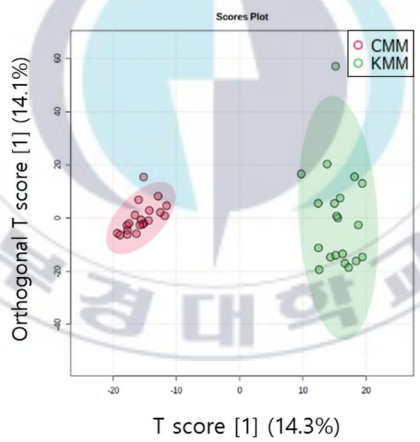
Williams, 2010). The unsupervised PCA score plots were constructed into two groups according to geographical origin in both ion modes (Fig. 2). The first two components of the PCA model explained 53.7% and 39.3% of the variance for positive and negative ion modes, respectively. However, there was no significant distinction between the two groups of mud loach in the PCA score plots. OPLS-DA was conducted to confirm a more clear separation. The supervised OPLS-DA score plots are a prediction and regression analysis method that can distinguish the two groups (Zhao et al., 2019). The OPLS-DA model with 0.90 for R2Y and 0.86 for Q2 in positive ion mode, 0.92 for R2Y and 0.9 for Q2 in negative ion mode is suitable and well validated. The Korean and Chinese mud loach were clearly clustered into two groups (Fig. 3).

3. Identification of differential metabolites as biomarker

The VIP score is used to indicate the contribution of the variable that divided into two groups. Potential biomarker metabolites were selected



(A)



(B)

Fig. 3. OPLS-DA score plots of mud loach samples in positive ion mode (A) and negative ion mode (B) derived from LC-MS. Abbreviation: KMM, Korean *Misgurnus mizolepis*; CMM, Chinese *Misgurnus mizolepis*.

based on VIP values > 1 from OPLS-DA (Chong and Jun, 2005). In addition, *p*-value < 0.05 and fold change were evaluated. A total of 9 ion features were confirmed, and metabolite identification was performed through the database (Table 5). Overall, it had a high discrimination potential in amino acid and fatty acid metabolites. Saccharin, glutathione, succinyl proline of Korean mud loach were significantly higher than those of Chinese. However, N-acetylhistidine, anserine, DL-carnitine hydrochloride, aspartame, stearidonic acid, DL-valine of Chinese mud loach were significantly higher than those of Korean (Fig. 4).

Finally, ROC curve was confirmed to discover biomarker of origin in mud loach. The ROC curve is a method of expressing the performance of a discriminant model. It is determined that the closer the AUC value, which is the calculated area under the ROC curve, to 1, the better the performance of the model (Izquierdo et al., 2001). The AUC value of N-acetylhistidine was 0.96 and anserine was 0.94. It provided a good quality score representing overall performance (Fig. 5). The AUC value of the multivariate ROC curve composed of the two selected metabolites was 0.99 (Fig. 6). The other metabolites were unsuitable as biomarker with

Table 5. Different metabolites between Korean and Chinese groups

No.	RT (min)	Identified metabolites	Adduct	*m/z*	Error (ppm)
1	0.82	Anserine	$[M+H]^+$	241.1304	3.7
2	0.87	Succinyl proline	$[M+H]^+$	216.0871	1.9
3	1.00	DL-Carnitine Hydrochloride	$[M+H]^+$	198.0883	-4.5
4	1.83	Glutathione	$[M+H]^+$	611.1448	0.1
5	1.96	DL-Valine	$[M+H]^+$	118.0865	1.9
6	3.19	Aspartame	$[M+H]^+$	295.1294	1.7
7	6.02	N-Acetylhistidine	$[M+H]^+$	198.0088	4.4
8	11.16	Saccharin	$[M+H]^+$	184.0066	1.7
9	11.39	Stearidonic acid	$[M+H]^+$	277.2164	0.7

