Thesis for the Degree of Doctor of Philosophy

Application of Metabarcoding for Fish Species Identification

by

Eun Soo Noh

Department of Microbiology

The Graduate School

Pukyong National University

August 2017

Application of Metabarcoding for Fish Species Identification

수산가공품의 원료종 식별을 위한 메타바코딩 응용에 관한 연구

Advisor: Prof. Kyoung-Ho Kim

by

Eun Soo Noh

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

Doctor of Philosophy

in the Department of Microbiology, Graduate School,

Pukyong National University

August 2017

Application of Metabarcoding for Fish Species Identification

A dissertation

by

Eun Soo Noh

Approved by

Chairman : Tae-Jin Choi

Member : Gun-Do Kim

Member : Young-Mog Kim

Member : Jung-Ha Kang

Member : Kyoung-Ho Kim

CONTENTS

List of Tables

Chapter II.

Chapter III.

Chapter IV.

List of Figures

Chapter II.

Chapter III.

Chapter IV.

Application of Metabarcoding for Fish Species Identification

Eun Soo Noh

Advisor: Prof. Kyoung-Ho Kim

Department of Microbiology, Graduate School, Pukyong National University

Abstract

The identification of raw materials in processed foods is a major issue in food safety management. Substitution of an expensive species by an inexpensive one (i.e., fraudulent labeling) is illegal and provides unfair profits to the manufacturer. Identification of species by their morphological characteristics is sufficient to identify fresh fish, but cannot easily be applied to processed foods (fish fillets, canned foods), which comprise a large proportion of the seafood consumed worldwide. Molecular technology has been widely used in recent years to overcome the limitations of morphology-based identification methods. Especially, DNA metabarcoding is a potentially useful method for quick assessing of the abundance of taxa in complex environments.

Metabarcoding using next generation sequencing (NGS) provides a good representation of species diversity. Here, the advantages and limitations of using

NGS for fish species identification from seafood products were evaluated. Two universal primer sets were designed based on the mitochondrial cytochrome c oxidase subunit I (COI) gene to amplify a barcode region of ~650 bp. Eleven seafood products were analyzed using the two primer sets in separate experiments; 56 and 32 fish species were identified. Similar amounts of the main species (e.g., Alaska Pollock and golden thread) were detected in most products by both primer sets. These new primer sets for metabarcdoing studies by NGS have shown that they can be used for species identification of processed seafood products.

NGS yields accurate results, but it is not considered cost-effective for practical use. For that reason, the denaturing gradient gel electrophoresis (DGGE) fingerprinting technology has been applied to the metabarcoding analysis of seafood products. DGGE uses a reduction in electrophoretic mobility according to the denaturation characteristics of amplicons to facilitate in-gel separation. Therefore, various fish species in seafood products can be identified by DGGE fingerprinting. A universal primer set was designed based on the COI gene to amplify a 214 bp fragment, and the fish species in various seafood products were identified using DGGE fingerprinting. A total of 30 bands were identified, and 20 species were identified from 20 seafood products. A smaller number of fish species were identified in comparison with former NGS results, but the major species were almost the same. Therefore, DGGE fingerprinting is suitable for the detection of various raw materials at a reasonable cost in seafood products.

Finally, a study was conducted a quantitative analysis of Alaska pollock, which is the most commonly used fish in processed seafood products. Droplet digital PCR (ddPCR) technology was used to determine the content of Alaska pollock.

This technique enables the absolute quantification of the number of copies of a target DNA sequence in a sample. A universal primer set and specific probes were designed based on the mitochondrial 16S rRNA gene. This gene was chosen because it exhibits low intraspecific genetic variation; thus, false-negative results due to genetic variation can be avoided. The linear relationship between sample weight and DNA concentration for both Alaska pollock and cutlassfish was analyzed. In addition, a linear relationship was observed between the DNA copy numbers obtained by ddPCR technology. These results enabled us to devise a formula for determining the weight of Alaska pollock based on the DNA copy number, and the content of Alaska pollock in 15 seafood products was successfully measured.

Chapter I.

General Introduction

General Introduction

1. Importance of species identification in seafood products

Identification of raw materials in processed food is a major issue in food safety management. Increased international trade and worldwide seafood consumption have affected the demand for, and supply of certain fish species, and it is used as an economic fraud in the world seafood trade (Jacquet & Pauly, 2008). A variety of seafood products are produced by the food-processing industry. The manufacturers of some of these products use low-value species rather than expensive major industrial species, and attempt to increase profits by using fake labels (Rasmussen & Morrissey, 2008). Identification of species by their morphological characteristics is sufficient to identify fresh fish, but cannot easily be applied to processed foods (fish fillets, canned foods), which comprise a large proportion of the seafood consumed worldwide (Rasmussen & Morrissey, 2009). Processed food lacks distinguishing morphological features, such as size, body and fin (Armani et al., 2012).

The European Union has established information labeling laws for seafood product, which require accurate traceability information such as species name, origin, and method of production. Seafood fraud and species substitutions are prohibited in the United States according to the Federal Food Drug and Cosmetic Act Section 343(b): A food shall be deemed to be misbranded if “it is offered for sale under the name of another food” (Schultz, 2012). The US Food and Drug Administration (FDA) provides a list of seafood species to facilitate correct labeling of seafood product (Cawthorn et al., 2012). However, enforcement of labeling regulations for processed food is difficult because the composition be

confirmed. In addition to the detrimental effects on the market caused by substitution of seafood product, the health of consumers may be at risk due to pollution by, or toxicity of, fraudulent products. Moreover, endangered species can be threatened (Armani et al., 2015). Therefore, seafood product should be classified as safe, and technology for quality control and identification of raw materials is required.

Identifying species in seafood is challenging. It is estimated that ~20,000 species of fish are consumed worldwide (Teletchea, 2009). Sequencing analysis of DNA polymorphisms is the most frequently used method for identification of fish species. However, identification can be hampered by the presence of several species with similar sequences or intraspecific genetic variations. Also, analysis of processed foods is hampered by DNA degradation during boiling and frying (Rasmussen & Morrissey, 2008). In addition, various factors present in processed foods can act as inhibitors of polymerase chain reaction (PCR) amplification (Espiñeira et al., 2009). Therefore, the method should be optimized for analysis of the products.

In general, species identification is achieved by sequencing of the mitochondrial COI, Cytb, or 16S rRNA regions (Ferri, Alù, Corradini, Licata, & Beduschi, 2009). Mitochondrial DNA (mtDNA) analysis has several advantages over nuclear DNA. First, a large amount of DNA can be obtained from a small number of samples because mtDNA has a higher DNA copy number than nuclear DNA (nDNA). Second, ambiguous species by heterozygous genotypes can be avoided because mtDNA is haploid with maternal inheritance (Alberts et al., 1994). Several time- and cost-effective PCR methods are used to identify species (Gil, 2007). Typically, restriction fragment length polymorphism-PCR (RFLP-

PCR) or random amplified polymorphic DNA (RAPD-PCR) (non-specific amplification), and major histocompatibility complex (MHC-PCR) (species-specific amplification) are used for rapid identification of target species by electrophoresis (Gagnaire et al., 2007; Gil, 2007; Slawomir et al., 2009). However, conventional methods are unable to distinguish species in mixed seafood products, such as surimi. This problem can be overcome by estimation of biodiversity using a metabarcoding approach.

2. Metabarcoding

DNA metabarcoding enables rapid assessment of the relative abundance of taxa in various environments (Yu et al., 2012). DNA barcoding relies on the partial genome sequences of various species, typically the mitochondrial (animal species) or chloroplast (plant species) genes (Taberlet et al, 2012). The sequence of the selected region to be used for species identification is termed a barcode. The barcode markers to be used for each taxon have been standardized by The Consortium for the Barcode of Life (CBOL); the standard barcode marker for animals is the mitochondrial cytochrome oxidase I gene (Ratnasingham & Hebert, 2007). Metabarcoding is a combination of two techniques; DNA-based species identification and massively parallel sequencing (MPS). The PCR-amplified barcode gene is subjected to next-generation sequencing (NGS), which yields abundant DNA sequences. These sequences are classified using analytical software and identified by comparison with reference database. Determination of presence or absence of particular species is possible, and metabarcoding analysis is more comprehensive and less dependent on professional taxonomic knowledge (Cowart et al., 2015).

In general, identification of species from a whole-specimen can be considered as DNA metabarcoding. In addition, metabarcoding differs from metagenomics because the former involves identifying specific taxa, and the latter determination of microbial diversity in environmental samples (Young et al., 2014). Here, metabarcoding was used to analyze the seafood product samples. Such samples usually comprise a limited number of taxa and yield higher quality DNA than environmental samples. Thus, utilization of genetic markers for analysis is

important for species identification. Despite its broad range of applications, metarbarcoding is somewhat unreliable. As mentioned earlier, metabarcoding provides only limited information on the biodiversity of a particular sample, and is unable to identify certain taxa because universal primer for the barcode gene do not cover all species (Leray et al., 2013). Therefore, indicator species, species of interest or ecologically/economically important species can be missed. Development of universal metabarcoding primers for specific taxa is thus vital (Deagle et al., 2014).

The majority of primers for metabarcoding have been developed for analysis of environmental samples. Because DNA can be degraded by adverse environmental conditions, primers are specific for short sequences in the barcode region; the resulting low resolution may limit the species discrimination ability (Miya et al., 2015). Generally, a sequence of 100 bp has a species identification resolution of ~90% (Meusnier et al., 2008). Therefore, development of optimized primers and methods for identification of aquatic species from seafood products is required.

3. Analytical methods

3.1 Next generation sequencing

In chapter 2, NGS was used for metabarcoding analysis of raw materials in seafood products. NGS, also known as MPS, enables rapid analysis of hundreds of thousands to millions of sequence reads (Van Dijk et al., 2014). NGS involves decomposition of a genome into smaller pieces, each of which is read simultaneously, followed by bioinformatics analysis (Ansorge, 2009). In this study, 454 Genome Sequencer FLX (Roche, Basel, Switzerland) was used to analyze DNA extracted from mixed seafood products and ensure that they are safe for consumption.

3.2 Denaturing gradient gel electrophoresis

In chapter 3, metabarcoding analysis of raw materials in seafood products was performed by denaturing gradient gel electrophoresis (DGGE). DGGE enables evaluation of molecular fingerprints by electrophoretic resolution of the PCR amplicons according to their nucleotide sequence (Ercolini, 2004b). As it migrates through the polyacrylamide gel, double-stranded DNA is denatured by the chemical denaturant, slowing migration considerably. Different DNA sequences produce different band patterns depending on the concentration of the denaturant; theoretically, each band represents the same gene (Peng et al., 2007). This technique has been used to analyze microbial in the human gut (Zhang et al., 2009), foodstuffs (Ercolini, 2004), water (Van Hannen et al., 1998) and soil

(Nakatsu et al., 1999). In this study, the feasibility of DGGE for metabarcoding analysis was demonstrated.

3.3 Droplet digital PCR

In chapter 4, metarbarcoding studies using the third-generation PCR technology, droplet digital PCR (ddPCR), confirmed the abundance of target species in seafood products. ddPCR enables quantification of nucleic acid with greater accurate than conventional quantitative real-time PCR (qRT-PCR) (Hindson, et al., 2013). Quantification by qRT-PCR utilizes the relationship between the cycle threshold value and the initial DNA template concentration using a standard curve (Vandesompele, et al., 2002), and differences in PCR efficiency can affect the results (Hayden, et al., 2012). In contrast, ddPCR facilitates approach to absolute quantification of nucleic acids by directly determining the number of target molecules without the need for a standard curve.

4. Objective of this study

The objective of this study was to develop a technique for identification of fish species in mixed seafood products containing unknown raw materials. To verify the accuracy and practicality of the metabarcoding approach, three analytical techniques were used.

(1) In chapter 2, NGS analysis was performed for metabarcoding study. Metabarcoding has been used for the first time to the analysis of processed seafood products and the possibility for the identification of raw materials was verified.

(2) In chapter 3, the development of the DGGE method for metabarcoding study is described. The possibility of practical use for food safety management was confirmed.

(3) In chapter 4, the ddPCR method for quantitative analysis of Alaska pollock from seafood products is described. The formula for calculate the raw sample weight was established based on the number of DNA copies and the accuracy and applicability of this method was verified.

These metabarcoding approaches facilitate rapid identification of the raw materials of mixed seafood products, which will enhance to food safety management and provision of safe foodstuffs to consumers.

5. References

Albert, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Molecular biology of the cell, 4th edition. *New York and Lundon: Garland Publishing Inc.*

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *New biotechnol*, *25*(4), 195-203.

Armani, A., Castigliego, L., & Guidi, A. (2012). Fish frauds: the DNA challenge. *CAB Rev*, *7*(71), 1-12.

Armani, A., Guardone, L., Castellana, R. L., Gianfaldoni, D., Guidi, A., & Castigliego, L. (2015). DNA barcoding reveals commercial and health issues in ethnic seafood sold on the Italian market. *Food Control*, *55*, 206-214.

Cawthorn, D. M., Steinman, H. A., & Witthuhna, R. C. (2012). DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Res. Int*, *46*(1), 30-40.

Cowart, D. A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., & Haond, S. A. (2015). Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS One*, *10*(2), e0117562.

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Letters*, *10*(9), 20140562.

Ercolini, D. (2004). PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J Microbiol. Meth*, *56*(3), 297-314.

Espiñeiran, M., Vieites, J. M., & Santaclara, F. J. (2009). Development of a genetic method for the identification of salmon, trout, and bream in seafood products by means of PCR–RFLP and FINS methodologies. *Eur. Food Res. Technol, 229*(5), 785-793.

Ferri, G., Alù, M., Corradini, B., Licata, M., & Beduschi, G. (2009). Specis identification through DNA "Barcodes". *Genet. Test. Mol. Bioma, 13*(3), 421-426.

Gagnaire, P. A., Tsukamoto, K., Aoyama, J., Minegishi, Y., Valade, P., & Berrebi, P. (2007). RFLP and semi-multiplex PCR-based identification of four eel species from the south-western Indian Ocean region. *J. Fish Biol, 71*, 279-287.

Gil, L. A. (2007). PCR-based methods for fish and fishery products authentication. *Trends Food Sci. Technol, 18*(11), 558-566.

Hayden, R. T., Gu, Z., Ingersoll, J., Abdul-Ali, D., Shi, L., Pounds, S., & Caliendo, A. M. (2012). Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. *J. Clin. Microbiol, 51*(2), 540-546.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., Vessella, R. L., & Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods, 10*(10), 1003-1005.

Jacquet, J. L., & Pauly, D. (2008). Trade secrets: Renaming and mislabeling of seafood. Mar. Policy, 32(3), 309-318.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set

targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool, 10,* 34.

Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics, 9*(1), 214.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. Open Sci, 2*(7), 150088.

Nakatsu, C. H., Torsvik, V., & Øvreås, L. (1999). Soil Community Analysis Using DGGE of 16S rDNA Polymerase Chain Reaction Products. *Soil Sci. Soc. Am. J, 64*(4), 1382-1388.

Peng, L., Chaoqun, H., Luping, Z., Chunhua, R., & Qi, S. (2007). Effects of DNA extraction and universal primers on 16S rRNA gene-based DGGE analysis of a bacterial community from fish farming water. *Chin. J. Oceanol. Limn, 25*(3), 310-316.

Rasmussen, R. S., & Morrissey, M. T. (2008). DNA-Based Methods for the Identification of Commercial Fish and Seafood Species. *Compr. Rev. Food Sci. F, 7*(3), 280-295.

Rasmussen, R. S., & Morrissey, M. T. (2009). Application of DNA-based methods to identify fish and seafood substitution on the commercial market. *Compr. Rev. Food Sci. F, 8*(2), 118-154.

Ratnasingham, S., & Hebert, P. D. N. (2007). Bold: The Barcode of Life Data

System (http://www.barcodinglife.org). *Mol. Ecol. Resour, 7*(3), 355-364.

Schultz, H. (2012). Food law handbook. Springer Science & Business Media.

Slawomir, K., Remigiusz, P., & Jolanta, K. (2009). Eel species identification by polymerase chain reaction followed by restriction fragment length polymorphism (PCR-RFLP). *Med. Wet, 65*(5), 315-318.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol, 21*(8), 2045-2050.

Teletchea, F. (2009). Molecular identification methods of fish species: reassessment and possible applications. *Rev. Fish Biol. Fisher, 19*(3), 265.

Vandesompele, J., Preter, K. D., Pattyn, F., Poppe, B., Roy, N. V., Paepe, A. D., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol, 3*(7), research0034-0031.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet, 30*(9), 418-426.

Van Hannen, E. J., Van Agterveld, M. P., Gons, H. J., & Laanbroek, H. J. (1998). Revealing genetic diversity of eukaryotic microorganisms in aquatic environments by denaturing gradient gel electrophoresis. *J. Phycol, 34*(2), 206-213.

Young, J. M., Weyrich, L. S., & Cooper, A. (2014). Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. *Forensic Sci. Int-Gen, 13*, 176-184.

