**Chapman University**
## Chapman University Digital Commons

Food Science Faculty Articles and Research

Food Science

3-2-2016

# Use of the Mitochondrial Control Region as a Potential DNA Mini-Barcoding Target for the Identification of Canned Tuna Species

Jacquelyn K. Mitchell
*Chapman University*

Rosalee S. Hellberg
*Chapman University*, hellberg@chapman.edu

Follow this and additional works at: http://digitalcommons.chapman.edu/food_science_articles

🗛 Part of the Aquaculture and Fisheries Commons, Food Biotechnology Commons, Food Microbiology Commons, Meat Science Commons, Other Animal Sciences Commons, and the Other Food Science Commons

### Recommended Citation

# Use of the Mitochondrial Control Region as a Potential DNA Mini-Barcoding Target for the Identification of Canned Tuna Species

1    **Use of the Mitochondrial Control Region as a Potential DNA Mini-Barcoding Target for the**

2    **Identification of Canned Tuna Species**

3

4    Jacquelyn K. Mitchell, Rosalee S. Hellberg*

5

6    Chapman University, Schmid College of Science and Technology, Food Science and Nutrition, One

7    University Drive, Orange, CA 92866

8

9    ***Corresponding Author:**

10   Rosalee S. Hellberg, Ph.D

11   Chapman University

12   Ph: 714-628-2811

13   Fax: 714-289-2041

14   Email: hellberg@chapman.edu

15

16   **Abstract**

17   In this study, a DNA mini-barcoding methodology was developed for the differentiation of species

18   commonly found in canned tuna. Primers were designed to target a 236-base pair (bp) fragment of the

19   mitochondrial control region (CR) and a 179-bp fragment of the first internal transcribed spacer region

20   (ITS1). Phylogenetic analysis revealed the ability to differentiate 13 tuna species on the basis of the CR mini-

21   barcode, except in a few cases of species introgression. Supplementary use of ITS1 allowed for

22   differentiation of introgressed Atlantic bluefin tuna (*Thunnus thynnus*) and albacore tuna (*Thunnus alalunga*),

23   while differentiation of introgressed Atlantic bluefin tuna and Pacific bluefin tuna (*Thunnus orientalis*)

24   requires a longer stretch of the CR. After primer design, a market sample of 53 commercially canned tuna

25   products was collected for testing. This mini-barcoding system was able to successfully identify species in 23

26   of the products, including albacore tuna, yellowfin tuna (*Thunnus albacares*), and skipjack tuna (*Katsuwonus*

27   *pelamis*). One instance of mislabeling was detected, in which striped bonito (*Sarda orientalis*) was identified

28   in a product labeled as tongol tuna (*Thunnus tonggol*). PCR amplification and sequencing was unsuccessful

29   in a number of products, likely due to factors such as the presence of PCR inhibitors and DNA fragmentation

30   during the canning process. Overall, CR and ITS1 show high potential for use in identification of canned tuna

31   products; however, further optimization of the assay may be necessary in order to improve amplification and

32   sequencing success rates.

33

34   **Keywords:** DNA mini-barcoding; canned tuna; species identification; mitochondrial control region; first

35   internal transcribed spacer region

36

37

38

39

40

41

42

43

44    **Introduction**

45       Fish species substitution is a type of misbranding that involves one fish species being substituted for

46    another and sold as a mislabeled product. Mislabeling of fish species has been known to occur on the

47    commercial market, with one U.S. market survey reporting that 33% of fish tested nationwide were

48    mislabeled (Warner et al. 2013) and another study finding that 25% of fish samples collected in North

49    America were potentially mislabeled (Wong and Hanner 2008). Mislabeling of fish products is carried out

50    for reasons such as economic gain or avoidance of trade restrictions (Rasmussen and Morrissey 2008). Fish

51    that are substituted or mislabeled are considered in violation of the Federal Food Drug and Cosmetic Act

52    Section 403: Misbranded Food (21 U.S.C. 343). The potential for seafood mislabeling on the commercial

53    market has become an increasing concern due to factors such as changes in the supply of particular fish

54    species, increasing international trade, and increased production of processed seafood (Rasmussen Hellberg

55    and Morrissey 2011). Whole, unprocessed fish can typically be identified by morphological characteristics.