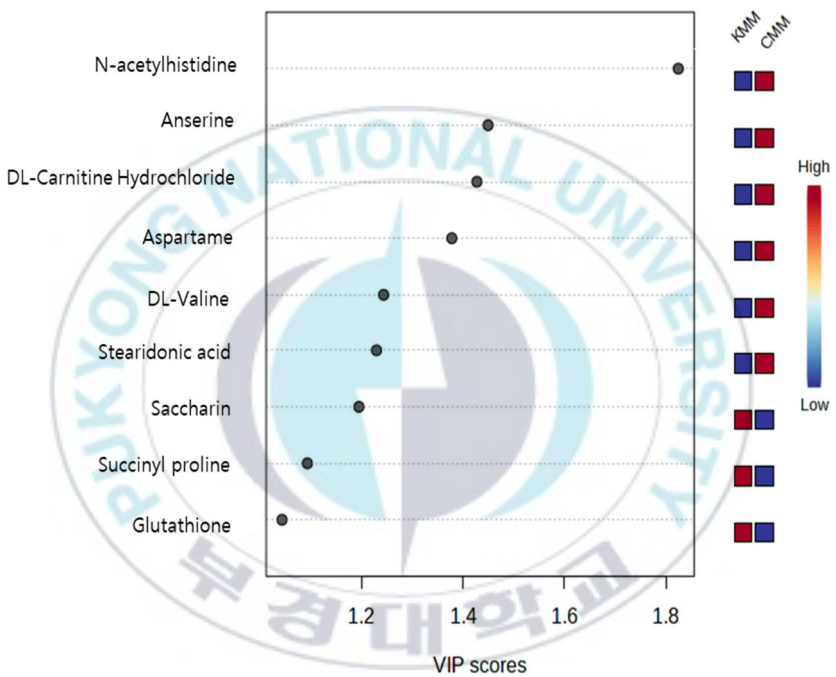
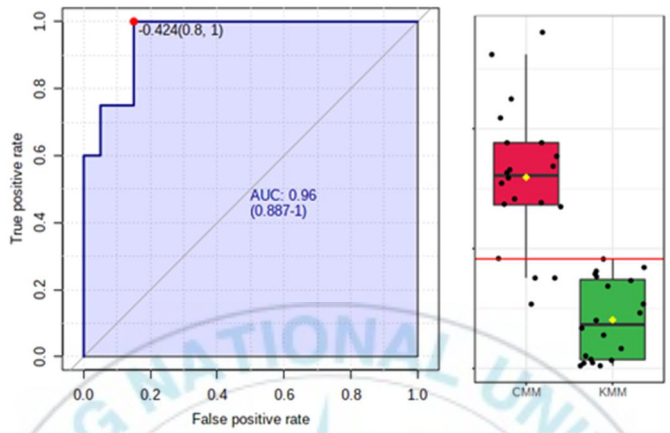
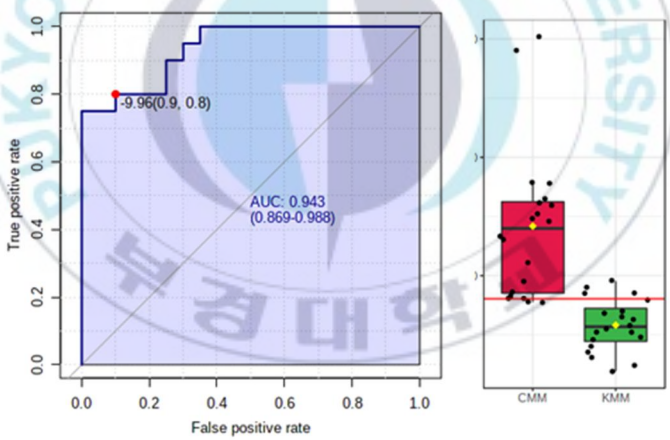


Fig. 4. VIP scores by OPLS-DA derived from LC-MS. Abbreviation: KMM, Korean *Misgurnus mizolepis*; CMM, Chinese *Misgurnus mizolepis*.



(A)



(B)

Fig. 5. ROC curve to discriminate between Korean (KMM) and Chinese mud loach (CMM). A, N-Acetlyhistidine; B, Anserine.

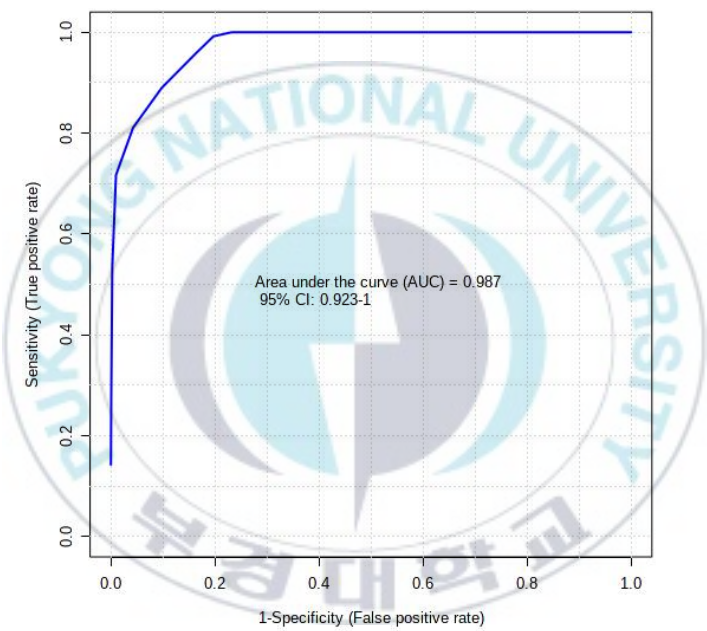


Fig. 6. Multivariate ROC curve of N-acetylhistidine and anserine to discriminate between Korean and Chinese mud loach.

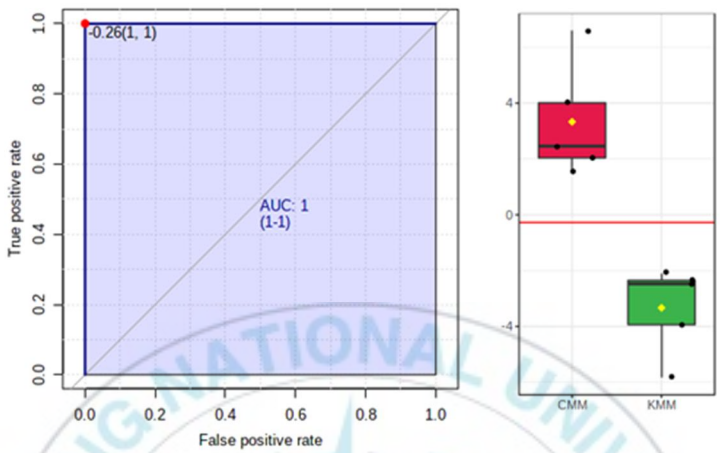
values below 0.9. Therefore, N-acetylhistidine and anserine were proposed as potential biomarkers for discrimination of geographical origin.