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity

assessment and biomonitoring. Methods *Ecol. Evol*, *3*(4), 613-623.

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., Parameswaran, P., Crowell, M. D., Wing, R., Rittmann, B. E., & Brown, R. K. (2009). Human gut microbiota in obesity and after gastric bypass. *P. Natl. Acad. Sci. USA*, *106*(7), 2365-2370.

Chapter II.

Metabarcoding Approach for Detection of Raw Material in Processed Seafood

Metabarcoding Approach for Detection of Raw Material in Processed Seafood

1. Abstract

The increased international trade and worldwide consumption of seafood has increased the possibility of economic fraud in the world seafood trade. Especially, the tremendous development of processed seafood has heavily restricted the application of classical identification methods. DNA matabarcoding is rapid method for identification of multiple species from an environmental sample. In this study, two DNA markers were developed in the mitochondrial cytochrome oxidase subunit I region, and the metabarcoding analysis of processed seafood products was performed using these markers. A total of 11 seafood products were analyzed, and the 56 and 32 species of fish were identified by each primer set, respectively. The results obtained by the two primer sets were compared and similar contents of the main species were obtained in most of the seafood products. These results demonstrate the potential of metabarcoding approach for seafood safety management.

Keyword: Metabarcoding; processed seafood; species identification; forensic; universal primer

2. Introduction

As a result of the increased demand for seafood and the globalization of the seafood industry, more species of fish are now available from the market. Although there are advantages for obtaining various seafood more easily and cost-effectively, the increased species diversity in the seafood market can cause problems in species identification (Wong & Hanner, 2008). In this regard, the ability to identify the species in raw or processed food is an important challenge for efforts to prevent commercial seafood fraud, and many studies have explored this issue. It is especially difficult for consumers to distinguish among species when morphological determination is not possible due to morphological similarities or processing (Cutarelli et al., 2013). The presence of harmful substances, such as toxins, can be potentially dangerous to consumers if substitute or mislabeled fish are available in markets (Rasmussen & Morrissey, 2008b). Therefore, species identification is essential for consumer protection and can also protect endangered species from illegal fishing.

Molecular technology is an effective species identification tool that can overcome problems with morphology-based identification methods. Specifically, metabarcoding technology has been a powerful way to explore biological diversity in complex environments (Hänfling et al., 2016). Direct sequencing of the DNA present in environmental samples allows for rapid and accurate detection of various taxa without morphological identification. The analysis of environmental DNA can increase understanding of biological diversity as it can identify organisms that had been undetected using conservative methods (Evans et al., 2015). Therefore, this technology is used mainly to investigate local fauna,

such as rare and endangered species (Thomsen et al., 2012). In recent years, metabarcoding research has been actively performed to analyze species diversity in various environmental samples, such as rivers (Civade et al., 2016; Shaw et al., 2016) and seas (Yamamoto et al., 2017; Zaiko et al., 2015). However, this is a relatively new approach for the seafood industry.

The success of metabarcoding analysis is based on the target gene region and the primer set because they affect the efficiency and accuracy of species identification (Leray et al., 2013). The coverage of primers for species identification depends on the purpose of the analysis; generally, specific primer sets for the target group are effective. However, using these specific primer sets on an unintended group can lead to a false positive or false negative result. Therefore, when analyzing samples containing numerous species, the primers should target conserved regions that are universally available (Miya et al., 2015).

Despite the difficulty of designing universal primers for numerous species, several mitochondrial gene regions have been amplified by universal primers in most animals. For example, the mitochondrial 16S rRNA, cytochrome b, and cytochrome c oxidase I genes have been used for evolutionary research(Ivanova et al., 2007; Kocher et al., 1989). However, analysis of the 16S rRNA region may have underestimated diversity due to its relatively slow rate of evolution compared to other regions(Xia et al., 2011). The Cyt b region is suitable for differentiation and identification of species with high rates of molecular evolution, but the accumulated taxonomic databases in this region remain limited(Tobe et al., 2010). The cytochrome c oxidase I (COI) gene has been selected as the standard barcode for taxon descriptions(Hebert et al., 2003). The relatively short sequence

of the COI region used in barcodes shows sufficient variation to accurately identify various species.

The aim of this study was to provide an efficient metabarcoding approach for detecting various fish species in mixed seafood such as surimi that does contain unknown raw material. To verify the precision and feasibility of this approach, two primer sets were used to amplify the partial COI gene. This method successfully detected a variety of species, and these results will contribute to food safety management by detecting the raw material in mixed seafood.

3. Materials and methods

3.1 Sample collection

Surimi products were purchased at local markets in Busan, South Korea. The samples were transported to the laboratory on ice and kept frozen at -20°C until use. General information for each sample is shown in Table 1.

3.2 DNA extraction

Each sample was homogenized using a mortar and pestle after being dried in an oven at 65°C for 24 hours. DNA extraction was performed in triplicate from each surimi product using a Qiagen DNeasy Blood and Tissue Kit (Qiagen Inc., Hilden, Germany) following the manufacturer's guidelines. Briefly, approximately 20 mg of dried sample was lysed with Buffer ATL containing proteinase K (Qiagen Inc.) at 56°C for 3 hours. The lysates were then mixed with Buffer TL (Omega Bio-tek, Norcross, GA, USA) and ethanol, and the DNA was purified using resin columns.

3.3 PCR amplification

To analyze the fish species based on the extracted DNA sample, the mitochondrial cytochrome oxidase I (COI) gene was selected as the sequencing target. Two sets of universal primers were designed to amplify approximately 740 and 700 base pair (bp) of COI gene fragment, respectively; the sequences are shown in Table 2. PCR was performed using an ABI Verity thermocycler

(Applied Biosystems Corp., Foster City, CA, USA). PCR mixtures were prepared using Takara ExTaq (Takara Bio Inc., Otsu, Japan). Each 20-μl mixture contained 2 μl of Takara ExTaq PCR buffer, 1.6 μl of Takara dNTPs (2.5 mM each), 10 pmol of each primer, 0.5 U of Takara ExTaq DNA polymerase, 10–20 ng of template DNA, and distilled water. The amplification conditions were: initial denaturation at 94°C for 7 minutes followed by 35 cycles of denaturation at 94°C for 1 minute, annealing at 54°C for 1 minute, and extension at 72°C for 1 minute, with a final extension at 72°C for 7 minutes. After amplification, the amplicons were purified using a DNA purification kit (GeneAll, Seoul, South Korea). The quality and length of each amplicon were checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

3.4 Library preparation for metabarcoding

The purified products were prepared for pyrosequencing using a GS FLX Titanium Rapid Library MID Adaptors Kit (454 Life Science, Roche, Basel, Switzerland). The DNA quantity was determined using a PicoGreen DNA assay (Invitrogen, Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions and pooled to equimolar concentrations. The DNA libraries were subjected to pyrosequencing with the 454 GS Junior system using titanium chemistry (Roche, Branford, CT, USA).

3.5 Data pre-processing

Pyrosequencing data were obtained using Genome Sequencer FLX system software v2.3 (Roche) and analyzed using CLC Genomic Workbench 9.0 (CLC Bio, Aarhus, Denmark). The pyrosequencing reads were sorted into individual samples according to the unique sequence tag inserted during library preparation. Sequence reads were trimmed to remove adaptor sequences and low-quality data with a quality score limit of 0.05. The required minimum read length was 250 bp; a max of two ambiguous nucleotides were allowed. After trimming, duplicate entries were deleted based on the 'merge reads' function. Consequently, quality-checked reads were converted into FASTA format for future use.

3.6 Taxonomic assignment

To identify taxa, the reads were compared against MitoFish database v3.08 (http://mitofish.aori.u-tokyo.ac.jp) using BLASTN. Species assignments were considered to be more than or equal to 97% sequence identity at an E-value threshold of 10^{-5}. When the sequence similarity was lower than 97%, they were assigned to the higher taxonomic group (e.g., genus level).

Table 1. Information of surimi products used in this study.

Sample	Production country of surimi	Proportion	Fish species labeled in product
A	America	82.76%	Alaska pollock
B	America	71.56%	Alaska pollock
C	Imported	70.67%	–
D	Imported	64.62%	White flesh fish
E	Imported	54.89%	Croaker
F	Imported	–	–
G	Imported	61.13%	Golden-thread, cutlassfish and croaker
H	America	30.40%	–
I	Korea	–	–
J	Imported	–	–
K	Imported	68.97%	–

"–" indicates that the information was not labeled

Table 2. Primer sequences used for genome sequencer FLX system in this study.

Set	Oligo ID	Oligo sequences (5'→3')	Fragment size
Set 1	1F	TCAGCCATCCTACCTGTGGC	737 bp
	1R	GGGTGGCCGAAGAATCAGAA	
Set 2	2F	CAACCAATCACAAAGACATCGGCAC	703 bp
	2R	ACTTCTGGGTGGCCGAAGAATCAGAA	

4. Results

4.1 Next generation sequencing (NGS) data statistics

In total, 97,619 partial mitochondrial COI sequences were obtained from 11 surimi samples using primer set 1. After pre-processing, including the removal of low-quality reads, the number of high-quality reads totaled 94,183. The average sequence length was 478.65 bp. Using primer set 2, a total of 138,497 reads were obtained, and the number of available reads was 80,686. The average length of the resulting reads was 437.15 bp. The average number of sequence reads was 7,948 per sample. The information obtained for each sample classified by barcode is shown in Tables 3 and 4.

4.2 Taxonomic assignment of primer set 1

After the BLASTN search against the MitoFish database, the validated reads of each sample were assigned to a fish taxonomy. For samples from primer set 1, 88,329 reads were classified at the species level based on more than 1% of the content. The taxonomic composition was also analyzed, and 56 species in 30 families were identified from 11 samples (Table 5 and Table 6). The main raw materials were identified by analyzing the results of the species constituting more than 5% of each sample (Table 7). Sequences of *Gadus chalcogrammus* were found to be most abundant in samples A–C, G, H, and J—especially in samples A, B, and H, where more than 99% of the sequence reads were assigned to *G. chalcogrammus*. Approximately 65% of the sequence reads were *G. chalcogrammus* in samples C, G, and J. In sample D, *Nemipterus bathybius* was

the highest proportion at 36.98%, with *Terapon jarbua* accounting for 9.93%, *Sphyraena chrysotaenia* for 8.79%, and *Ariomma indicum* for 6.04%. In sample E, *Cypselurus hiraii* accounted for 42.44%, *Sphyraena flavicauda* for 16.12%, *Scomber japonicus* for 5.66%, *Exocoetus volitans* for 5.37%, and *Trachurus japonicus* for 5.13%. Sample F contained the least amount of main species. *Decapterus maruadsi* and *Istiophorus platypterus* accounted for 16.49% and 15.56%, respectively, while *Upeneus moluccensis* accounted for 9.66%, *A. indicum* for 6.61%, and *Scomber australasicus* for 5.46%. In sample K, *N. bathybius* accounted for the highest percentage at 76.04%, and the rest was *G. chalcogrammus* at 15.39%.

4.3 Taxonomic assignment with primer set 2

Of the 80,686 validated reads obtained with primer set 2, 76,695 were classified at the species level, and all reads were assigned to one of 32 species in 16 families (Table 5 and Table 8). The highest number of *G. chalcogrammus* sequences were found in samples A–C, G, and H. Samples A, B, and H contained more than 99% *G. chalcogrammus*, while samples C and G contained 82.08% and 67.58%, respectively. In sample D, *Nemipterus randalli* accounted for 56.67%, while *Sarurida umeyoshii* accounted for 16.92%. In sample E, a relatively low number of reads were identified, but the sequences included 28.48% *Larimichthys polyactis*, 18.57% *Upeneus vittatus*, and 16.73% *Trichiurus japonicus*. Sample F consisted of 39.95% *Selar crumenophthalmus*, 18.52% *Saurida undosquamis*, 7.49% *D. maruadsi*, and 6.86% *T. japonicus*. In samples I and J, *Engraulis japonicus* accounted for 42.45% and 68.95%, respectively.

Trichiurus japonicus accounted for 26.53% and *G. chalcogrammus* accounted for 24.96% in sample I, whereas only *G. chalcogrammus* was found in sample J, where it accounted for 27.24% of the sequences. Finally, N. randalli accounted for the highest content in (42.22%) sample K, with *G. chalcogrammus*, *N. bathybius*, and *S. undosquamis* accounting for 28.08%, 6.62%, and 6.59%, respectively (Table 7).

4.4 Comparison of the two primer sets

The ability of the two primer sets to distinguish fish species from each surimi produc was compared; the results are shown in Figures 1 and 2. Similar results were obtained for the two primer sets in samples A–C, G, and H, but there were significant differences in the other samples. In total, 18 and 13 main fish species were identified using primer sets 1 and 2, respectively. However, only three species, *G. chalcogrammus*, *N. bathybius*, and *D. maruadisi*, were identified as the same species by both primer sets. *G. chalcogrammus* was confirmed in eight (A–C and G–K) of the eleven samples, but only in sample I, it was identified when using the primer set 2. *N. bathybius* was identified in sample K by both primer sets, and different species of the same genus were detected in samples D and K by both primer sets. *D. maruadisi* was identified in sample K by both primer sets. *L. polyactis* was not detected in sample E by primer set 1, but it was detected using primer set 2. In samples I and J, *E. japonicus* was detected by primer set 2 only. Many other differences were found between primer sets in these samples.

Table 3. Number of sequences recovered using primer set 1.

Sample	Raw Avg Length	Trim Avg Length	Raw Reads	Trim Reads
A	483.3	466.8	13,131	12,992
B	483.5	467.8	12,243	12,237
C	496.0	479.5	10,190	10,188
D	512.2	497.6	11,495	11,495
E	508.8	490.9	10,854	10,854
F	515.9	498.5	11,053	11,051
G	467.7	472.8	5,209	4,890
H	472.8	466.0	5,987	5,797
I	388.1	476.0	5,742	4,273
J	411.6	464.5	5,528	4,566
K	485.6	484.8	6,187	5,840

Table 4. Number of sequences recovered using primer set 2.

Sample	Raw Avg Length	Trim Avg Length	Raw Reads	Trim Reads
A	196.1	429.0	23,982	6,834
B	334.7	432.1	11,008	6,603
C	353.0	434.1	11,314	7,083
D	384.1	447.5	10,601	6,825
E	387.9	440.2	9,413	6,426
F	403.8	453.3	9,754	6,679
G	399.4	433.7	9,316	6,817
H	385.5	436.2	11,705	8,281
I	389.5	429.0	6,564	4,735
J	341.7	429.6	21,436	12,952
K	321.3	443.9	13,386	7,451

Table 5. Taxonomic assignment and numerical abundance of sequences generated from surimi samples using each primer set. When the sequence similarity was lower than 97%, they were classified as “others”.