56    However, species identification becomes more challenging after commercial processing, when distinguishing

57    external features of the fish have been removed.

58       Canned tuna is among the top-three consumed seafoods in the United States (NFI 2014), and has high

59    potential to be the target of intentional or unintentional mislabeling. There are fourteen species listed in the

60    Code of Federal Regulations (CFR) that can legally be sold in the United States as canned tuna fish (21 CFR

61    161.190). Canned tuna labeled as "white" can only contain albacore (*Thunnus alalunga*) with a Munsell

62    value of 6.3 or higher while "light" tuna can contain any species listed in 21 CFR 161.190, as long as the

63    tuna has a Munsell value ≥ 5.3. Different tuna species have varying quality, value, availability, and

64    restrictions, leading to the potential for fraudulent species substitution (Chuang et al. 2012; Jacquet and Pauly

65    2008). For example, the average 2014 ex-vessel price paid in the U.S. for commercial landings of skipjack

66    tuna (*Katsuwonus pelamis*) was US$1.50/kg while the average price paid for bigeye tuna (*Thunnus obesus*)

67    was US$6.78/kg (NMFS 2015). Besides economic deception, tuna species mislabeling can also present a risk

68    to human health. For example, sushi products labeled as "white tuna" have been reported to instead contain

69    escolar (*Lepidocybium flavorunneum*) (Lowenstein et al. 2009; Warner et al. 2013). Escolar contains high

70    levels of wax esters and is banned for sale in Japan and Italy because it can cause gastrointestinal distress

71    (EFSA 2004). Proper labeling of canned tuna is also needed to allow at-risk consumers to properly follow the

3

72  U.S. Food and Drug Administration (FDA)/U.S. Environmental Protection Agency (EPA) guidelines for fish

73  consumption (FDA/EPA 2014). In order to limit mercury exposure, children and pregnant women are

74  advised to eat no more than six ounces per week of canned albacore, but are encouraged to eat eight to twelve

75  ounces of fish lower in mercury, such as light canned tuna.

76      Due to the processed nature of canned tuna, DNA-based testing is typically required for species

77  identification (Espiñeira et al. 2009; Quinteiro et al. 1998). DNA barcoding is one of the major DNA-based

78  tests used to identify fish species, and it has been adopted by the FDA for testing of regulatory samples

79  (Handy et al. 2011a; Handy et al. 2011b). This method is a sequencing-based test that differentiates between

80  animal species based on a standardized gene fragment (Hebert et al. 2003). In fish, the standard fragment

81  used for DNA barcoding is 655 base pairs (bp) in length and is located near the 5' end of the cytochrome *c*

82  oxidase subunit 1 (COI) mitochondrial gene (Handy et al. 2011a). DNA barcoding has proven to be a highly

83  successful method for differentiating most fish species, but the COI gene fragment cannot always

84  successfully differentiate between closely-related tuna species due to low genetic divergences (Cawthorn et

85  al. 2011; Ward et al. 2009). Furthermore, the high-heat treatment that occurs during the canning process

86  degrades DNA to fragments roughly 100 to 360 bp in length (Rasmussen and Morrissey 2009), often

87  preventing species identification with the full-length COI barcode.

88      DNA mini-barcoding, which targets shorter regions within the full-length barcode, can be employed to

89  differentiate fish species even after heavy processing (Rasmussen Hellberg and Morrissey 2011; Shokralla et

90  al. 2015). A COI mini-barcoding system has been developed for the identification of fish species; however,

91  canned tuna products could not consistently be differentiated at the species level with this method and the use

92  of alternative genetic markers was suggested (Shokralla et al. 2015). Previous studies have reported some

93  success in differentiating canned tuna using short fragments of the mitochondrial gene coding for cytochrome

94  *b* (Espiñeira et al. 2009; Unseld et al. 1995); however, these studies did not consider the possibility of

95  introgression, which has been reported to occur in a small percentage of cases and results in identical or

96  extremely similar mitochondrial DNA sequences across multiple species (Viñas and Tudela 2009). The

97  mitochondrial DNA control region (CR), which is a non-coding stretch of DNA that shows high levels of