The composition of fish changes according to the period and climatic environments, such as spawning season and water temperature (Champion et al., 2022). Accordingly, seasonal variation was taken into account. As a result of considering VIP score, *p*-value, and fold change in four seasons, significant differences were found in N-acetylhistidine and anserine (Table 6). N-Acetylhistidine showed an AUC value of 1, indicating the excellent performance of the discriminant model (Fig. 7). In the case of anserine, it showed that AUC value greater than 0.9 in all four seasons, so the environment variable did not have a significant effect (Fig. 8).

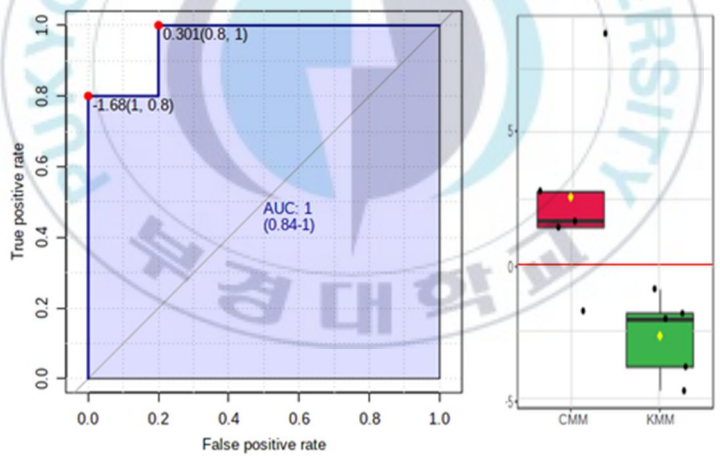
N-Acetylhistidine, one of the imidazole compounds, is a histidine derivative that is L-histidine having an acetyl substituent on the alpha-nitrogen. It was reported to exist as a major non-protein nitrogen component in the skeletal muscle of vertebrates (Yamada et al., 2009). N-Acetylhistidine decreased in muscle tissue under prolonged starvation (Yamada et al., 1994), but was fully recovered after the resumption of feeding. In addition, increasing water temperature was shown to increase

Table 6. Univariate statistics of biomarker metabolites in four seasons

Metabolites	Season	VIP score	Fold change	*p*-value
N-Acetylhistidine	Spring	1.76	3.3919	4.14×10^{-4}
	Summer	1.25	2.8695	2.19×10^{-2}
	Autumn	1.55	45.414	2.38×10^{-3}
	Winter	1.36	4.2004	1.51×10^{-4}
Anserine	Spring	1.65	4.0851	1.49×10^{-3}
	Summer	1.45	1.9810	3.09×10^{-3}
	Autumn	1.21	1.5187	4.72×10^{-2}
	Winter	1.16	1.6695	6.48×10^{-3}

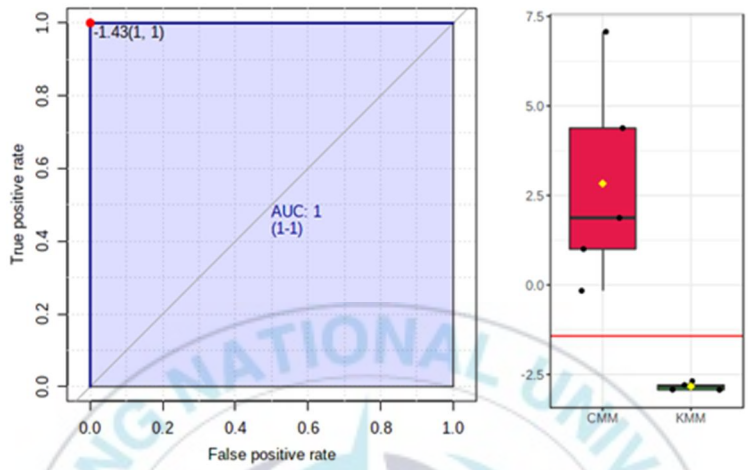


(A)

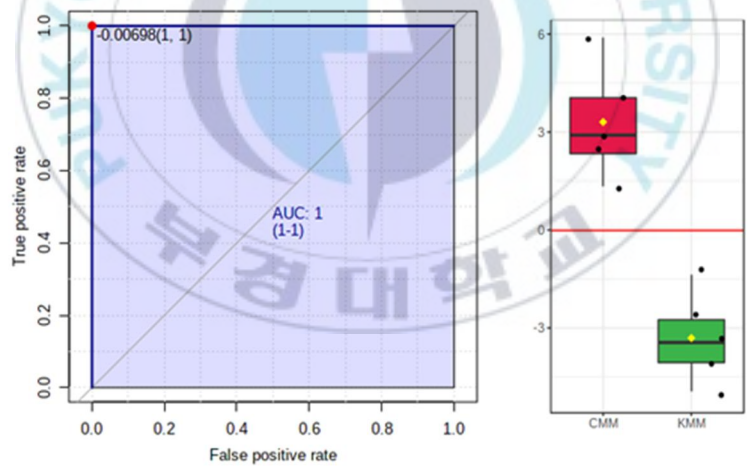


(B)

Fig. 7. ROC curve of N-acetylhistidine to discriminate between Korean (KMM) and Chinese mud loach (CMM). A, spring; B, summer; C, Autumn; D, Winter.