Sample	Primer set 1			Primer set 2		
	No.	Scientific name	Reads	No.	Scientific name	Reads
A	1	*Gadus chalcogrammus*	12987	1	*Gadus chalcogrammus*	6827
		Others	5		Others	7
B	1	*Gadus chalcogrammus*	12230	1	*Gadus chalcogrammus*	6555
		Others	7		Others	48
C	1	*Gadus chalcogrammus*	6654	1	*Gadus chalcogrammus*	5814
	2	*Ariomma indicum*	642	2	*Epinephelus akaara*	528
	3	*Hilsa kelee*	500	3	*Saurida umeyoshii*	249
	4	*Nemipterus bathybius*	295	4	*Nemipterus randalli*	105
	5	*Alepes kleinii*	279	5	*Silurus asotus*	96
	6	*Mene maculata*	274		Others	33
	7	*Alepes djedaba*	259			
	8	*Decapterus maruadsi*	256			
	9	*Terapon jarbua*	241			
	10	*Sphyraena chrysotaenia*	161			
	11	*Upeneus vittatus*	152			
		Others	1782			
D	1	*Nemipterus bathybius*	4251	1	*Nemipterus randalli*	3868
	2	*Terapon jarbua*	1142	2	*Saurida umeyoshii*	1155
	3	*Sphyraena chrysotaenia*	1010	3	*Selar crumenophthalmus*	216
	4	*Ariomma indicum*	694	4	*Trichiurus japonicus*	178
	5	*Hilsa kelee*	434	5	*Engraulis japonicus*	172
	6	*Nemipterus japonicus*	383	6	*Plecoglossus altivelis*	139
	7	*Alepes kleinii*	347	7	*Upeneus vittatus*	138
	8	*Scomber australasicus*	336	8	*Pomadasys maculatus*	134
	9	*Alepes djedaba*	315	9	*Priacanthus hamrur*	115
	10	*Gerres filamentosus*	253	10	*Kareius bicoloratus*	78
	11	*Upeneus vittatus*	238	11	*Megalaspis cordyla*	73
	12	*Mene maculata*	215		Others	160
	13	*Nemipterus mesoprion*	206			
	14	*Uranoscopus archionema*	178			
	15	*Chirocentrus nudus*	164			
	16	*Decapterus maruadsi*	129			
	17	*Decapterus macarellus*	125			
	18	*Pomadasys maculatus*	117			

		Others	516			
E	1	*Cypselurus hiraii*	4606	1	*Larimichthys polyactis*	1830
	2	*Sphyraena flavicauda*	1750	2	*Upeneus vittatus*	1193
	3	*Scomber japonicus*	614	3	*Trichiurus japonicus*	1075
	4	*Exocoetus volitans*	583	4	*Saurida undosquamis*	306
	5	*Trachurus japonicus*	557	5	*Selar crumenophthalmus*	274
	6	*Scomber australasicus*	540	6	*Cheilopogon pitcairnensis*	217
	7	*Upeneus moluccensis*	451	7	*Ostorhinchus dispar*	180
	8	*Decapterus maruadsi*	211	8	*Trachurus declivis*	186
	9	*Ostorhinchus dispar*	187	9	*Ostichthys trachypoma*	144
	10	*Lepidotrigla multispinosa*	177	10	*Selar boops*	110
	11	*Upeneus mascareinsis*	123	11	*Upeneus tragula*	85
	12	*Ostichthys japonicus*	116	12	*Engraulis japonicus*	79
	13	*Alepes vari*	109	13	*Epinephelus akaara*	69
	14	*Dactyloptena peterseni*	109		Others	931
		Others	918			
F	1	*Decapterus maruadsi*	1822	1	*Selar crumenophthalmus*	2668
	2	*Istiophorus platypterus*	1719	2	*Saurida undosquamis*	1237
	3	*Upeneus moluccensis*	1067	3	*Decapterus maruadsi*	500
	4	*Ariomma indicum*	731	4	*Trichiurus japonicus*	458
	5	*Carangoides malabaricus*	517	5	*Carangoides malabaricus*	225
	6	*Scomber australasicus*	603	6	*Upeneus tragula*	163
	7	*Alepes vari*	472	7	*Istiophorus platypterus*	146
	8	*Amphiprion percula*	399	8	*Upeneus vittatus*	107
	9	*Decapterus macrosoma*	371	9	*Clupea pallasii pallasii*	81
	10	*Mene maculata*	274	10	*Epinephelus areolatus*	76
	11	*Sphyraena flavicauda*	245	11	*Notropis volucellus*	69
	12	*Dactyloptena peterseni*	231		Others	949
	13	*Selaroides leptolepis*	183			
	14	*Cypselurus hiraii*	170			
	15	*Atule mate*	161			
	16	*Carangoides chrysophrys*	159			
	17	*Sardinella albella*	153			
	18	*Iniistius verrens*	128			
	19	*Istiompax indica*	114			
	20	*Scalicus amiscus*	114			
		Others	1418			
G	1	*Gadus chalcogrammus*	2978	1	*Gadus chalcogrammus*	4607
	2	*Istiompax indica*	221	2	*Selar crumenophthalmus*	415
	3	*Alepes melanoptera*	126	3	*Priacanthus tayenus*	381
	4	*Decapterus maruadsi*	107	4	*Saurida undosquamis*	340
	5	*Seriola dumerili*	105	5	*Decapterus maruadsi*	98
	6	*Lota lota*	65	6	*Scolopsis bimaculata*	77
	7	*Carangoides equula*	61		Others	899
	8	*Nemipterus bathybius*	61			
	9	*Makaira mazara*	58			
	10	*Polydactylus plebeius*	53			

		Others	1055			
H	1	*Gadus chalcogrammus*	5766	1	*Gadus chalcogrammus*	8230
		Others	31		Others	51
I	1	*Biwia zezera*	2600	1	*Engraulis japonicus*	2010
	2	*Rhodeus sericeus*	300	2	*Trichiurus japonicus*	1256
	3	*Nemipterus bathybius*	212	3	*Gadus chalcogrammus*	1182
	4	*Upeneus tragula*	154	4	*Trachurus declivis*	111
	5	*Lutjanus johnii*	98	5	*Decapterus macrosoma*	96
	6	*Amphiprion perideraion*	96		Others	80
	7	*Seriola dumerili*	74			
	8	*Lagocephalus inermis*	59			
		Others	680			
J	1	*Gadus chalcogrammus*	3030	1	*Engraulis japonicus*	8931
	2	*Decapterus maruadsi*	670	2	*Gadus chalcogrammus*	3528
	3	*Jaydia carinatus*	538	3	*Decapterus maruadsi*	197
	4	*Sebastes trivittatus*	67		Others	296
	5	*Lota lota*	56			
		Others	205			
K	1	*Nemipterus bathybius*	4441	1	*Nemipterus randalli*	3146
	2	*Gadus chalcogrammus*	899	2	*Gadus chalcogrammus*	2092
	3	*Seriola dumerili*	219	3	*Nemipterus bathybius*	493
	4	*Decapterus maruadsi*	110	4	*Saurida undosquamis*	491
		Others	171	5	*Engraulis japonicus*	196
				6	*Priacanthus tayenus*	180
				7	*Scolopsis bimaculata*	175
				8	*Decapterus maruadsi*	171
					Others	507

Table 6. Diversity of fish species by primer set 1 inferred using MitoFish database.

Family (30)	Genus (41)	Species (56)
Apogonidae	*Jaydia*	*Jaydia carinatus*
	Ostorhinchus	*Ostorhinchus dispar*
Ariommatidae	*Ariomma*	*Ariomma indicum*
Carangidae	*Alepes*	*Alepes djedaba*
		Alepes kleinii
		Alepes melanoptera
		Alepes vari
	Atule	*Atule mate*
	Carangoides	*Carangoides chrysophrys*
		Carangoides equula
		Carangoides malabaricus
	Decapterus	*Decapterus macarellus*
		Decapterus macrosoma
		Decapterus maruadsi
	Selaroides	*Selaroides leptolepis*
	Seriola	*Seriola dumerili*
	Trachurus	*Trachurus japonicus*
Chirocentridae	*Chirocentirus*	*Chirocentrus nudus*
Clupeidae	*hilsa*	*Hilsa kelee*
	Sardinella	*Sardinella albella*
Cyprinidae	*Rhodeus*	*Rhodeus sericeus*
Dactylopteridae	*Dactyloptena*	*Dactyloptena peterseni*
Gadidae	*Gadus*	*Gadus chalcogrammus*
Gerreidae	*Gerres*	*Gerres filamentosus*
Exocoetidae	*Cypselurus*	*Cypselurus hiraii*
	Exocoetus	*Exocoetus volitans*
Gobioninae	*Biwia*	*Biwia zezera*
Holocentridae	*Ostichthys*	*Ostichthys japonicus*
Haemulidae	*Pomadasys*	*Pomadasys maculatus*
Istiophoridae	*Istiompax*	*Istiompax indica*
	Istiophorus	*Istiophorus platypterus*
	Makaira	*Makaira mazara*
Labridae	*Iniistius*	*Iniistius verrens*
Lotidae	*Lota*	*Lota lota*
Lutjanidae	*Lutjanus*	*Lutjanus johnii*

Menidae	*Mene*	*Mene maculata*
Mullidae	*Upeneus*	*Upeneus mascareinsis*
		Upeneus moluccensis
		Upeneus tragula
		Upeneus vittatus
Nemipteridae	*Nemipterus*	*Nemipterus bathybius*
		Nemipterus japonicus
		Nemipterus mesoprion
Peristediidae	*Scalicus*	*Scalicus amiscus*
Polynemidae	*Polydactylus*	*Polydactylus plebeius*
Pomacentridae	*Amphiprion*	*Amphiprion percula*
		Amphiprion perideraion
Scombridae	*Scomber*	*Scomber australasicus*
		Scomber japonicus
Sebastidae	*Sebastes*	*Sebastes trivittatus*
Sphyraenidae	*Sphyraena*	*Sphyraena chrysotaenia*
		Sphyraena flavicauda
Terapontidae	*Terapon*	*Terapon jarbua*
Tetraodontidae	*Lagocephalus*	*Lagocephalus spadicelus*
Triglidae	*Lepidotrigla*	*Lepidotrigla multispinosa*
Uranoscopidae	*Uranoscopus*	*Uranoscopus archionema*

Table 7. Taxonomic composition of fish species from each sample by two primer sets. Fish species are only shown with a relative abundance of more than 5%.

Sample	Primer set 1			Primer set 2		
	No.	Scientific name	Content	No.	Scientific name	Content
A	1	*Gadus chalcogrammus*	99.96 %	1	*Gadus chalcogrammus*	99.90 %
B	1	*Gadus chalcogrammus*	99.94 %	1	*Gadus chalcogrammus*	99.27 %
C	1	*Gadus chalcogrammus*	65.31 %	1	*Gadus chalcogrammus*	82.08 %
	2	*Ariomma indicum*	6.30 %	2	*Epinephelus akaara*	7.45 %
D	1	*Nemipterus bathybius*	36.98 %	1	*Nemipterus randalli*	56.67 %
	2	*Terapon jarbua*	9.93 %	2	*Saurida umeyoshii*	16.92 %
	3	*Sphyraena chrysotaenia*	8.79 %			
	4	*Ariomma indicum*	6.04 %			
E	1	*Cypselurus hiraii*	42.44 %	1	*Larimichthys polyactis*	28.48 %
	2	*Sphyraena flavicauda*	16.12 %	2	*Upeneus vittatus*	18.57 %
	3	*Scomber japonicus*	5.66 %	3	*Trichiurus japonicus*	16.73 %
	4	*Exocoetus volitans*	5.37 %			
	5	*Trachurus japonicus*	5.13 %			
F	1	*Decapterus maruadsi*	16.49 %	1	*Selar crumenophthalmus*	39.95 %
	2	*Istiophorus platypterus*	15.56 %	2	*Saurida undosquamis*	18.52 %
	3	*Upeneus moluccensis*	9.66 %	3	*Decapterus maruadsi*	7.49 %
	4	*Ariomma indicum*	6.61 %	4	*Trichiurus japonicus*	6.86 %
	5	*Scomber australasicus*	5.46 %			
G	1	*Gadus chalcogrammus*	60.90 %	1	*Gadus chalcogrammus*	67.58 %
				2	*Selar crumenophthalmus*	6.09 %
				3	*Priacanthus tayenus*	5.59 %
H	1	*Gadus chalcogrammus*	99.47 %	1	*Gadus chalcogrammus*	99.38 %
I	1	*Biwia zezera*	60.84 %	1	*Engraulis japonicus*	42.45 %
	2	*Rhodeus sericeus*	7.03 %	2	*Trichiurus japonicus*	26.53 %
				3	*Gadus chalcogrammus*	24.96 %
J	1	*Gadus chalcogrammus*	66.36 %	1	*Engraulis japonicus*	68.95 %
	2	*Decapterus maruadsi*	14.67 %	2	*Gadus chalcogrammus*	27.24 %
	3	*Jaydia carinatus*	11.78 %			
K	1	*Nemipterus bathybius*	76.04 %	1	*Nemipterus randalli*	42.22 %
	2	*Gadus chalcogrammus*	15.39 %	2	*Gadus chalcogrammus*	28.08 %
				3	*Nemipterus bathybius*	6.62 %
				4	*Saurida undosquamis*	6.59 %

Table 8. Diversity of fish species by primer set 2 inferred using MitoFish database.

Family (16)	Genus (24)	Species (32)
Apogonidae	*Ostorhinchus*	*Ostorhinchus dispar*
Carangidae	*Carangoides*	*Carangoides malabaricus*
	Decapterus	*Decapterus maruadsi*
		Decapterus macrosoma
	Megalaspis	*Megalaspis cordyla*
	Nemipterus	*Nemipterus bathybius*
		Nemipterus randalli
	Selar	*Selar boops*
		Selar crumenophthalmus
	Trachurus	*Trachurus declivis*
		Trichiurus japonicus
	Notropis	*Notropis volucellus*
Clupeidae	*Clupea*	*Clupea pallasii pallasii*
Engraulidae	*Engraulis*	*Engraulis japonicus*
Exocoetidae	*Cheilopogon*	*Cheilopogon pitcairnensis*
Gadidae	*Gadus*	*Gadus chalcogrammus*
Haemulidae	*Pomadasys*	*Pomadasys maculatus*
Holocentridae	*Ostichthys*	*Ostichthys trachypoma*
Istiophoridae	*Istiophorus*	*Istiophorus platypterus*
Nemipteridae	*Scolopsis*	*Scolopsis bimaculata*
Plecoglossidae	*Plecoglossus*	*Plecoglossus altivelis*
Pleuronectidae	*Kareius*	*Kareius bicolaratus*
Priacanthidae	*Priacanthus*	*Priacanthus hamrur*
		Priacanthus tayenus
	Saurida	*Saurida umeyoshii*
		Saurida undosquamis
Serranidae	*Epinephelus*	*Epinephelus akaara*
		Epinephelus areolatus
	Larimichthys	*Larimichthys polyactis*
Siluridae	*Silurus*	*Silurus asotus*
Mullidae	*Upeneus*	*Upeneus tragula*
		Upeneus vittatus

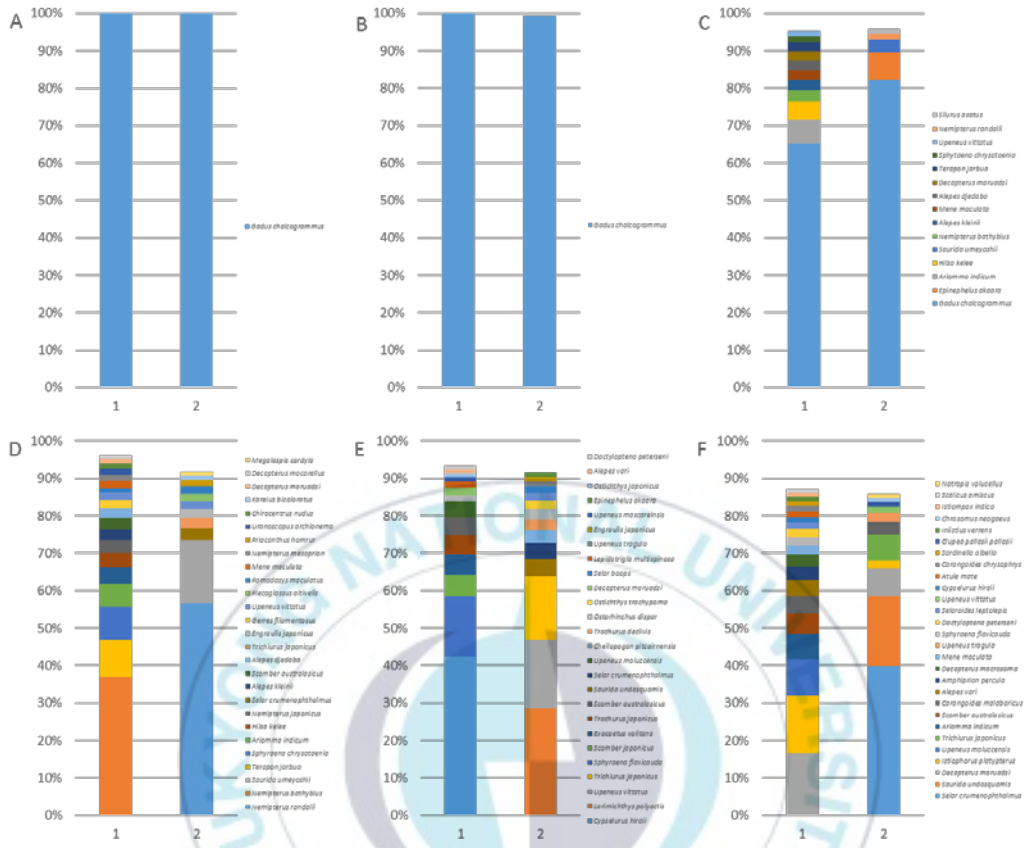

Figure 1. The composition and taxonomic affiliations of each sample by two primer sets (1). Fish species are only shown with a relative abundance of more than 1%.