98  genetic variation, is a promising option for differentiating tuna species using DNA mini-barcoding. Previous

99  studies have reported the ability to reliably identify tuna species based on sequence variation in a fragment of

100    the CR approximately 450 bp in length (Cawthorn et al. 2011; Viñas and Tudela 2009). This region has also

101    been studied with introgressed tuna sequences and a secondary nuclear fragment targeting the  first internal

102    transcribed spacer region (ITS1;~600-650 bp) has been identified for supplemental species differentiation

103    (Chow et al. 2006; Viñas and Tudela 2009). Despite the success of these genetic markers in differentiating

104    tuna species, the fragments targeted by previous studies are too long to be reliably recovered from canned

105    tuna products. Therefore, the objective of this study was to develop a DNA mini-barcoding system for tuna

106    species identification based on the CR and ITS1 and to test this system against a variety of canned tuna

107    products.

108    **Materials and Methods**

109    *Primer design and optimization*

110        Primers were designed to target a short (< 250 bp) fragment of the CR in tuna fish species listed in the

111    CFR for canned tuna (21 CFR 161.190). A total of 1,580 CR sequences were downloaded from GenBank for

112    the following species: Atlantic bluefin tuna (*Thunnus thynnus*), albacore tuna (*Thunnus alalunga*), bigeye

113    tuna (*Thunnus obesus*), yellowfin tuna (*Thunnus albacares*), Southern bluefin tuna (*Thunnus maccoyii*),

114    longtail tuna (*Thunnus tonggol*), blackfin tuna (*Thunnus atlanticus*), skipjack tuna (*Katsuwonus pelamis*),

115    slender tuna (*Allothunnus fallai*), bullet tuna (*Auxis rochei*), frigate tuna (*Auxis thazard*), kawakawa

116    (*Euthynnus affinis*), and Pacific bluefin tuna (*Thunnus orientalis*). Although *T. orientalis* is not listed in the

117    CFR for canned tuna, it is now considered to be a separate species from *T. thynnus* (Collete 1999; ITIS 2015).

118    The downloaded CR sequences included introgressed individuals of *T. thynnus* and *T. orientalis*, as identified

119    in previous studies (Alvarado Bremer et al. 2005; Carlsson et al. 2007; Carlsson et al. 2004; Viñas and

120    Tudela 2009). CR sequences were not available in GenBank for two of the species listed in the CFR for

121    canned tuna: spotted tunny (*Euthynnus alletteratus*) and black skipjack tuna (*Euthynnus lineatus*). All

122    sequences were aligned with ClustalW using the default settings in MEGA 5.2 (Tamura et al. 2011). The

123    sequences were then manually examined in BioEdit Sequence Alignment Editor, v.7.1.3.0 (Hall 1999) and

124    potential primer-binding sites were identified by searching for conserved regions flanking highly variable

125    regions. Parameters such as primer-dimer potential, %GC, and annealing temperatures were assessed using

126    the Thermo Scientific Multiple Primer Analyzer online tool

127    (http://www.thermoscientificbio.com/webtools/multipleprimer/). Based on the results of *in silico* analyses, a

128    cocktail of three primers was designed for amplification of a 236-bp region of the CR (Table 1). Two reverse

129    primers were designed to account for differences in the primer-binding region among the target species.

130    Phylogenetic analysis of this genetic region was carried out in MEGA 5.2 (Tamura et al. 2011) using a subset

131    of sequences representing each target species. Genetic divergence was calculated using the Kimura 2-

132    parameter distance method (Kimura 1980) with pairwise deletion for missing data, and a neighbor-joining

133    (NJ) tree was compiled (Saitou and Nei 1987). The robustness of the tree was evaluated using nonparametric

134    bootstrap analysis with 1,000 iterations. In order to allow for differentiation of albacore-like *T. thynnus* and *T.*

135    *alalunga*, an additional primer set was designed to target a short (<250 bp) region of ITS1 in both species

136    (Table 1). Primers were designed based on ITS1 sequences for *T. alalunga*, *T. thynnus*, and albacore-like *T.*

137    *thynnus* published previously (Chow et al. 2006; Viñas and Tudela 2009). Primer design and optimization, as

138    well as phylogenetic analysis, were carried out using the same methodology described above for the CR.