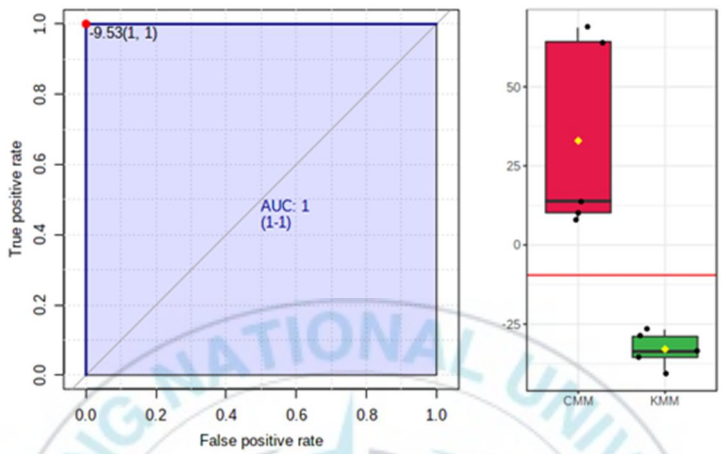


(C)

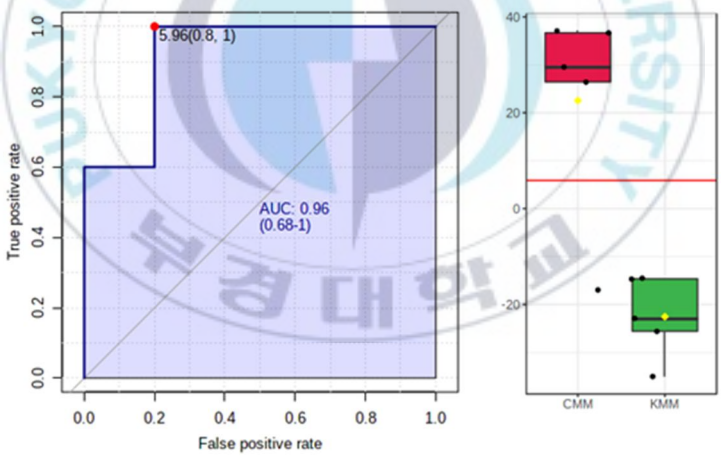


(D)

Fig. 7. Continued.

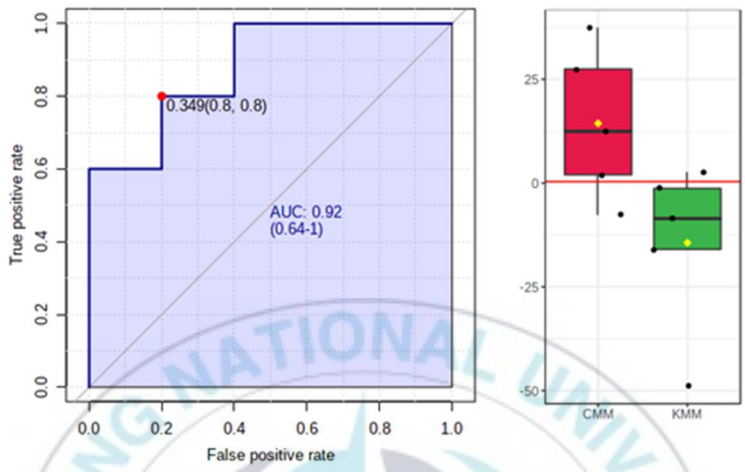


(A)

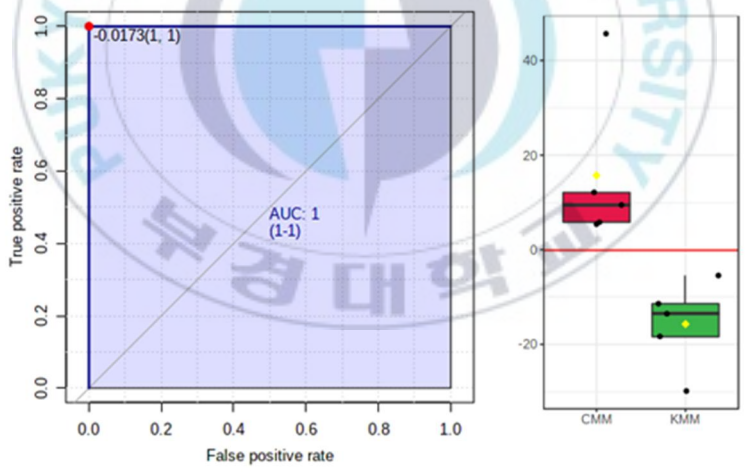


(B)

Fig. 8. ROC curve of anserine to discriminate between Korean (KMM) and Chinese mud loach (CMM). A, spring; B, summer; C, Autumn; D, Winter.