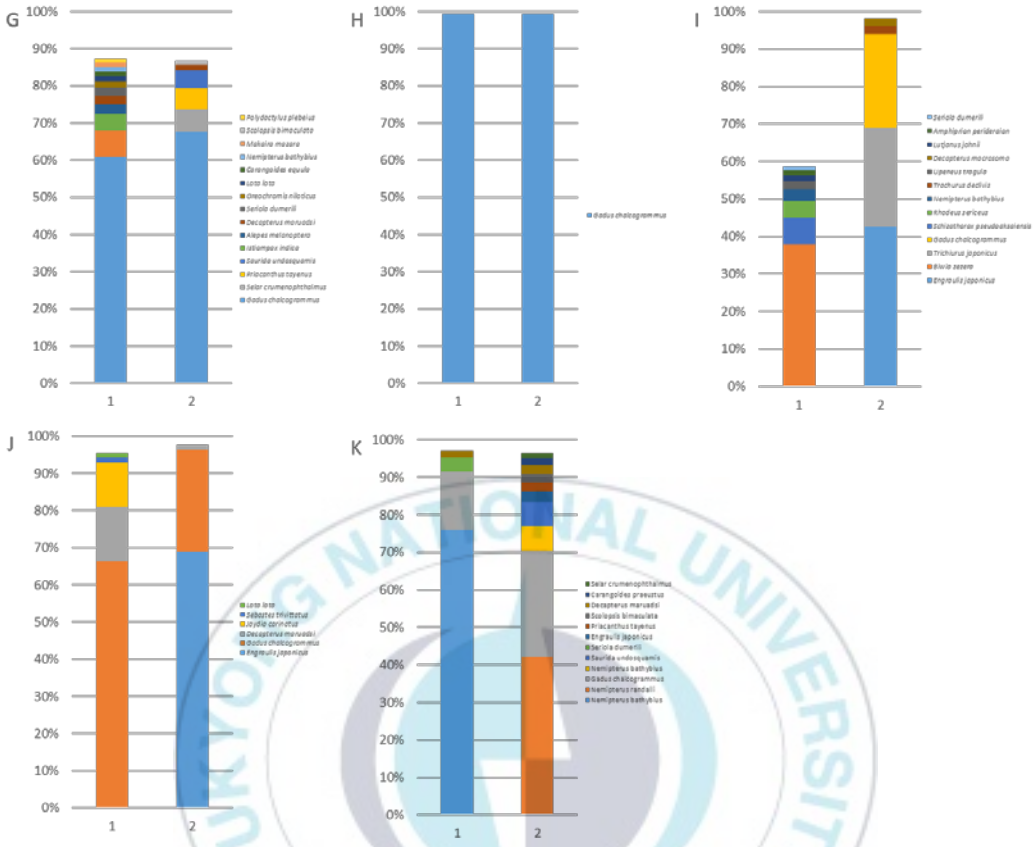

Figure 2. The composition and taxonomic affiliations of each sample by two primer sets (2). Fish species are only shown with a relative abundance of more than 1%.

5. Discussion

Metarbarcoding using NGS is much faster and more reliable than morphological analyses in complex environments; indeed, it allows for the automatic processing of multiple samples simultaneously. Moreover, this technique can identify a variety of species that cannot be distinguished with the naked eye. Due to these advantages, metabarcoding can be used as a substitute tool for the identification of food ingredients. The present study revealed some methodological limitations to the use of metabarcoding in foods.

The high level of genetic variability in the COI region is a problem when designing PCR primers within the barcode region. Several studies, including mini-barcode (Meusnier et al., 2008), have designed universal primers for COI fragments, but these have been difficult to apply to entire taxonomic groups. Another problem is the incompleteness of the reference sequence database. The reference sequence in the MitoFish database used in this study includes 1,324 complete mitochondrial genome sequences and 2,953 partial COI sequences. However, discovered fish species include 27,977 species in 1,827 genera (Nelson, 2006). Although there are relatively few known edible fish species, supplementation of this database is needed for taxonomic coverage.

The overall purpose of this study was to validate the efficacy of the metabarcoding approach, which is commonly used in environmental studies of seafood products. The diversity and content of species used as raw materials have been successfully analyzed. These results are consistent with those of previous studies in that primer selection can affect species detection rates. In this study, only one primer set identified *L. polyactis*, *T. japonicus*, and *E. japonicus*. In the

previous metabarcoding approach, the reason for species detection failures was likely primer specificity (Kelly et al., 2014). To determine the availability of primer sets, optimal amplification conditions should be studied, as well as the probability of individual PCRs and false negatives due to primer mismatches.

A comparison of primer and target sequences revealed one to four mismatches in the forward primer and two mismatches in the reverse primer of primer set 1. For primer set 2, there were two to four mismatches in the forward primer and two to three mismatches in the reverse primer (data not shown). However, primer mismatch alone cannot explain the failure in species discrimination. One hypothesis is that DNA damage occurred in the forward primer-binding site of primer set 1. Each reverse primer was designed at the same position, so the reverse primer was not the source of the problem. The forward primer in set 1 was located about 30 bp upstream of that in set 2. Generally, food processing can cause DNA damage, which can result in an inability to amplify DNA from several species.

Unfortunately, it is difficult to determine whether primer set 2 is more suitable for species identification in mixed seafood products. The species diversity detected using primer sets 1 and 2 included 56 and 32 species, respectively. Therefore, the use of various primers is required to confirm the diversity of raw materials from mixed seafood products, and new primers that consider the possible DNA degradation caused by food processing such as boiling and frying are needed. However, despite these difficulties, metabarcoding has tremendous potential for analyzing mixed seafood raw materials. Furthermore, several species could be confirmed by amplifying sequences longer than 650 bp in processed food.

Moreover, these results illustrate the potential of metabarcoding for food safety management. Samples A and B were labeled as premium surimi products using only Alaskan Pollock. The results show that more than 99% of samples A and B were *G. chalcogrammus*. In addition, sample H was primarily *G. chalcogrammus*. *G. chalcogrammus* is a premium raw material for making surimi, and these products do meet standards. The raw materials for these products were imported from the United States, and given the fact that they were manufactured at the same time, the raw materials are assumed to be the same. Sample E was labeled as containing 9% croaker, and this was confirmed in the present study. Although there was an error in the content rate, this error may be due to problems with the amplification or DNA extraction of each fish species.

More than half of the products used numerous species as raw materials; such products are rated as relatively low quality compared to products using a single species. There are several reasons why various fish species are identified in surimi products. Firstly, contamination during surimi manufacturing process can be a problem. Various seafood products are manufactured in the same manufacturing plant and several fish species can be mixed if they are not cleaned properly (Sampels, 2015). Secondly, pure fish pellets were not used for manufacture of surimi products. The production of surimi requires the use of clean pellets, but sometimes fish that do not have intestine removed are used (Ueki et al., 2016). Lastly, a variety of fish can actually be mixed and used to make surimi products. Commonly, these surimi products are made by Southeast Asian nations. However, most surimi products do not precisely label their raw materials, so it is difficult for consumers to identify and purchase what they desire.

The development of metabarcoding for seafood should not be impeded by methodological challenges; however, more experimental research is needed. Here, new primer and modified universal primer were used for metabarcoding studies and showed how these primers can be used for metabarcoding analyses of mixed seafood products.

6. References

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J. C., Bonin, A., Taberlet, P., & Pont, D. (2016). Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLoS One, 11*(6), e0157366.

Cutarelli, A., Amoroso, M. G., Roma, A. D., Girardi, S., Galiero, G., Guarino, A., & Corrado, F. (2013). Italian market fish species identification and commercial frauds revealing by DNA sequencing. *Food Control, 37,* 46-50.

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., Mahon, A. R., Pfrender, M. E., Lamberti, G. A., & Lodge, D. M. (2015). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Mol. Ecol. Resour, 16*(1), 29-41.

Hänfling, B., Handley, L. L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R. C., Oliver, A., & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Mol. Ecol, 25*(13), 3101-3119.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *P. Roy. Soc. Lond. B Bio, 270*(1512), 313-321.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Resour, 7*(4), 544-548.

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using

environmental DNA to census marine fishes in a large mesocosm. *PLoS One, 9*(1), e86175.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *P. Natl. Acad. Sci. USA, 86*(16), 6196-6200.

Leray, M., author, E., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool, 10*, 34.

Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics, 9*(1), 214.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. Open. Sci, 2*, 150088.

Nelson, J. S. (2006). Fishes of the world (4th ed.). Hoboken, N.J.: John Wiley.

Rasmussen, R. S., & Morrissey, M. T. (2008). DNA-based methods for the identification of commercial fish and seafood species. *Compr. Rev. Food Sci. F, 7*(3), 280-295.

Sampels, S. (2015). The effects of processing technologies and preparation on the final quality of fish products. *Trends Food Sci. Tech, 44*(2), 131-146.

Shaw, J. L. A., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., & Coopera, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biol. Conserv, 197*, 131-138.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol, 21*(11), 2565-2573.

Tobe, S. S., Kitchener, A. C., & Linacre, A. M. T. (2010). Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome b and cytochrome oxidase subunit I mitochondrial genes. *PLoS One, 5*(11), e14156.

Ueki, N., Wan, J., & Watabe, S. (2016). Deterioration of white croaker (*Pennahia argentata*) meat thermally-induced gel products caused by proteolytic enzymes in the contaminated intestine and kidney. *Food Chem, 199*, 416-422

Wong, E. H. K., & Hanner, R. H. (2008). DNA barcoding detects market substitution in North American seafood. *Food Res. Int, 41*(8), 828-837.

Xia, Y., Gu, H. F., Peng, R., Chen, Q., Zheng, Y. C., Murphy, R. W., & Zeng, X. M. (2011). COI is better than 16S rRNA for DNA barcoding Asiatic salamanders (Amphibia: Caudata: Hynobiidae). *Mol. Ecol. Resour, 12*(1), 48-56.

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., & Miy, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Sci. Rep, 7*, 40368.

Zaiko, A., Samuilovienea, A., Ardura, A., & Vazquez, E. G. (2015). Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Mar. Pollut. Bull, 100*(1), 53-59.

Chapter III.

Development of Primer Set for the Identification of Fish Species in Surimi Products using Denaturing Gradient Gel Electrophoresis

Development of Primer Set for the Identification of Fish Species in Surimi Products using Denaturing Gradient Gel Electrophoresis

1. Abstract

The purpose of this study was to develop a primer set for the identification of fish species in processed surimi products. Primer set was designed based on the mitochondrial cytochrome c oxidase subunit I gene, and fish species in surimi products were identified using a molecular fingerprinting technique, denaturing gradient gel electrophoresis (DGGE); the results were subjected to sequence-based analysis. The DGGE profiles indicated the presence in surimi products of a greater diversity of fish species than reported previously: 20 species belonging to 16 genera were identified. Therefore, this method facilitates the simple and rapid detection and identification of fish species in seafood products produced from minced fish.

Keyword: Surimi; denaturing gradient gel electrophoresis; identification; universal primer

2. Introduction

Surimi is a processed seafood that is simple to prepare at low cost (Yin et al., 2014). Imitation crab meat, which is made from white fish, such as pollock and cod, is an example of a surimi product. Alaska pollock (*Gadus chalcogramma*) is a major raw material for surimi (Poowakanjana & Park, 2013). Because of the decline in the Alaska pollock catch rate from 250,000 MT in 2003 to about 125,000 MT in 2010, other fish species have been considered for surimi production (Poowakanjana & Park, 2013). Consequently, Pacific whiting (*Merluccius productus*), northern blue whiting (*Micromesistius poutassou*), southern Blue whiting (*Micromesistius autralis*), atka mackerel (*Pleurogrammus azonus*), threadfin bream (*Nemipterus sp.*) and jack mackerel (*Trachurus murphy*) are now in use as raw materials for production of surimi (Park, 2005).

Food companies and consumers are focused on the safety and quality of food, and surimi quality is influenced by the type of fish included (Shiku et al., 2004). In general, whiteness and texture of white flesh fish result in a high-quality product with high-protein and low-fat (Martin-Sanchez et al., 2009). Therefore, identifying the fish species in surimi is important for quality assurance. Some companies seek to make a profit by replacing higher-priced with lower-priced fish (Keskin & Atar, 2012). Indeed, substituting expensive fish species with those of lower cost is easy to use for unfair profits and illegal sales such as fraudulent labeling because consumers are unable to identify the fish species in surimi products (Huxley-Jones et al., 2012). A study of fish fillets reported that lower-cost fish species were used in place of those specified on the label (Pinto et al., 2015). According to the U.S. Food and Drug Administration and the European Union guidelines, the most important factor for seafood quality control is

identification of the fish species therein (Galal-Khallaf et al., 2016; Keskin & Atar, 2012). Rapid and accurate identification of fish species is, therefore, essential for food safety management. Additionally, the profit motive and rapid increases in demand have resulted in overfishing and in various fish species becoming endangered (Galal-Khallaf et al., 2016). Therefore, the identification of the fish species in surimi is required for the management of overfishing and the conservation of endangered species.

DNA-based analysis is required for the identification of fish species in surimi, as it is virtually impossible to distinguish fish species based on their morphology (Huxley-Jones et al., 2012). Several recent studies have employed molecular techniques to identify the fish species in processed seafood products (Galal-Khallaf et al., 2016; Keskin & Atar, 2012; Pinto et al., 2015; Zhao et al., 2013). However, because those studies were focused on identifying only a single species, the methods are unsuitable for use with surimi.

Denaturing gradient gel electrophoresis (DGGE) uses the reduction in electrophoretic mobility according to the denaturing characteristics of PCR products to facilitate their in-gel separation (Muyzer & Smalla, 1998). Therefore, DGGE enables the identification of the various fish species in a minced fish product (Boon et al., 2002; Muyzer et al., 1993). DGGE analysis has been adapted by environmental microbiologists for bacterial community characterization (Griffiths et al., 2000; Muyzer, 1999). This method has also been widely used to characterize microbial community diversity and composition as well as structural changes in various environments, including foodstuffs (Arcuri et al., 2013; Ercolini, 2004; Hong et al., 2007). However, to date, no study has applied DGGE to identify the fish species in a minced foodstuff.

The aims of the present study were to design DGGE primers targeting the cytochrome c oxidase subunit I (COI) gene to identify the fish species in surimi products. This study is the first attempt to analyze fish species using a DGGE method.

3. Materials and methods

3.1 Sample preparation

Twenty surimi products were purchased from markets in Busan, South Korea in February 2015 and transported to the laboratory under refrigerated conditions. Sample information is provided in Table 1.

3.2 DNA extraction

Dried surimi product (200 mg) was subjected to DNA extraction in triplicate using a Qiagen DNeasy Blood & Tissue Kit (Qiagen, Valencia, California) according to the manufacturer's instructions. DNA was quantified using the NanoVue system (GE Healthcare Europe, Munich, Germany), and triplicate samples were pooled into a single sample.

3.3 Primer design and optimization

Whole-length mitochondrial sequences from five fish species used in surimi products—*Nemipterus virgatus* (KR701906), *Gadus chalcogramma* (AB094061), *Micromesistius poutassou* (FR751401), *Pleurogrammus azonus* (AB744047), and *Trachurus japonicus* (AP003092)—were obtained from the GenBank database at the NCBI (http://www.ncbi.nlm.gov) and were aligned using the ClustalW software in the BioEdit platform version 7.0.9.0 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html, Hall 1999) (Table 2). The COI gene was used as the target for identification of fish species. Design of the primers took into consideration the GC content, nucleotide composition at the 3' end, melting temperature, secondary structure, product size, and coverage of fish species. A primer pair targeting a 213-bp region of the COI gene was designed (Table 3). Gradient PCR was

performed at 46, 48, 50, 52, 54, and 56°C to identify the optimum amplification temperature. The PCR products were separated in a 1% agarose gel and visualized using 1× Redsafe Nucleic Acid Staining Solution (iNtRON, South Korea).

3.4 Primer test

To evaluate the coverage of the primer pair, 152 genomic DNA samples from 76 fish species of 28 families were obtained from the Fisheries' Genetic Resources Management Center (National Institute of Fisheries Science, South Korea); these are presented in the Table 4. The reference sequences were aligned and compared with those of the ShortFish-F and ShortFish-R primers. Reference DNA was amplified by PCR with the designed primers and a 46 to 56°C temperature gradient. A neighbor-joining tree was constructed using the MEGA 5.0 software to evaluate the resolution of the 213-bp amplicons.

3.5 DGGE-PCR amplification

To increase primer specificity, the touchdown PCR method was used. A GC-clamp (CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G**)** was attached to the 5' end of the forward primer to stabilize the PCR product during DGGE analysis (Ferris, Muyzer, & Ward, 1996). The 20 μl PCR reaction volume contained 1× Ex Taq buffer with 1.5 mM $MgCl_2$, 0.2 mM dNTP mixture, 0.5 μM each primer, 0.5 units TaKaRa Ex Taq polymerase (TaKaRa Shuzo, Shiga, Japan), and 1 μl of template DNA (50 ng/μl). The touchdown thermocycling conditions were: initial denaturation at 94°C for 7 min using 35 cycles of amplification comprising denaturation at 94°C for 1 min, annealing at

52°C to 48°C for 1 min (the annealing temperature decreased by 1°C every five cycles and was then maintained at 48°C), and extension at 72°C for 1 min; this was followed by a final extension at 72°C for 7 min. Negative controls without a DNA template were included in each analysis. PCR products were visualized by electrophoresis in a 1% agarose gel, followed by visualization with 1× Redsafe Nucleic Acid Staining Solution (iNtRON, South Korea).

3.6 DGGE analysis

PCR products were purified using a GeneAll Expin PCR SV Kit (GeneAll Biotechnology Co., South Korea). Purified DNA was analyzed using the DCode Mutation Detection System (Bio-Rad, Hercules, USA) in 8% (w/v) polyacrylamide gels with denaturing gradients of 20% to 50% and 30% to 60% (100% denaturing solution contained 7 M urea and 40% formamide). Electrophoresis was performed at 20 V for 10 min and then at 80 V and 60°C for 14 h in 1× TAE buffer. After washing with distilled water, the gel was stained with 2× SYBR gold (Invitrogen, USA) for 30 min and imaged using the Molecular Imager Gel Doc System (ATTO E-graph, TaKaRa, Japan).