139         Following phylogenetic analyses, all newly designed primers were synthesized with M13 tails to facilitate

140    DNA sequencing (Handy et al. 2011a). Primers were optimized using gradient polymerase chain reaction

141    (PCR) with annealing temperatures of 45º-65ºC (temperature increasing in 2ºC increments) using a

142    Mastercycler nexus gradient thermal cycler (Eppendorf, Hamburg, Germany). Additional PCR parameters

143    are described in the "PCR and sequencing" section. Optimal reaction conditions were determined based on

144    amplification success with canned *K. pelamis* (CR primers only), canned *T. alalunga*, and fresh/frozen *T.*

145    *alalunga*. Following optimization, primers were tested against commercially canned tuna samples as

146    described below.

147    *Sample collection*

148         A total of 53 canned tuna fish products representing a variety of commonly sold species were collected

149    for use in testing the DNA mini-barcoding system developed here (Table 2). Products were purchased from 6

150    online retail sources and 8 retail outlets in Orange County, CA. Fish tissue (~10 mg) was collected from each

151    canned tuna sample using sterile forceps and placed into a sterile 1.5 ml microcentrifuge tube for DNA

152    extraction. The remaining portion of each sample was stored at -80ºC.

153    *DNA extraction*

154         DNA extraction was carried out for all fish samples using the DNeasy Blood and Tissue Kit, Spin-

155    Column protocol (Qiagen, Valencia, CA) with modifications described previously (Handy et al. 2011a;

156     Handy et al. 2011b). Buffer ATL (50 µl) and Proteinase K (5.56 µl) were added to each sample tube

157     described above and tissues were lysed at 56ºC for 1-3 h, with vortexing approximately every 30 min. After

158     lysis, Buffer AL (55.6 µl) and 95% ethanol (55.6 µl) were added and each sample was vortexed. Samples

159     were then transferred to silica spin columns, centrifuged at 8,000 rpm for 1 min, and transferred to fresh

160     collection tubes. Wash buffer AW1 (140 µl) was then aliquoted into each spin column and samples were

161     again centrifuged at 8,000 rpm for 1 min before being placed in a fresh collection tube. Wash buffer AW2

162     (140 µl) was then added to each spin column followed by a centrifugation step of 14,000 rpm for 3 min.

163     Finally, the silica columns were placed in 1.5 ml microcentrifuge tubes and 50 µl of pre-heated Buffer AE

164     (37ºC) were added to each column. Samples were incubated at room temperature for 1 min, followed by

165     centrifugation at 8,000 rpm for 1 min to allow elution of the DNA. The extracted DNA was stored at -80ºC

166     until PCR and sequencing. Reagent blanks with no fish tissue added were included as negative controls for

167     each set of samples extracted.

168         Samples that failed sequencing underwent a repeat DNA extraction that incorporated the use of the MP

169     FastPrep-24 Tissue and Cell Homogenizer (MP Biochemicals, Solon, Ohio). One sample of fish tissue (20-25

170     mg) was collected from each sample and placed into an MP Lysing Matrix A tube (MP Biochemicals).

171     Buffer ATL (180 µl) from the DNeasy Blood and Tissue Kit was added to each Lysing Matrix A tube and the

172     tubes were homogenized in the FastPrep-24 instrument at 6.0 m/s for 40 s. Samples were then spun down

173     briefly and Proteinase K (20 µl) was added to each sample. The tissues were lysed for 2 h with vortexing

174     approximately every 30 min. After lysis, DNA extraction was carried out using the DNeasy Blood and Tissue

175     Kit according to the manufacturer's instructions. DNA was eluted in 50 µl of preheated (37ºC) Buffer AE.

176     Reagent blanks with no fish tissue added were included as negative controls for each set of samples extracted.

177     *PCR and sequencing*

178         The DNA samples extracted from each product underwent PCR along with reagent blanks and non-

179     template controls. Each reaction included the following components: 0.5 OmniMix HS PCR bead (Cepheid,

180     Sunnyvale, CA), 0.5 µl of each 10 µM primer or primer cocktail (Table 1), 2-3 µl DNA template, and

181     molecular grade water for a total volume of 25 µl. The CR mini-barcode region was amplified using one 10