(C)



(D)

Fig. 8. Continued.

the N-acetylhistidine of fish (Geda et al., 2017). So, it proves the hypothesis that the N-acetylhistidine level is higher in China because the water temperature is higher in China than in Korea. Anserine is a dipeptide comprising β-Alanine and 3-Methyl-L-histidine. It was known to be present in skeletal muscle in fish (Abe, 1983; Ogata and Murai, 1994). In a previous study, there was no difference in anserine contents for a feed of rainbow trout (*Oncorhynchus mykiss*) that had diets containing different protein contents. The different anserine contents were affected by muscle buffering capacity and were influenced by swimming activity. That is, the anserine deficiency was related to the pH of the muscle (Aksnes et al., 2006; Baslow, 1998). Therefore, due to the physical environment that occurs during the import and distribution process, there seems to be a difference in the anserine levels of Korean and Chinese mud loach.

4. Quantitative comparison between two groups of different geographical origins

N-Acetylhistidine and anserine were quantified using calibration curves obtained from the standard compound. Typical chromatograms show a good baseline separation (Fig. 9, Fig. 10). The retention times of N-acetylhistidine and anserine were 8.76 and 6.98, respectively. The correlation coefficient (R^2) was 0.9985 for N-acetylhistidine and 0.9949 for anserine. In the MS/MS spectrum, product ions at 109.0 *m/z* and 169.9 *m/z* was detected as high intense fragment ion from the precursor ion 241.11 *m/z*. These results indicated that the method for quantification of the compound is exact and reliable. N-Acetylhistidine and anserine quantitative levels in the muscle tissue of mud loach are presented in Table 7 and Table 8. The amount of N-acetylhistidine was significantly different between Korean and Chinese samples in all four seasons. The average for Korean mud loach was 1.62±0.95 μmol/g and for Chinese mud loach was 3.27±0.88 μmol/g, which is about two times higher. The level of N-acetylhistidine decreased toward winter, this was due to the lower water

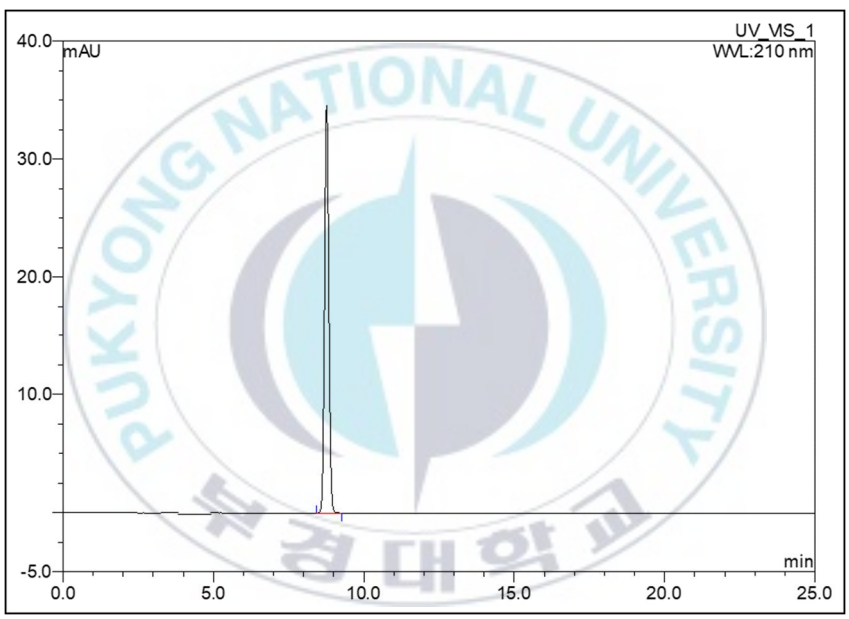


Fig. 9. Standard chromatograms of N-acetylhistidine.