3.7 Identification of DGGE bands

For taxonomic classification, 30 DGGE bands were isolated from the gel and identified by re-amplification, sequencing, and sequence comparison (band numbers and their taxonomic positions are shown in Fig. 1 and Table 5, respectively). The gel segments were placed into 1.5-ml tubes, resuspended in 100 μl of distilled water, and then stored at 4°C overnight. Extracted DNA was used as a template for re-amplification with the ShortFish-F (lacking the GC

clamp) and ShortFish-R primers in a PCR reaction volume of 20 μl. The reaction temperature profile consisted of denaturation at 96°C for 2 min, followed by 25 cycles of denaturation at 96°C for 15 s, annealing at 50°C at 5 s, and extension at 60°C for 2 min. The PCR fragments were next separated on an ABI 3500 Genetic Analyzer (Applied Biosystems). The PCR products were subjected to Sanger sequencing using the BigDye Terminator v. 1.1 chemistry (Applied Biosystems). Sequencing reactions were carried out using 1 μl of purified PCR product, 2 μl of 5× BigDye Terminator v. 1.1 sequencing buffer, 0.8 μl of BigDye terminator reaction mix v. 1.1, and 1 μl of 0.1 pM ShortFish-F forward primer in a total volume of 10 μl.

Table 1. Characteristics of surimi products used in this study.

Sample	Production country of surimi	Proportion	Fish species labeled in product
A	imported	80.96 %	–
B	Pakistan	64.81 %	–
C	China	54.84 %	Cutlassfish
D	imported	65.34 %	–
E	imported	66.51 %	–
F	imported	66.84 %	–
G	imported	80.69 %	White flesh fish
H	imported	–	–
I	imported	66.51 %	–
J	imported	80.30 %	Horse mackerel 21.41%
K	imported	72.90 %	Croaker
L	imported	64.65 %	Cutlassfish and golden-thread
M	imported	64.65 %	Cutlassfish and golden-thread
N	imported	64.65 %	Cutlassfish and golden-thread
O	imported	62.00 %	–
P	imported	64.36 %	–
Q	imported	78.40 %	–
R	imported	81.05 %	–
S	imported	50.04 %	–
T	imported	54.16 %	–

"–" indicates that the information was not labeled

Table 2. Alignment of primer regions with complete mitochondrial sequences of reference species.

	5641	5642	5643	5644	5645	5646	5647	5648	5649	5650	5651	5652	5653	5654	5655	5656	5657	5658	5659
N. virgatus	C	A	C	A	A	A	G	A	C	A	T	C	G	G	C	A	C	C	C
G. chalcogramma	C	A	C	A	A	A	G	A	C	A	T	T	G	G	C	A	C	C	C
M. poutassou	C	A	C	A	A	A	G	A	C	A	T	T	G	G	C	A	C	C	C
P. azonus	C	A	C	A	A	A	G	A	C	A	T	T	G	G	C	A	C	C	C
T. japonicus	C	A	C	A	A	A	G	A	C	A	T	C	G	G	C	A	C	C	C

	5834	5835	5836	5837	5838	5839	5840	5841	5842	5843	5844	5845	5846	5847	5848	5849	5850	5851	5852	5853
N. virgatus	G	G	C	G	G	G	T	T	C	G	G	A	A	A	C	T	G	A	C	T
G. chalcogramma	G	G	A	G	G	C	T	T	T	G	G	A	A	A	C	T	G	A	C	T
M. poutassou	G	G	C	G	G	C	T	T	C	G	G	G	A	A	C	T	G	A	C	T
P. azonus	G	G	C	G	G	T	T	T	C	G	G	G	A	A	C	T	G	A	C	T
T. japonicus	G	G	A	G	G	C	T	T	T	G	G	A	A	A	C	T	G	A	C	T

Table 3. Universal primers designed from conserved region of COI gene.

Primer name	Sequence	Fragment size	Developed for
ShortFish-F	5'- CACAAAGACATTGGCACCC -3'	213 bp	Entire fish
ShortFish-R	5'- AGTCAGTTTCCGAACCCTCC -3'		

4. Results and discussion

Food-processing conditions, such as temperature and pH, may result in the degradation of DNA (Gryson, 2010). Amplifying >200-bp DNA fragments from surimi products is problematic because of DNA degradation during storage and processing (e.g., frying, boiling, and broiling) (Galal-Khallaf et al., 2016). Although shorter DNA fragments include less information, analysis of degraded DNA is needed to identify the raw materials in processed food and develop novel markers for DNA amplification. In general, ~100-bp DNA sequences yield only 90% species resolution (Meusnier et al., 2008). Consequently, a novel, higher-resolution DNA marker is required. The mitochondrial COI gene has been shown to be a robust genetic marker in systematic phylogenetic research (Hebert et al., 2003; Hebert et al., 2003; Smith et al., 2008). Therefore, the conserved regions of various COI sequences were compared to develop the forward primer ShortFish-F (5'- CAC AAA GAC ATT GGC ACC C -3') and reverse primer ShortFish-R (5'- AGT CAG TTT CCG AAC CCT CC -3') (Table 3).

The coverage of the ShortFish-F/ShortFish-R primer pair was tested using sequence matching. Some species showed mismatched sequences comprising 0–3 bases compared with the corresponding positions of the COI region of the total of 76 species. However, as the mismatched bases were not located at the 3' end, the primer was tolerant of mismatched sequences (Neff et al., 2002). The sequence of the reverse primer ShortFish-R was also compared with reference sequences. Zero to four mismatched bases were identified in the reverse primer, but the 3' end sequences matched perfectly; therefore, this primer was also tolerant of mismatched sequences. Gradient PCR using reference DNA indicated that the original 213-bp DNA fragments could be amplified from all samples at annealing

temperatures of 48–52°C. The phylogenetic analysis showed that most species could be distinguished using short sequences, with the exception of five species in two families—Carangidae (*Gadus* sp. and *Trachiotus* sp.) and Tetraodontidae (*Takifugu* sp.). Each family constitutes a clade with 100% similarity (Fig. 2). This finding supports the results of a previous study that made use of short sequences to examine the genetic relationships among five major monophyletic clades. The whole-length COI sequence enabled identification of 97% of species, and 95% and 90% resolutions were reported for 250-bp and 100-bp sequences, respectively (Meusnier et al., 2008). In this study, the 213-bp sequence yielded a 93% identification rate. Therefore, this novel biomarker is suitable for the identification of fish species.

Genomic DNA from the surimi products was used as a template for DGGE-PCR. PCR amplification using the primers GC-ShortFish-F (5'- CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GCA CAA AGA CAT TGG CAC CC -3')/ShortFish-R was performed using a touchdown cycling protocol. Twenty PCR products were separated by DGGE in 8% polyacrylamide gels. The gel with a 20% to 50% denaturing gradient showed a resolution superior to that of the gel with a 30% to 60% denaturing gradient (data not shown).

Twenty samples showed diverse DGGE fingerprints; however, several showed similar or identical patterns (A-H, J, O, P, R and S in Fig. 1) and were clustered in groups, the members of which shared raw materials. Therefore, this analysis could be used to differentiate raw materials. The average number of DGGE bands was 10. Based on the position of each band, a total of 30 bands (6–15 per lane) were identified (Fig. 1).

The DGGE bands facilitated species identification using reference sequences in the GenBank database (Table 5). The results revealed that surimi products comprise a considerable variety of fish species (e.g., *Nemipterus virgatus*, *Gadus chalcogramma*, *Micromesistius poutassou*, *Pleurogrammus azonus*, and *Trachurus japonicas*). The sequences of 30 separate bands corresponded to 20 fish species. Seven species were affiliated with two or three sequences (17 in total): *Nemipterus japonicas* (DGGE band nos. 14 and 15), *Trachurus japonicus* (nos. 2, 3, and 4), *Gadus chalcogrammus* (nos. 12 and 13), *Trichiurus lepturus* (nos. 17 and 18), *Selar crumenophthalmus* (nos. 19, 20, and 21), *Megalaspis cordyla* (nos. 23 and 24), and *Saurida tumbil* (nos. 27, 28, and 29). Therefore, 20 fish species could be detected by DGGE using the primer set designed in this study. The results indicate the presence of a variety of fish species in surimi products.

Band 14, which was *Nemipterus japonicus*, Japanese threadfin bream, was present in all surimi samples. This species is a common ingredient in surimi products sold in Korea, but its proportion differed among the samples, as indicated by the variation in band intensity. Because DGGE is a semi-quantitative method, the abundance of specific fish species in surimi products can be estimated (Cani, 2013; Muyzer & Waal, 1994).

Band 1 *(Nemipterus randalli*) was present in all surimi samples, with the exception of B and S. *Nemipterus* spp. are important in the fishery industry; their catches ranked 20th in 2011 and 2012 (Moffitt & Cajas-Cano, 2014). Another *Nemipterus* sp., *N. bipunctatus*, was also detected, but its band position in other samples was not clear.

Most bands were present in only one sample. For example, bands 2, 3, 4, and 5 were present only in sample J, and they were affiliated with the genus *Trachurus*. Bands 12 and 13 were present in samples O, Q, R, and T, and these formed a distinct pattern including 30 unique ingredients.

In some cases, the same species occupied different band positions, as has been reported previously (Gonzalez-Arenzana et al., 2013; Hu et al., 2009). This was likely due to intraspecies heterogeneity, such as with regard to haplotype (Case et al., 2007). In this study, all bands were identified using PCR-cloning, and the major disadvantage of DGGE (Van-Moreira et al., 2013) (i.e., a single band for multiple organisms) was not encountered. This was probably because of the presence of few DNA sequences in a single sample, which differs from other studies of microbial diversity. The detection sensitivity of DGGE has been reported to be 1–2% of the total population (Kan et al., 2006); however, this is unimportant for identifying the major fish species used in surimi.

Due to the increasing demand for fish products, determination of the species composition has become an important issue for both consumers and food regulatory authorities. The novel primer set and DGGE method developed in this study enabled the semi-quantitative identification of fish species (Andorra et al., 2010). Therefore, this DGGE method is suitable for a variety of applications, such as detection of various ingredients for make fish feed. In conclusion, the method developed for the rapid identification of fish species that will enhance the ability to control the quality of minced foodstuffs.

Table 4. List of 76 species obtained from the NIFS.

No	Scientific name	Region	Accession No
1	*Eptatretus stoutii*	COI	GU440317
2	*Himantura* sp.	COI	JX263423
3	*Engraulis japonicus*	COI	JF952723
4	*Ilisha elongate*	COI	HM030780
5	*Sardinops melanostictus*	COI	JQ266230
6	*Gadus macrocephalus*	COI	JQ354101
7	*Gadus chalcogrammus*	COI	JF952737
8	*Lophius litulon*	COI	EU660706
9	*Cololabis saira*	COI	JQ354059
10	*Sebastes fasciatus*	COI	KC015912
11	*Sebastes alutus*	COI	HQ712757
12	*Pleurogrammus monopterygius*	COI	JQ354278
13	*Anthias nicholsi*	COI	JQ774959
14	*Acanthistius patachonicus*	COI	EU074304
15	*Branchiostegus albus*	COI	EU861053
16	*Carangoides equula*	COI	AY541645
17	*Trachurus japonicus*	COI	JF952880
18	*Pagrus major*	COI	GU207340
19	*Chrysophrys auratus*	COI	DQ107829
20	*Pagrus caeruleostictus*	COI	JN868714
21	*Larimichthys polyactis*	COI	HQ385794
22	*Larimichthys croces*	COI	FJ595214
23	*Pseudotolithus elongates*	COI	KF965495
24	*Pseudotolithus typus*	COI	KF965520
25	*Miichthys miiuy*	COI	JQ738461
26	*Micropogonias undulates*	COI	JQ841936
27	*Micropogonias furnieri*	COI	GU225148
28	*Atrobucca* sp.	COI	JF492920
29	*Atractoscion* sp.	COI	DQ107824
30	*Trichiurus japonicus*	COI	JN990871
31	*Trichiurus* sp.	COI	JX124916
32	*Scomber scombrus*	COI	AB120717
33	*Paralichthys isosceles*	COI	JQ365476
34	*Limanda aspera*	COI	JX183913
35	*Cynoglossus senegalensis*	COI	EU513631
36	*Cynoglossus lingua*	COI	KF965355
37	*Cynoglossus arel*	COI	KF965470

38	*Cynoglossus macrolepidotus*	COI	KF965350
39	*Cynoglossus bilineatus*	COI	KF965375
40	*Mustelus mosis*	COI	HQ149887
41	*Mustelus asterias*	COI	KJ205083
42	*Okamejei acutispina*	COI	EU334812
43	*Himantura gerrardi*	COI	JF493648
44	*Himantura astra*	COI	DQ108157
45	*Himantura undulata*	COI	JX263423
46	*Muraenesox bagio*	COI	JN021234
47	*Sardinella aurita*	COI	JQ266230
48	*Sardinella maderensis*	COI	JQ266230
49	*Helicolenus barathri*	COI	DQ108056
50	*Sebastes ciliates*	COI	KF930415
51	*Platycephalus indicus*	COI	HM180794
52	*Lateolabrax maculatus*	COI	JQ343911
53	*Epinephelus septemfasciatus*	COI	FJ594966
54	*Ephinehelus fuscoguttatus*	COI	HQ174861
55	*Chloroscombrus chrysurus*	COI	KP641366
56	*Trachinotus anak*	COI	KJ642220
57	*Trachurus novaezelandiae*	COI	EF609485
58	*Brama brama*	COI	EU074367
59	*Nemipterus bipunctatus*	COI	JQ350137
60	*Macrospinosa cuja*	COI	JX260908
61	*Pseudotolithus brachygnathus*	COI	JF494251
62	*Pseudotolithus* sp.	COI	DQ885031
63	*Otolithes ruber*	COI	JF494030
64	*Sciaenops ocellatus*	COI	KF461230
65	*Trichiurus gangeticus*	COI	FJ265828
66	*Scomber japonicus*	COI	JQ738502
67	*Lepidopsetta polyxystrapolyxystra*	COI	HQ712518
68	*Cynoglossus monodi*	COI	EU513629
69	*Ephippion guttifer*	COI	KJ093731
70	*Lagocephalus gloveri*	COI	JQ681796
71	*Lagocephalus guentheri*	COI	KF442241
72	*Lagocephalus wheeleri*	COI	EU595161
73	*Takifugu chinensis*	COI	AP009534
74	*Takifugu pseudommus*	COI	AP009534
75	*Takifugu rubripes*	COI	HM102315
76	*Takifugu xanthopterus*	COI	JQ681824

Table 5. Sequence analysis of bands excised from DGGE gels.

Band	Nearest match	Homology	Size	Accession no.
1	*Nemipterus randalli*	100%	172	KM538438
2	*Trachurus japonicus*	99%	174	KC970408
3	*Trachurus japonicus*	98%	174	KP267655
4	*Trachurus japonicus*	98%	174	KP267655
5	*Trachurus novaezelandiae*	95%	162	KM006769
6	*Oreochromis niloticus*	100%	165	KT307783
7	*Branchiostegus argentatus*	100%	165	KP267650
8	*Dactyloptena orientalis*	99%	174	FJ583314
9	*Scolopsis taenioptera*	100%	164	KF809419
10	*Lutjanus bengalensis*	96%	166	EU600136
11	*Pennahia macrocephalus*	99%	165	KP722759
12	*Gadus chalcogrammus*	99%	174	KT321137
13	*Gadus chalcogrammus*	99%	174	KT321137
14	*Nemipterus japonicus*	99%	174	KF009634
15	*Nemipterus japonicus*	97%	172	KF009634
16	*Nemipterus bipunctatus*	97%	165	JQ350137
17	*Trichiurus lepturus*	99%	168	JQ681420
18	*Trichiurus lepturus*	95%	174	JN242479
19	*Selar crumenophthalmus*	97%	164	JF494494
20	*Selar crumenophthalmus*	99%	173	JF494494
21	*Selar crumenophthalmus*	97%	164	JF494494
22	*Mene maculate*	95%	174	KJ202178
23	*Megalaspis cordyla*	98%	164	KR011052
24	*Megalaspis cordyla*	98%	174	KR011052
25	*Decapterus maruadsi*	96%	174	JX261479
26	*Saurida undosquamis*	99%	166	KP266852
27	*Saurida tumbil*	99%	172	KP267628
28	*Saurida tumbil*	97%	174	KM459006
29	*Saurida tumbil*	98%	170	KM459006
30	*Larimichthys polyactis*	96%	160	JQ738596