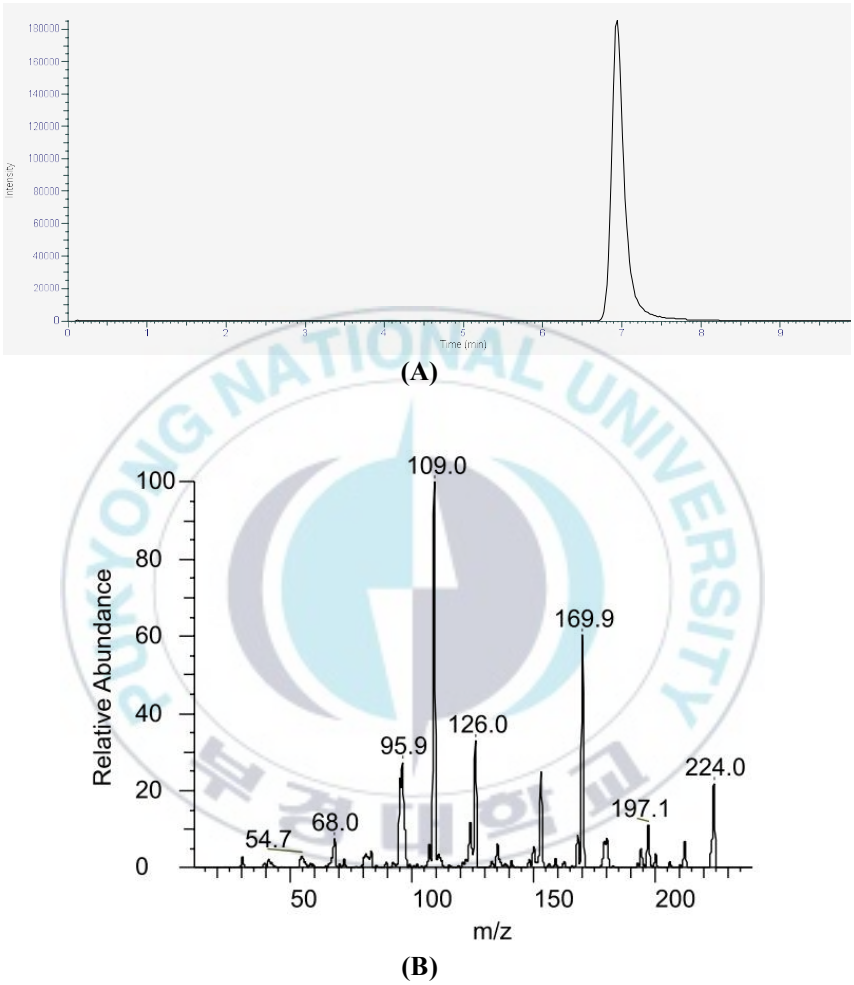


Fig. 10. Standard chromatograms (A) and MS/MS spectrum (B) of anserine acquired in positive mode.

Table 7. The amounts of N-acetylhistidine in muscles tissue of mud loach in Korea and China

Season	N-Acetylhistidine (μmol/g tissue)	
	Korea	China
Spring	2.73±0.27[1)a2)]	3.83±0.58[b]
Summer	2.29±0.39[a]	3.88±0.41[b]
Autumn	0.57.±0.18[a]	3.24±0.51[b]
Winter	0.87±0.19[a]	2.14±0.58[b]
Mean S.D	1.62±0.95[a]	3.27±0.88[b]

[1)] Mean±S.D (n=5)

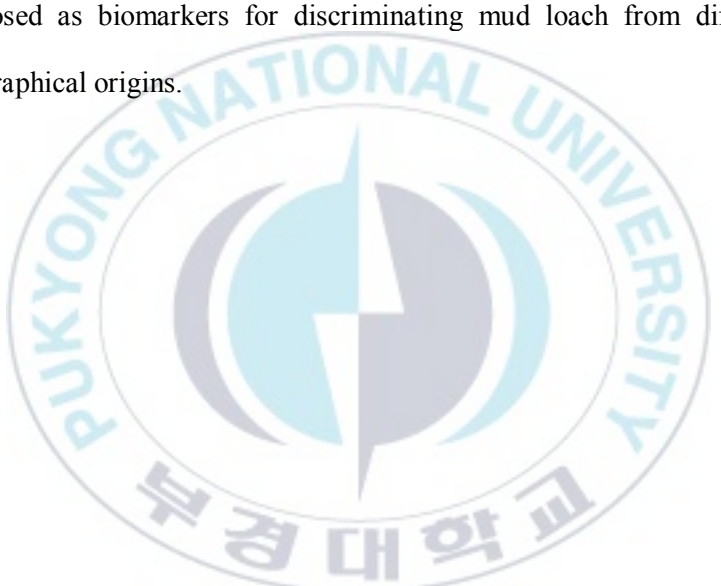
[2)] Same letter in line are not significantly different at the 5% level (p<0.05).

Table 8. The amounts of anserine in muscles tissue of mud loach in Korea and China

Season	Anserine (μg/g tissue)	
	Korea	China
Spring	113.40±10.36[a]	184.48±14.98[b]
Summer	130.48±16.01[a]	161.39±12.69[b]
Autumn	140.53±7.80[a]	154.81±6.60[b]
Winter	111.01±3.17[a]	123.50±13.55[b]
Mean S.D	123.86±16.05[a]	156.05±25.05[b]

[1)] Mean±S.D (n=5)
[2)] Same letter in line are not significantly different at the 5% level (p<0.05).

temperature. The contents of anserine in Chinese mud loach were significantly higher than those of Korean. The average for Korean mud loach was 123.86±16.05 μg/g and for Chinese mud loach was 156.05±25.05 μg/g. Therefore, N-acetylhistidine and anserine were proposed as biomarkers for discriminating mud loach from different geographical origins.