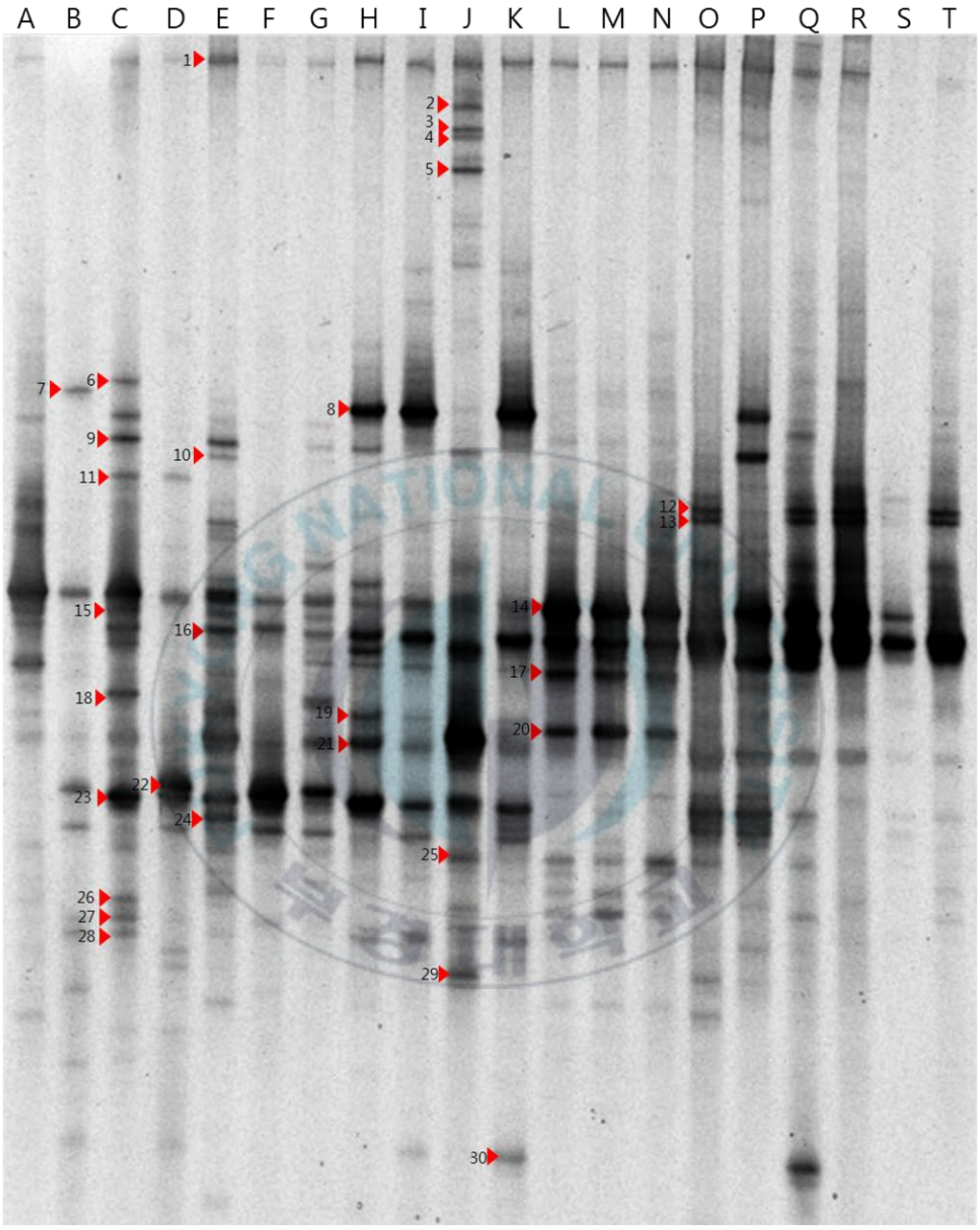

Figure 1. DGGE profiles of the mitochondrial COI region in 20 surimi products. Sequenced bands are numbered.

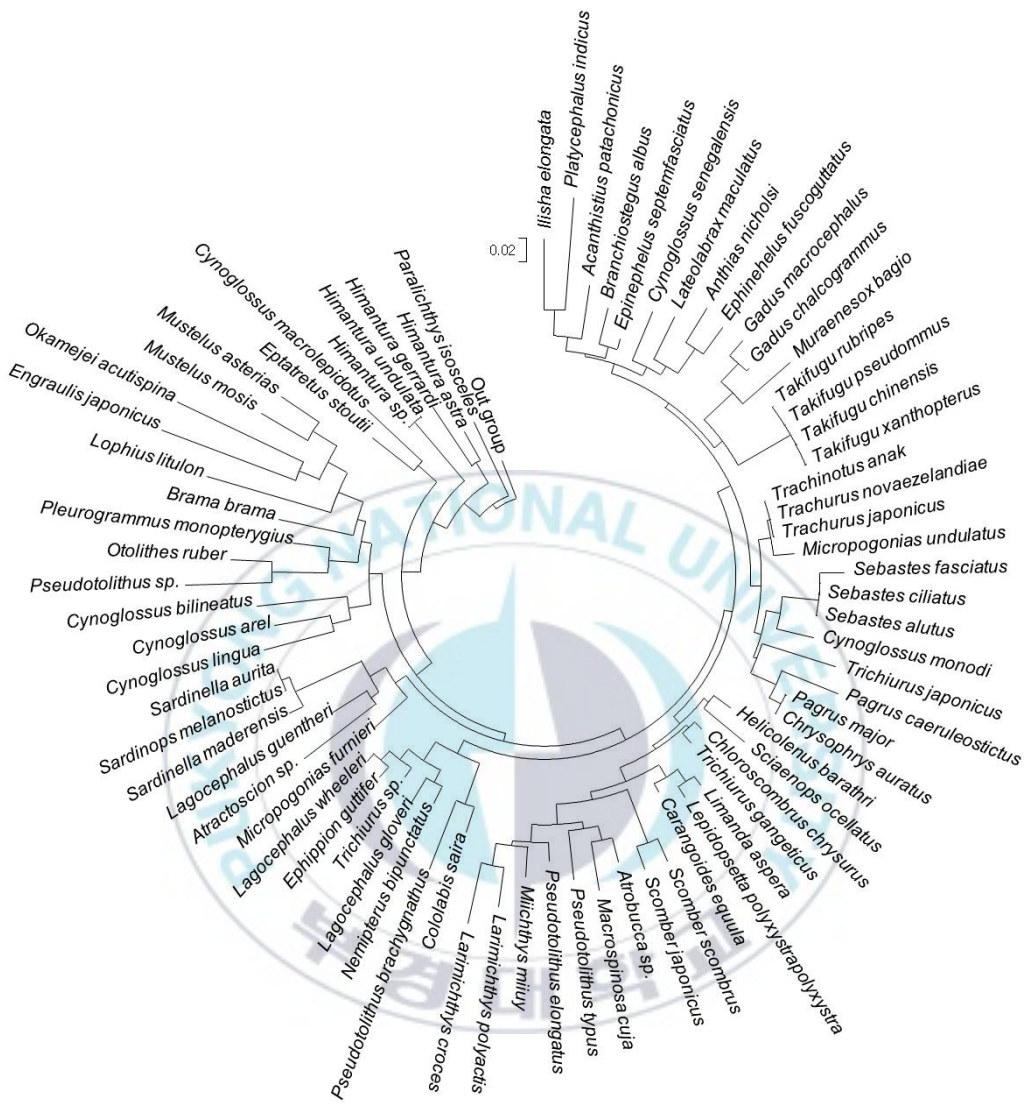

Figure 2. Phylogenetic tree showing relationships among the 76 reference species. The neighbor-joining tree is based on a 213-bp region of the COI gene. Scale bar represents 0.02 substitutions per nucleotide position.

5. References

Andorra, I., Landi, S., Mas, A., Esteve-Zarzoso, B., & Guillamon, J. M. (2010). Effect of fermentation temperature on microbial population evolution using culture-independent and dependent techniques. *Food Res. Int, 43*(3), 773-779.

Arcuri, E. F., Sheikha, A. F. E., Rychlik, T., Piro-Metayer, I., & Montet, D. (2013). Determination of cheese origin by using 16S rDNA fingerprinting of bacteria communities by PCR-DGGE: Preliminary application to traditional Minas cheese. *Food Control, 30*(1), 1-6.

Boon, N., Windt, W., Verstraet, W., & Top, E. (2002). Evaluation of nested PCR-DGGE with group-specific 16S rRNA primers for the analysis of bacterial communities from different wastewater treatment plants. *FEMS Microbiol. Ecol, 39*, 101-112.

Cani, P. D. (2013). Gut microbiota and obesity: lessons from the microbiome. *Brief. Funct. Genomics, 12*, 381-387.

Case, R. J., Boucher, Y., Dahllof, I., Holmstorm, C., Doolittle, W. F., & Kjelleberg, S. (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl.Environ. Microb, 73*(1), 278-188.

Ercolini, D. (2004). PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J. Microbiol. Meth, 56*(3), 297-314.

Ferris, M. J., Muyzer, G., & Ward, D. M. (1996). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. A new molecular approach to analyze the genetic diversity of mixed microbial communities. *Appl. Environ.*

Microbiol, 62(2), 340-346.

Galal-Khallaf, A., Ardura, A., Borrell, Y. J., & Garcia-Vazquez, E. (2016). Towards more sustainable surimi? PCR-cloning approach for DNA barcoding reveals the use of species of low trophic level and aquaculture in Asian surimi. *Food Control, 61*, 62-69.

Gonzalez-Arenzana, L., Lopez, R., Santamaria, P., & Lopez-Alfaro, I. (2013). Dynamics of lactic acid bacteria populations in Rioja wines by PCR-DGGE, comparison with culture-dependent methods. *Appl. Microbiol. Biotechnol, 97*, 6931-6941.

Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., & Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial Community. *Appl. Environ. Microbiol, 66*(12), 5488-5491.

Gryson, N. (2010). Effect of food processing on plant DNA degradation and PCR-based GMO analysis: a review. *Anal. Bioanal. Chem, 396*(6), 2003-2022.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *P. Roy. Soc. Lond. B Bio, 270*, 313-322.

Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *P. Roy. Soc. Lond. B Bio, 270*, S96-S99.

Hong, H., Pruden, A., & Reardon, K. F. (2007). Comparison of CE-SSCP and DGGE for monitoring a complex microbial community remediationg mine drainage. *J. Microbiol. Meth, 69*, 52-64.

Hu, P., Zhou, G., Xu, X., Li, C., & Han, Y. (2009). Characterization of the predominant spoilage bacteria in sliced vacuum-packed cooked ham based on 16S rDNA-DGGE. *Food Control, 20*(2), 99-104.

Huxley-Jones, E., Shaw, J. L. A., Fletcher, C., & Parnell, J. (2012). Use of DNA barcoding to reveal species composition of convenience seafood. *Conserv Biol, 26*(2), 367-371.

Kan, J., Wang, K., & Chen, F. (2006). Temporal variation and detection limit of an estuarine bacterioplonkton community analyzed by denaturing gradient gel electrophoresis (DGGE). *Aquat. Microb. Ecol, 42*, 7-18.

Keskin, E., & Atar, H. H. (2012). Molecular identification of fish species from surimi-based products labeled as Alaskan pollock. *J. Appl. Ichthyol, 28*, 811-814.

Martin-Sanchez, A. M., Navarro, C., Perez-Alvarez, J. A., & Kuri, V. (2009). Alternatives for efficient and sustainable production of surimi: a review. *Compr. Rev. Food Sci. F, 8*(4), 359-374.

Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. BMC Genomics, 9, 214.

Moffitt, C. M., & Cajas-Cano, L. (2014). Blue Growth: The 2014 FAO State of World Fisheries and Aqaculture. *Am. Fish. S, 39*(11), 552-553.

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystem. *Curr. Opin. Microbiol, 2*(3), 317-322.

Muyzer, G., & Smalla, K. (1998). Application of Denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *A. van. leeuw. J. Microb, 73*(1), 127-141.

Muyzer, G., Waal, E. C. de., & Uitterlinden, A. G. (1993). Profiling of complex microbial population by DGGE analysis of polymerase chain reaction amplified genes encoding for 16S rRNA. *Appl. Environ. Microbiol, 59*, 695-700.

Muyzer, G., & Waal, E. C. de. (1994). Determination of the genetic diversity of microbial communities using DGGE analysis of PCR-amplified 16S rDNA. *Microbial Mats, 35*, 207-214.

Neff, M. M., Turk, E., & Kalishman, M. (2002). Web-based primer design for single nucleotide polymorphism analysis. *Trends in Genet, 18*(12), 613-615.

Park, J. W. (2005). Surimi and surimi seafood (2nd ed.). Boca Raton, FL: Taylor & Francis/CRC Press.

Pinto, A. D., Marchetti, P., Mottola, A., Bozzo, G., bonerba, E., Ceci, E., Bottaro, M., & Tantillo, G. (2015). Species identification in fish fillet products using DNA barcoding. *Fish. Res, 170*, 9-13.

Poowakanjana, S., & Park, J. W. (2013). Biochemical characterisation of Alaska pollock, Pacific whiting, and threadfin bream surimi as affected by comminution conditions. *Food Chem, 138*(1), 200-207.

Shiku, Y., Hamaguchi, P. Y., Benjakul, S., Visessanguan, W., & Tanaka, M. (2004). Effect of surimi quality on properties of edible films based on Alaska pollack. *Food Chem, 86*(4), 493-499.

Smith, P. J., Mcveach, S. M., & Steinke, D. (2008). DNA barcoding for the identification of smoked fish products. *J. Fish Biol, 72*(2), 464-471.

Van-Moreira, I., Conceicao, E., Nunes, O. C., & Manaia, C. M. (2013). Bacterial diversity from the source to the tap: a comparative study based on 16S

rRNA gene-DGGE and culture-dependent methods. *FEMS Microbiol. Ecol, 83*, 361-374.

Yin, T., & Park, J. W. (2014). Effects of nano-scaled fish bone on the gelation properties of Alaska pollock surimi. *Food Chem, 150*, 463-468.

Yoon, M., Kim, J.-S., Kim, D., Jo, J., & Cho, S. (2014). Optimization of the processing conditions for the preparation of surimi products containing rice flour. *Fisheries and Aquatic Sciences, 17*(2), 167-173.

Zhao, W., Zhao, Y., Pan, Y., Wang, X., Wang, Z., & Xie, J. (2013). Authentication and traceability of Nibea albiflora from surimi products by species-specific polymerase chain reaction. *Food Control, 31*, 97-101.

Chapter IV.

Quantitative Analysis of Alaska Pollock in Seafood Products by Droplet Digital PCR

Quantitative Analysis of Alaska Pollock in Seafood Products by Droplet Digital PCR

1. Abstract

A highly accurate quantitative method based on the new quantitative PCR technique, droplet digital PCR (ddPCR), was applied to determine the content of Alaska pollock in seafood products. Using the ddPCR method, a linear relationship among raw sample weight, DNA concentration, and DNA copy number was identified. The formula to calculate the raw sample weight was established based on the number of DNA copies. To confirm the accuracy and applicability of this method, mixed samples of known composition were analyzed and verified. The results of this study indicated that ddPCR is highly suitable for quantifying Alaska Pollock in seafood products. Therefore, ddPCR method can be applied to various field of biological research, such as food safety management and forensics.

Keyword: Droplet digital PCR; quantitative analysis; alaska pollock; seafood prodcuts.

2. Introduction

Identification of species and estimation of content are important for evaluating food safety (Rasmussen & Morrissey, 2008). Consumer concern over health and nutrition have increased in recent years, which has led to increased consumption of fish. Fish species with distinct morphological characteristics can be identified by consumers. However, fish products are frequently processed to increase their palatability, which hampers discrimination of fish species (Sotelo et al., 1993). This is exacerbated by mashing and mixing of fish before processing. Processing can result in different raw materials having similar textures, making it difficult for consumers to identify and evaluate the proportions of raw materials in mixed processed foods (Wong & Hanner, 2008).

High-quality raw materials for food are, for reasons of cost, sometimes substituted by inexpensive and inferior materials (Hellberg & Morrissey, 2011). European Council (EC) Regulation No. 104/2000 (17 December 1999) on the common organization of seafood and aquaculture product markets states that aquatic products cannot be sold for retail purposes unless labeled with the species name, production method, and country of origin (Regulation, 2000). Therefore, a quality inspection system is needed to guarantee the safety of mixed processed food, and reliable qualitative and quantitative detection methods should be developed.

Species-specific PCR primers can be used to rapidly and accurately identify target species in food (Trotta et al., 2005). Quantitative real-time PCR (qRT-PCR) uses the relationship between the cycle threshold value and the initial DNA template concentration, and the results are verified in real time (Hayden et al.,

2012). However, qRT-PCR has some important limitations for quantitative analysis. Producing the reference DNA for the standard curve is costly and time-consuming. The accuracy and reproducibility of qRT-PCR are limited by indirect quantitation using the cycle threshold value (Hindson et al., 2013). In addition, PCR inhibitors in extracted DNA can affect amplification efficiency and result in bias (Dingle et al, 2013).

Droplet digital PCR (ddPCR) enables absolute quantification of the number of copies of target DNA in a sample using limiting dilution, PCR, and Poisson analysis (Pinheiro et al., 2011). The droplet generator divides the PCR mixture into ~20,000 droplets, some of which contain one or more copy of the target DNA. A PCR proceeds in each droplet, and the amplicons are confirmed using fluorescent hydrolysis probes specific for the target species (Hindson et al., 2013). The total number of fluorescent positive (1) or negative (0) individual droplets is counted by a droplet reader. The absolute number of target DNA molecules can be calculated using the Poisson distribution. ddPCR has excellent accuracy and reproducibility because it is based on absolute quantification and is not subject to amplification efficiency bias. In addition, the method is more sensitive than qRT-PCR, and accurate results can be obtained from even low-DNA-concentration samples (Doi et al., 2015).

ddPCR has been used with environmental (Doi et al., 2015), food (Morisset et al., 2013), and clinical samples (Strain et al., 2013). Moreover, ddPCR has been applied to estimate fish species diversity and biomass using environmental DNA (eDNA); a correlation between eDNA and biomass was reported. Also, ddPCR was applied to quantitative analysis of raw materials in mixed meats, a highly processed blended food, and showed a high level of accuracy (Floren et al, 2015).

However, no previous study has performed a quantitative analysis of seafood products.

The purpose of this study was to evaluate the ability of ddPCR to accurately quantify raw materials in surimi, which typically contains Alaska pollock (*Gadus chalcogrammus*). The results suggested that the ddPCR method enabled quantification of Alaska pollock in blended seafood products, and that it could be applied for quantification of other fish species in diverse products. This method will enhance seafood safety management.

3. Materials and Methods

3.1 Sample preparation

Fresh fish typically used in whole, commercially available surimi products were obtained from the local market. Also, fifteen surimi products were purchased from local market (Table 1). Fresh Alaska pollock (*Gadus Chalcogrammus*), cutlassfish (*Trichiurus lepturus*), and surimi products were chopped into small pieces and dried in a dry oven at 65℃ for 24 hours. The dried samples were ground in liquid nitrogen using a mortar and pestle. The Alaska pollock and cutlassfish were mixed with known composition (from 80% to 20%) to verify the suitability and sensitivity of the method.

3.2 DNA extraction

Genomic DNA was extracted and purified from 5 mg of powdered sample using a Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) as recommended by the manufacturer. The final volume of DNA solution was 100 µl.

3.3 Primers and probes

The mitochondrial 16S rRNA region was selected as the target detection sequence. Oligomers were designed manually from conserved sequences based on the alignment of 16S rRNA sequences in the NCBI database (http://www.ncbi.nlm.nih.gov). To derive the conserved sequence, the nucleotide frequency at each position was determined using a custom script in CLC

Genomics Workbench 8.0.1(CLC bio, Denmark). Sequences of conserved regions with 100% sequence matches were used for oligomer design. The universal species probe (All-P) was specific for a conserved region and the Alaska Pollock probe (Gad-P) was designed target a sequence specific to this species (Table 2). The universal probe was labeled with 6-carboxyfluorescein (FAM) and black-hole quencher (BHQ1). The Alaska pollock probe was conjugated to hexachlore-6-carboxyfluorescein (HEX) and BHQ1. The sequences targeted by the primers and probes are shown in Figure 1.

3.4 ddPCR procedure

Each 20 µl PCR reaction contained 10 µl of ddPCR Master Mix (Bio-Rad, Hercules, CA, USA), 900 nM of each primer, 125 nM of each probe, 2 µl of genomic DNA, and distilled water to a volume of 20 µl. The 20 µl PCR mixture was divided into 15,000 to 20,000 droplets by a Bio-Rad QX100 droplet generator (Bio-Rad, Hercules, CA, USA). PCR amplification was performed using a Tetrad thermal cycler (Bio-Rad, Hercules, CA, USA). PCR conditions were as follows: pre-denaturation for 10 min at 95℃, followed by 40 cycles of denaturation for 30 sec at 95℃, annealing and extension for 90 sec at 60℃, enzyme inactivation for 10 min at 95℃ and a hold at 4℃. After PCR amplification, TaqMan fluorescence was detected by the droplet reader.

3.5 Standard curve generation

Powdered samples 1 - 7 mg were weighed using an electronic balance (Sartorius CPA225D, Germany). Genomic DNA was extracted and quantified using a NanoVue spectrophotometer (GE Healthcare, Uppsala, Sweden). Genomic DNA was analyzed using ddPCR and the correlation coefficient between the sample weight and DNA concentration was calculated using Microsoft Excel 2016 (Microsoft Corp., Redmond, WA, USA).

Table 1. Information of surimi products used in this study.

Sample	Production country of surimi	Proportion	Fish species labeled in product
A	Imported	–	–
B	Imported	–	–
C	America	82.76%	Alaska pollock
D	America	71.56%	Alaska pollock
E	Imported	70.67%	–
F	Imported	64.62%	White flesh fish
G	Imported	54.89%	Croaker
H	Imported	–	–
I	Vietnam	57.67%	White flesh fish
J	Imported	61.13%	Golden-thread, cutlassfish and croaker
K	America	30.70%	–
L	America	–	Alaska pollock
M	Korea	–	–
N	Imported	–	–
O	Imported	68.97%	–

"–" indicates that the information was not labeled

Table 2. Description of primers and probes

Oligo ID	Sequence	Target
Forward	GTACCTTTTGCATCATGATT	16S rRNA
Reverse	TGGCTGCTTTTARGCCCA	
All-P	FAM-GCAAAAGAGTGGGAAGA-BHQ1	Entire species
Gad-P	HEX-TCACCCATGCTTACGCTAAA-BHQ1	Alaska pollock

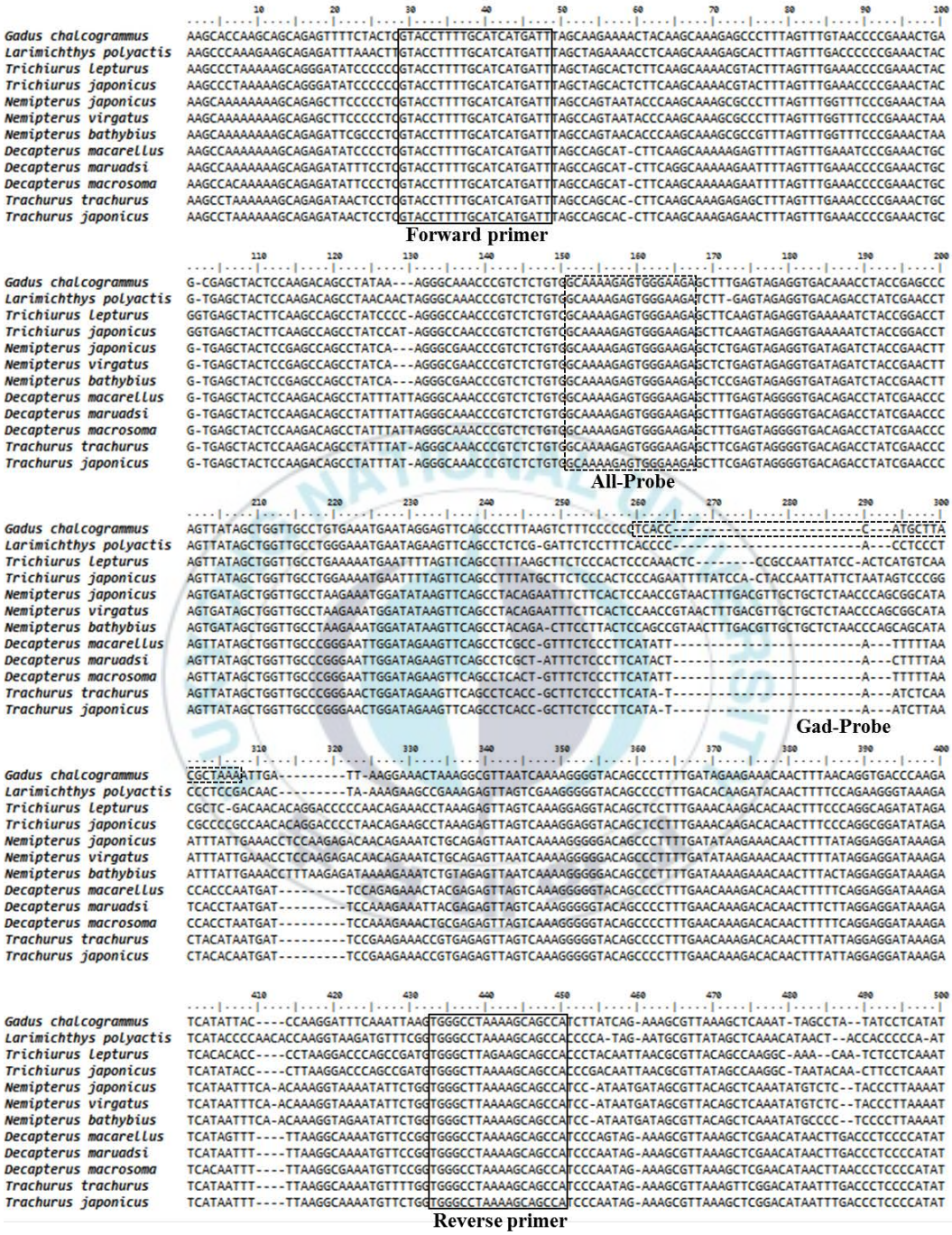

Fig 1. DNA sequences of the twelve fish species. Line boxes indicate the targets of the primer. Line boxes indicate the gargets of each probe.

4. Results

4.1 Specificity of primers and probes

The specificity of the primers and probes was first determined. The genome sequences of the main fish species in surimi – *G. chalcogrammus* (NC004449), *Larimichthys polyactis* (GU586227), *Trichiurus lepturus* (NC018791), *T. japonicus* (EU339148), *Nemipterus japonicus* (NC023972), *N. virgatus* (KU933270), *N. bathybius* (NC029938), *Decapterus macarellus* (NC026718), *D. maruadsi* (NC024556), *D. macrosoma* (NC023458), *Trachurus trachurus* (AB108498), and *T. japonicus* (AP003092) – were obtained from the NCBI database, and compared with those of the primers and probes (Fig 1). The forward primer exhibited perfect matches with all of the above species, and one and two mismatches in the reverse primer with genus *Trichiurus* and *Nemipterus*, respectively. In addition, the primers did not result in non-specific amplification. The universal species probe showed perfect matches with all species. The Alaska pollock-specific probe perfectly matched *G. chalcogrammus*, but not the other species.

4.2 DNA extraction efficiency

DNA concentration was measured to confirm the correlation between the amount of sample and DNA concentration. In three independent experiments, the extracted concentration exhibited a linear relationship with sample weight. The correlation coefficients were 0.997 for Alaska pollock, and 0.998 for cutlassfish (Fig 2).

4.3 Target DNA detection by ddPCR

ddPCR was performed on Alaska pollock samples containing 11.3 - 225.8 ng of DNA, and cutlassfish samples containing 13.3 ng to 72.3 ng of DNA. The concentration for analysis was determined by the detection limit of the ddPCR device and DNA was diluted to 1/100 and used for analysis. Each analysis data was obtained from three replicates per sample. The correlation coefficients between DNA concentration and the Alaska pollock or cutlassfish DNA copy number were 0.998 and 0.997 in the FAM channel, and 0.997 and zero in the HEX channel, respectively (Fig 3).

4.4 Analysis of mixed sample of known weight

To demonstrate its accuracy and applicability, the ddPCR method was applied to four mixed samples of known composition. DNA was extracted three times from 5 mg samples using the same method. The three independent samples showed similar results. The DNA concentration and copy number of each sample are shown in table 4. The concentrations of the mixed samples were not different from the sum of the standard DNA concentration, and the ratio of the total DNA to Alaska pollock DNA copies was similar to the mixing ratio. The estimated Alaska pollock contents of the four samples were 4.03, 3.09, 1.58, and 0.98 mg, respectively.

4.5 Analysis of commercial surimi products.

The Alaska pollock contents of 15 commercial surimi products were analyzed by ddPCR. The proportion of Alaska pollock was determined by comparing the

number of DNA copies with reference data. The Alaska pollock content of the nine samples was 0.79 - 2.87 mg/5 mg (Table 5).

Table 3. The results of quantification of standard samples.

Sample	Sample weight (mg)	DNA concentration (ng)	Copy number	
			Alaska pollock	Entire species
Alaska pollock (*G. chalcogrammus*)	1	11.3	4.1	4.3
	2	30.2	11.3	11.7
	3	78.2	22.7	22.3
	4	118.8	31.5	30.5
	5	156.7	42.2	42.1
	6	175.0	45.1	45.9
	7	225.8	55.2	56.6
Cutlassfish (*T. lepturus*)	1	13.3	0.1	1.9
	2	19.5	0.0	4.1
	3	26.8	0.0	8.3
	4	42.0	0.1	12.8
	5	48.7	0.1	17.9
	6	62.2	0.0	23.1
	7	72.3	0.0	27.8

Table 4. The results of quantification of the mixed samples with known proportion.

Sample	Mixture proportion		DNA concentration	Copy number		Measure weight	Deviation
	Alaska pollock	Cutlassfish	(ng)	Alaska pollock	Entire species	Alaska pollock	
1	4 mg	1 mg	143.0	31.0	34.9	4.08 mg	2.0%
2	3 mg	2 mg	102.8	23.5	27.7	3.21 mg	7.0%
3	2 mg	3 mg	57.2	9.6	21.6	1.58 mg	-21.0%
4	1 mg	4 mg	52.7	4.4	16.3	0.98 mg	2.0%

Table 5. Quantitative analysis of Alaska pollock from commercial surimi products. 5 mg of powdered sample was used for quantitative analysis.

Sample	DNA concentration (ng)	Copy number		Measure weight
		Alaska pollock	Entire species	Alaska pollock
A	3.5	0.0	0.0	–
B	9.5	0.7	0.0	–
C	78.0	20.6	22.0	2.87 mg
D	63.2	15.9	14.9	2.32 mg
E	35.5	8.5	12.7	1.46 mg
F	4.9	0.7	2.9	–
G	11.2	0.9	2.7	–
H	12.2	0.5	2.6	–
I	21.7	0.0	5.4	–
J	30.0	6.7	11.6	1.25 mg
K	16.1	9.1	10.1	1.53 mg
L	17.3	5.4	12.5	1.09 mg
M	16.6	7.8	10.5	1.37 mg
N	24.7	10.4	12.0	1.68 mg
O	25.8	2.8	11.6	0.79 mg

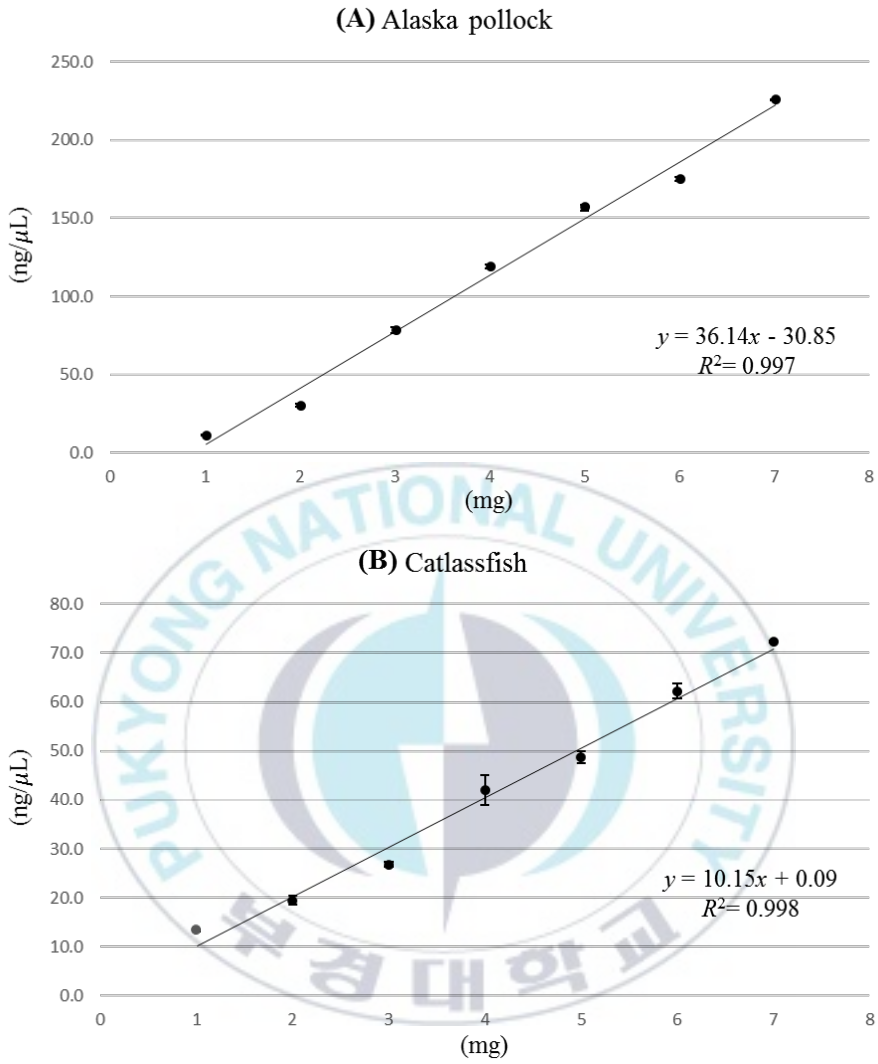

Fig 2. Linear relationship between weight (mg) and DNA concentration (ng). The correlation coefficient (R^2) for the sample weight and DNA concentration were (A) 0.997 for Alaska pollock and (B) 0.998 for catlassfish.