. Conclusion

This study identified metabolites of mud loach from different geographical origins using LC-QTOF/MS. Multivariate statistical analysis was used to select metabolites with significant differences between the two groups. In addition, it was confirmed and verified as a biomarker through quantitative analysis. A model for discriminating different geographical origins was established with PCA and OPLS-DA. Considering the VIP score, fold change, and *p*-value, significant compounds were identified. Finally, N-acetylhistidine and anserine were selected as biomarkers of geographical origin that discriminate the two groups through the ROC curve. Quantitative analysis validated the availability of metabolite markers, indicating a significant difference between the two groups. The result of this study is expected to prevent confusion in the distribution system and regulate inferior food incidents. Also, it will contribute to the development of aquatic product metabolomics.

. Summary

Mud loach has long been a consumed species in Korea due to its nutritional excellence and good flavor. There is a case in which a Chinese mud loach is deceived into an expensive Korean mud loach because of its taste and preference. This study analyzed metabolites of different geographical origins of mud loach using LC-QTOF/MS. PCA was performed to identify the group association of mud loach. OPLS-DA from LC-MS data sets showed a clear distinction between Korean and Chinese mud loach. The OPLS-DA model with 0.90 for R2Y and 0.86 for Q2 in positive ion mode, 0.92 for R2Y and 0.9 for Q2 in negative ion mode is suitable and well validated. Univariate statistical analysis results showed significantly different metabolites between Korean and Chinese samples. N-acetylhistidine and anserine were selected as a biomarker for discrimination of origin through the ROC curve. Quantitative analysis of biomarkers showed that two metabolites of Chinese mud loach were significantly higher than those of Korean. This results indicated that the

metabolic analysis could discriminate the geographical origins of mud loach.

. References

Abdi H, Williams LJ. 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433-459.

Abe H. 1983. Distribution of free L-histidine and its related compounds in marine fishes. Bulletin of the Japanese Society of Scientific Fisheries, 49, 1683-1687.

Aksnes A, Hope B, Albrektsen S. 2006. Size-fractionated fish hydrolysate as feed ingredient for rainbow trout (*Oncorhynchus mykiss*) fed high plant protein diets. II: Flesh quality, absorption, retention and fillet levels of taurine and anserine. Aquaculture, 261, 318-326.

Baslow MH. 1999. The existence of molecular water pumps in the nervous system: a review of the evidence. Neurochemistry International, 34, 77-90.

Bondia-Pons I, Savolainen O, Törrönen R, Martinez JA, Poutanen K, Hanhineva K. 2014. Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. Food Research International, 63, 132-138.

Champion C, Frank D, Taylor MC, Kaczmarska K, Piyasiri U, Broadhurst MK, Shalders TC, Coleman MA. 2022. Climate-driven changes to taste and aroma determining metabolites in an economically valuable portunid (*Portunus armatus*) have implications for future harvesting. Frontiers in Marine Science, 9, 973801.

Choi BK, Choe JR, Lee JH, Park JS, Kim JE, Hwang JA, Kim HS. 2020. Fish farm monitoring and growth performances of Chinese Muddy Loach (*Misgurnus mizolepis*) under indoor or outdoor aquaculture conditions, Korea. Korean Journal of Ichthyology, 32, 232-238.

Chong IG and Jun CH. 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems, 78, 103-112.

Dettmer K, Aronov PA, Hammock BD. 2007. Mass spectrometry□based metabolomics. Mass Spectrometry Reviews, 26, 51-78.

Geda F, Declercq AM, Remø SC, Waagbø R, Lourenço M, Janssens GP. 2017. The metabolic response in fish to mildly elevated water temperature relates to species-dependent muscular concentrations of imidazole compounds and free amino acids. Journal of Thermal Biology, 65, 57-63.

Granato D, Santos JS, Eschera GB, Ferreira BL, Maggio RM. 2018. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. Trends in Food Science and Technology, 72, 83-90.

Hoffmann JF, Carvalho IR, Barbieri RL, Rombaldi CV, Chaves FC. 2017. *Butia* spp. (Arecaceae) LC-MS-based metabolomics for species and geographical origin discrimination. Journal of Agricultural and Food Chemistry, 65, 523-532.

Izquierdo MS, Fernandez-palacios H, Tacon AGJ. 2001. Effect of broodstock nutrition on reproductive performance of fish. Aquaculture, 197, 25-42.

Jung Y, Lee J, Kwon J, Lee KS, Ryu DH, Hwang GS. 2010. Discrimination of the geographical origin of beef by 1H NMR-based metabolomics. Journal of Agricultural and Food Chemistry, 58, 10458-10466.

Kang C, Zhang Y, Zhang M, Qi J, Zhao W, Gu W, Gu J, Guo W, Li Y. 2022. Screening of specific quantitative peptides of beef by LC–MS/MS coupled with OPLS-DA. Food Chemistry, 387, 132932.

Kim EJ, Kwon J, Park SH, Park CY, Seo YB, Shin HK, Kim HK, Lee KS, Choi SY, Ruy DH, Hwang GS. 2011. Metabolite profiling of *Angelica gigas* from different geographical origins using 1H NMR and UPLC-MS analyses. Journal of Agricultural and Food Chemistry, 59, 8806-8815.

Kim EM, Dong CM, Lee MN, Noh JK, Noh ES, Nam BH, Kim YO, Jung HS. 2022. Development of multiplex species-specific PCR for the simultaneous identification of three closely related species in the genera *Misgurnus* and *Paramisgurnus*. Aquaculture Reports, 24, 101144.