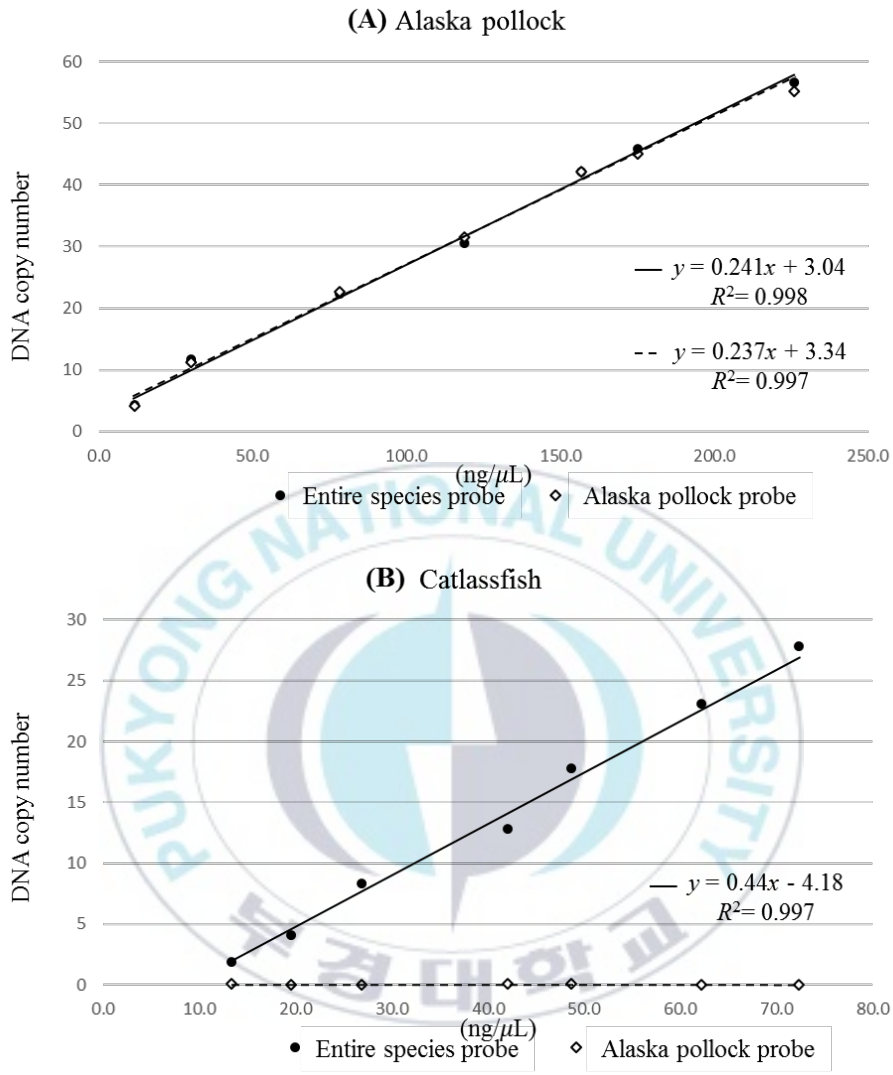

Fig 3. Linear relationship between DNA concentration (ng) and the target DNA copy number. (A) The correlation coefficient (R^2) for the DNA concentration and the DNA copy number of Alaska pollock were 0.998 and 0.997 in each probe, respectively. (B) The correlation coefficient of catlassfish were 0.997 and zero in each probe, respectively.

5. Discussion

Our results suggest that ddPCR can be used to accurately quantify the Alaska pollock content of surimi products. Fresh samples were used to ensure a consistent of DNA content, and the mitochondrial 16S rRNA region, which exhibits low intraspecies genetic variation, was selected as the target to proven false negative results due to genetic variation (Xia et al., 2011).

A primer set and two probes were designed; the latter targeted a sequence with no intraspecies divergence. The specificity of the primers and probes was tested by fluorescence melting curve analysis (FMCA) and real-time PCR method (Huang et al., 2011). Using FMCA, fluorescence was emitted when the probe and target template DNA exhibit 100% sequence identity.

Samples were dried and powdered, and DNA extraction was performed three times independently for ensure accuracy. To ensure the accuracy of the quantification method, each ddPCR reaction generated and analyzed more than ~15,000 effective drops per experiment. Both Alaska pollock and cutlassfish showed a linear relationship between raw sample weight (mg) and DNA concentration (ng). Alaska pollock and cutlassfish samples exhibited an almost-linear correlation between DNA concentration and target DNA copy number.

DNA extraction from mixed seafood products can be affected by various factors; e.g., tissue composition and DNA degradation (Cai et al., 2014). Therefore, the concentration of DNA may not reflect the weight percentage of raw materials. To confirm its accuracy, four mixed samples of known composition were analyzed using the ddPCR methods; the results were similar to those of standard samples. Although the amount of extracted DNA varied

depending on the fish species, the relationship between sample weight and DNA concentration did not differ between single or mixed samples. The deviations were very low, suggesting that ddPCR analysis is highly accurate. In addition, the Alaska pollock content was measured in 15 commercially available surimi products; the results demonstrated the feasibility of ddPCR.

This study is the first to quantify raw materials in mixed seafood products by ddPCR. This method has the potential to be used routinely to quantify various fish species. Therefore, it can be applied to various fields of biological research, such as food safety management and forensics.

6. References

Cai, Y., Li, X., Lv, R., Yang, J., Li, J., He, Y., & Pan, L. (2014). Quanti tative analysis of pork and chicken products by droplet digital PCR. *Biomed. Res. Int*, 810209.

Dingle, T. C., Sedlak, R. H., Cook, L., & Jerome, K. R. (2013). Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory s ubstance. *Clin. Chem, 59*(11), 1670-1672.

Doi, H., Takahara, T., Minamoto, T., Matsuhashi, S., Uchii, K., & Yamana ka, H. (2015). Droplet digital polymerase chain reaction (PCR) outp erforms real-time PCR in the detection of environmental DNA from an invasive fish species. *Environ. Sci. Technol. Lett, 49*(9), 5601-56 08.

Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H., & Minam oto, T. (2015). Use of droplet digital PCR for estimation of fish ab undance and biomass in environmental DNA surveys. *PLoS One, 10* (3), e0122763.

Floren, C., Wiedemann, I., Brenig, B., Schütz, E., & Beck, J. (2015). Spec ies identification and quantification in meat and meat products using droplet digital PCR (ddPCR). *food chem, 173*(15), 1054-1058.

Hayden, R. T., Gu, Z., Ingersoll, J., Abdul-Ali, D., Shi, L., Pounds, S., & Caliendo, A. M. (2012). Comparison of droplet digital PCR to real -time PCR for quantitative detection of cytomegalovirus. *J. Clin. Mi crobiol, 51*(2), 540-546.

Hellberg, R. S. R., & Morrissey, M. T. (2011). Advances in DNA-Based

Techniques for the Detection of Seafood Species Substitution on the Commercial Market *J. Lab. Autom, 16*(4), 308-321.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., Vessella, R. L., & Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods, 10*(10), 1003-1005.

Huang, Q., Liu, Z., Liao, Y., Chen, X., Zhang, Y., & Li, Q. (2011). Multi plex fluorescence melting curve analysis for mutation detection with dual-labeled, self-quenched probes. *PLoS One, 6*(4), e19206.

Morisset, D., Stebih, D., Milavec, M., Gruden, K., & Zel, J. (2013). Quant itative analysis of food and feed samples with droplet digital PCR. *PLoS One, 8*(5), e62583.

Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., & Emslie, K. R. (2011). Evaluation of a droplet dig ital polymerase chain reaction format for DNA copy number quantifi cation. *Anal. Chem, 84*(2), 1003-1011.

Rasmussen, R. S., & Morrissey, M. T. (2008). DNA-Based Methods for th e Identification of Commercial Fish and Seafood Species. *Compr. R ev. Food Sci. F, 7*(3), 280-295.

Regulation, C. (2000). No 104/2000 of 17 December 1999 on the common organization of the markets in fishery and aquaculture products. *Off. J. Eur. Commun. L, 17*, 22-52.

Sotelo, C. G., Piñeiro, C., Gallardo, J. M., & érez-Martin, R. I. (1993). Fi sh species identification in seafood products. *Trends Food Sci. Tech nol, 4*(12), 395-401.

Strain, M. C., Lada, S. M., Luong, T., Rought, S. E., Gianella, S., Terry, V. H., Spina, C. A., Woelk, C. H., & Richman, D. D. (2013). Hig hly precise measurement of HIV DNA by droplet digital PCR. *PLo S One*, *8*(4), e55943.

Trotta, M., Schonhuth, S., Pepe, T., Cortesi, M. L., Puyet, A., & Bautista, J. M. (2005). Multiplex PCR method for use in real-time PCR for identification of fish fillets from grouper (*Epinephelus* and *Myctero perca* species) and common substitute species. *J. Agr. Food Chem.*, *53*(6), 2039-2045.

Wong, E. H. K., & Hanner, R. H. (2008). DNA barcoding detects market substitution in North American seafood. *Food. Res. Int*, *41*(8), 828-8 37.

Xia, Y., Gu, H. F., Peng, R., Chen, Q., Zheng, Y. C., Murphy, R. W., & Zeng, X. M. (2011). COI is better than 16S rRNA for DNA barco ding Asiatic salamanders (Amphibia: Caudata: Hynobiidae). *Mol. Eco l. Resour*, *12*(1), 48-56.

Korean Summary (국문요약)

현재 수산물 먹거리의 안전관리를 위한 가장 중요한 쟁점 중 하나는 혼합 가공식품에서의 식품원료 및 원재료의 식별이라 할 수 있다. 전세계적으로 수산물 가공식품 시장이 점차 확대되면서 경제적 이득을 취하기 위해 고가의 원료종을 의도적으로 저가의 원료종으로 혼합 또는 대체하거나 표시사항을 허위로 기재하여 판매하는 가짜식품(Economically Motivated Adulteration, EMA)의 제조와 판매가 급증하고 있다. 수산물 가공식품의 경우 다양한 형태의 원료 가공으로 인해 형태학적으로 원료종을 식별할 수 없기 때문에 원재료의 판별은 매우 어려운 실정이다. 따라서, 본 연구에서는 형태학적인 종식별의 한계성과 문제점을 극복할 수 있는 분자생물학적 기법을 적용하여 다양한 형태의 수산가공식품에서 원료종을 식별할 수 있는 분석방법을 개발하였다.

분자생물학적 기술은 형태학 기반의 종 식별 방법의 한계를 극복하기 위해 최근 널리 사용되고 있다. DNA 기반의 분석법은 높은 특이성과 민감도를 가지며 일관된 결과를 제공할 수 있는 방법이다. 특히, DNA 메타바코딩 기술은 복잡한 환경에서 생물종의 다양성을 빠르게 분석할 수 있는 유용한 방법이다. 생물종을 식별하기 위하여 사용되는 유전자 영역을 바코드라 부르며, 일반적으로 동물종의 식별에는 미토콘드리아 유전자가 사용된다. 이러한 바코드 유전자에 대한 범용 프라이머는 모든 생물 종을 커버할 수 없기 때문에 메타바코딩 연구를 위해서는 특정 분류군을 위한 새로운 범용 프라이머의 개발이 매우 중요하다.

차세대 염기서열 분석 (Next Generation Sequencing, NGS) 기술을 이용한 메타바코딩 연구는 생물종의 다양성을 분석하기 위한 최적의 방법으로 여겨지고 있다. 본 연구에서는 수산물 가공식품으로부터 주원료의 식별을 위해 NGS 기술을 적용하였다. 미토콘드리아 DNA의 cytochrome c oxidase subunit I (COI) 유전자로부터 약 700 bp의 바코드 영역을 증폭시키기 위한 두 세트의 범용 프라이머를 설계 하였다. 개발된

프라이머를 이용하여 총 11개의 수산물 가공식품을 대상으로 분석을 수행하였으며, 프라이머 세트에 따라 각각 56종과 32종의 어류를 확인할 수 있었다. 분석된 어종의 다양성을 통해서 제작된 프라이머의 범용성에는 확연한 차이가 있음을 확인할 수 있었으나, 수산가공식품의 주원료로 가장 많이 사용되는 것으로 알려진 명태와 실꼬리돔은 두 세트의 프라이머 모두에서 확인이 가능하였다. 분석된 어종의 비교를 통해 새로 제작된 범용 프라이머가 수산가공식품의 주원료 식별을 위한 활용 가능성을 확인할 수 있었다.

NGS 기술을 이용한 메타바코딩 연구는 정확한 분석 결과를 얻을 수 있는 이점이 있지만, 분석의 실용화를 위해서는 비용적인 측면에서 어려움이 남아있다. 이러한 이유로 수산물 가공식품에서의 메타바코딩 기술의 실용화를 위해 변성 구배 젤 전기영동 (Denaturing gradient gel electrophoresis, DGGE) 기술을 적용하였다. DGGE는 DNA 염기서열의 변성 특성에 따른 전기영동의 이동속도 차이를 이용하여 젤 상에서 다양한 DNA를 쉽게 분리할 수 있는 방법이다. 분석을 위한 범용프라이머 세트는 COI 유전자를 기반으로 약 200 bp를 증폭하도록 설계 하였으며, 총 20개의 수산가공식품을 대상으로 DGGE 분석을 수행하였다. 각각의 DNA 밴드 위치에 따라 총 30개의 밴드를 선택 및 유전자 서열 분석을 수행하였으며, 총 20종의 어류를 확인할 수 있었다. NGS 기술에 비하여 비교적 적은 수의 어종이 확인 되었으나 주요 어종에 대해서는 높은 식별능을 가지고 있음을 확인할 수 있었다.

최종적으로 사전 연구를 바탕으로 수산물 가공식품에서 가장 많이 이용되는 명태를 대상으로 원료의 함량 분석 연구를 수행하였다. 본 분석에는 3세대 중합효소 연쇄 반응 (Polymerase chain reaction, PCR)으로 일컬어지는 droplet digital PCR (ddPCR) 기술을 사용하였으며, 이는 주어진 표본에서 표적 DNA의 수를 절대적으로 정량화 할 수 있는 방법이다. 분석을 위한 범용프라이머 세트와 특이 프로브는 미토콘드리아 DNA의 16S rRNA 유전자를 기반으로 설계하였다. 16S rRNA 유전자는 다른 유전자 영역에 비해 비교적 종내 유전적 변이율이 낮기 때문에 프로브 반응에서의 위음성의

결과를 피할 수 있는 유전자영역이다. 분석을 통해 명태의 무게와 DNA의 농도, 그리고 DNA의 농도와 DNA의 복제수 사이의 선형 관계를 확인할 수 있었다. 이 결과를 통해 DNA의 복제수를 기준으로 명태의 무게를 추정할 수 있는 수식을 고안하였으며, 총 15개의 수산물 가공식품으로부터 명태의 함량을 성공적으로 추정할 수 있었다.

본 연구의 최종 목적은 주원료를 확인할 수 없는 수산물 가공식품의 원료종을 식별하기 위한 기술을 개발하는 것이었다. 수산가공식품을 대상으로한 DNA 메타바코딩 분석의 정확성과 실용성을 확인하기 위해 NGS, DGGE, 그리고 ddPCR을 사용하였으며, 원료종의 정성·정량 분석을 가능케 하였다. 본 연구에서 확립된 메타바코딩 기술들은 향후 식품 안전 관리를 강화하고, 소비자에게 안전한 식품을 제공하기 위해 유용하게 활용될 수 있을 것이라 기대한다.

Acknowledgements

이 논문이 완성되기까지 많은 도움과 격려를 주신 모든 분들께 감사의 마음을 전합니다. 긴 학위 기간 동안 부족한 제가 스스로 깨우치고 배워나갈 수 있도록 이끌어 주시고, 따뜻하게 격려해 주신 강정하 박사님과 김경호 교수님께 깊은 감사와 존경의 마음을 드립니다. 또한, 논문의 심사를 위해 조언과 충고를 아끼지 않고 많은 관심을 가져주신 최태진, 김군도, 김영목 교수님께도 감사드립니다.

여러 해 동안 함께 일하며 많은 힘이 되어주신 국립수산과학원 안철민, 박중연 박사님께 깊은 감사의 마음을 전합니다. 그리고 항상 가까이에서 많은 조언을 아끼지 않으신 노재구, 김영옥, 남보혜, 공희정, 김동균, 김은미, 김주원 박사님 그리고 강성경 선생님과 동료 연구원분들께도 감사의 마음을 전합니다. 또한, 저에게 관심과 격려를 아끼지 않고 많은 도움을 주신 모든 분들께 진심으로 감사의 마음을 전합니다.

마지막으로, 지금의 제가 있기까지 항상 옆에서 변함없는 사랑으로 지켜봐 주시고 힘이 되어 주신 아버지 노영섭님, 어머니 허유님 그리고 누나 노제인님께 깊은 감사와 사랑의 마음을 표합니다. 감사합니다.