Kim TJ, Park JG, Kim HY, Ha SH, Lee B, Park SU, Seo WD, Kim JK. 2020. Metabolite profiling and chemometric study for the discrimination analyses of geographic origin of perilla (*Perilla frutescens*) and sesame (*Sesamum indicum*) seeds. Foods, 9, 989.

Lee JE, Lee BJ, Chung JO, Kim HN, Kim EH, Jung SH, Lee HS, Lee SJ, Hong YS. 2015. Metabolomic unveiling of a diverse range of green tea (*Camellia sinensis*) metabolites dependent on geography. Food Chemistry, 174, 452-459.

Mihailova A, Kelly SD, Chevallier OP, Elliott CT, Maestroni BM, Cannavan A. 2021. High-resolution mass spectrometry-based metabolomics for the discrimination between organic and conventional crops: A review. Trends in Food Science and Technology, 110, 142-154.

O'Dowd JJ, Cairns MT, Trainor M, Robins DJ, Miller DJ. 1990. Analysis of carnosine, homocarnosine, and other histidyl derivatives in rat brain. Journal of Neurochemistry, 55, 446-452.

Ogata H, Murai T. 1994. White muscle of masu salmon, *Oncorhynchus masou masou*, smolts possesses a strong buffering capacity due to a

high level of anserine. Fish Physiology and Biochemistry, 13, 285-293.

Rivera-Pérez A, Romero-González R, Frenich AG. 2021. Application of an innovative metabolomics approach to discriminate geographical origin and processing of black pepper by untargeted UHPLC-Q-Orbitrap-HRMS analysis and mid-level data fusion. Food Research International, 150, 110722.

Shin JY, Yang JH, Cha EJ, Kim HS, Lee YH, Kim S, Choi I, Yang JY. 2021. Analyzing the metabolomic profile of yellowtail (*Seriola quinquerdiata*) by capillary electrophoresis–time of flight mass spectrometry to determine geographical origin. Metabolites, 11, 793.

Song HY, Moon SJ, Kim KS, Bang IC. 2017. Genetic species identification by sequencing analysis of nuclear and mitochondrial genes for albino *Misgurnus* species from Korea. Korean Journal of Ichthyology, 29, 139-145.

Soonchunhyang University Industry Academy Cooperation Foundation. Primers for distinguishing *Misgurnus mizolepis* and method using the same. Patent No. 10-2012-0077094, July 16, 2012.

Sun B, Zheng AH, Zhang F, Wei KS, Chen Q, Luo Y, Zhang Y, Wang XR, Lin FC, Yang J, Tang HR. 2018. Metabolic profiles of Cuibi-1 and Zhongyan-100 flue-cured tobacco leaves in different growing regions by gas chromatography/mass spectrometry. Royal Society Open Science, 5, 180261.

Wang CY, Li YR, Pan C, Chen J, Jiang W, Li WN, Yan XJ. 2021. Quantitative analysis of carnosine, anserine, and homocarnosine in skeletal muscle of aquatic species from east China sea. Biochemistry and Biophysics Reports, 25, 100880.

Wang Y, Li C, Li L, Yang X, Chen S, Wu Y, Zhao Y, Wang J, Wei Y, Yang D. 2019. Application of UHPLC-Q/TOF-MS-based metabolomics in the evaluation of metabolites and taste quality of

Chinese fish sauce (Yu-lu) during fermentation. Food Chemistry, 296, 132-141.

Yamada S, Kawashima K, Baba K, Oku T, Ando S. 2009. Occurrence of a novel acetylated amino acid, *Nα*-acetylhistidine, in skeletal muscle of freshwater fish and other ectothermic vertebrates. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 152, 282-286.

Yamada S, Tanaka Y, Sameshima M, Furuichi M. 1994. Effects of starvation and feeding on tissue *Nα*-acetylhistidine levels in Nile tilapia *Oreochromis niloticus*. Comparative Biochemistry and Physiology Part A: Physiology, 109, 277-283.

Yang JH, Shin JY, Cha EJ, Kim HS, Lee YH, Kim S, Choi I, Yang JY. 2022. Analysis of metabolites of red seabream (*Pagrus major*) from different geographical origins by capillary electrophoresis time-of-flight mass spectrometry. PloS One, 17, e0270280.

Zhao G, Zhai X, Qu M, Tong C, Li W. 2019. Sulfated modification of the polysaccharides from *Crassostrea gigas* and their antioxidant and hepatoprotective activities through metabolomics analysis. International Journal of Biological Macromolecules, 129, 386-395.

Zhao G, Zhao W, Han L, Ding J, Chang Y. 2020a. Metabolomics analysis of sea cucumber (*Apostichopus japonicus*) in different geographical origins using UPLC–Q-TOF/MS. Food Chemistry, 333, 127453.

Zhao S, Liu H, Su Z, Khoo C, Gu L. 2020b. Identifying cranberry juice consumers with predictive OPLS□DA models of plasma metabolome and validation of cranberry juice intake biomarkers in a double□blinded, randomized, placebo□controlled, cross□over study. Molecular Nutrition and Food Research, 64, 1901242.