182     µM forward primer (CR_F) and two reverse primers mixed together (CR_R1 and CR_R2, each at a

183     concentration of 10 µM in the mixture), while the ITSI target was amplified with one 10 µM forward primer

184 (ITS1_F) and one 10 μM reverse primer (ITS1_R). A subset of samples was initially tested with both 2 and 3

185 μl of DNA. Increased sequencing success was achieved with 3 μl DNA and this volume was used in

186 subsequent testing. PCR for the CR mini-barcode region was carried out under the following conditions:

187 94ºC for 2 min; 35 cycles of 94ºC for 30 s, 49ºC for 40 s, and 72ºC for 1 min; and a final extension of 72ºC

188 for 10 min. Thermocycling for ITS1 included the same parameters except that the annealing temperature was

189 adjusted to 61ºC. Thermocycling was carried out using a Mastercycler nexus gradient thermal cycler

190 (Eppendorf). PCR products (10 μl) were loaded onto pre-cast 2.0% E-Gel agarose gels (Life Technologies,

191 Carlsbad, CA) and the gels were run for 10 min with an E-Gel Powerbase (Life Technologies). The results

192 were photographed with a Transilluminator FBDLT-88 (Fisher Scientific, Hampton, New Jersey) and a

193 FOTO/Analyst Express (Fotodyne, Hartland, WI) and visualized with PCIMAGE (version 5.0.0.0 Fotodyne).

194 PCR products then underwent cleanup using ExoSAP-IT (Affymetrix, Santa Clara, CA) according to the

195 manufacturer's instructions. All samples and negative controls were then sent to GenScript (Piscataway, NJ)

196 for bi-directional sequencing using M13 primers (Table 1) with BigDye Terminator v3.1 Cycle Sequencing

197 Kit (Life Technologies) and a 3730xl Genetic Analyzer (Life Technologies).

198 *Sequence analysis*

199 Sequences were assembled and edited using Geneious R7 (Biomatters, Ltd., Auckland, New Zealand)

200 (http://www.geneious.com, Kearse et al. 2012). Consensus sequences were aligned using ClustalW with the

201 default settings in Geneious R7 and trimmed to the target CR or ITS1 fragments. The consensus sequence

202 lengths, number of ambiguities, and percent high quality bases (HQ%) were recorded for each sample.

203 Sequences were queried in GenBank using the Basic Local Alignment Search Tool (BLAST) and the top

204 species matches were recorded. Samples that could not be identified at the species level with BLAST were

205 subjected to phylogenetic analysis using MEGA 5.2, as described above. Any sample found to be potentially

206 mislabeled was subjected to a second round of DNA extraction, PCR, and sequencing to confirm the initial

207 result.

208 **Results and Discussion**

209 *Phylogenetic analysis*

210 Phylogenetic analysis of the CR mini-barcode targeted by the primers developed in this study revealed

211 clear differentiation for ten of the thirteen tuna species for which sequences were available, with strong

212    bootstrap support (Fig. 1). While non-introgressed sequences showed clear differentiation across all thirteen

213    species, inclusion of the introgressed sequences resulted in an inability to separate *T. thynnus* from *T.*

214    *orientalis* and an inability to differentiate albacore-like *T. thynnus* from *T. alalunga.* Although the inability

215    to differentiate introgressed individuals of *T. thynnus* and *T. orientalis* is a limitation of this methodology,

216    bluefin tuna are highly-valued and are almost exclusively prepared as a delicacy in sashimi and sushi dishes

217    (SeafoodHealthFacts 2015). Since neither of these species is typically found in canned tuna products, a

218    longer CR fragment (~450 bp) previously found to differentiate these two species could be used for

219    identification of bluefin tuna in raw or lightly processed tuna products (Viñas and Tudela 2009). On the

220    other hand, *T. alalunga* is commonly found in canned tuna and, although most reference sequences of this

221    species grouped in a species-specific cluster (Fig. 1), one sequence grouped more closely to an albacore-like

222    *T. thynnus* sequence. The inability to differentiate *T. alalunga* and albacore-like *T. thynnus* has been reported

223    previously for the ~450 bp region of CR targeted by Viñas and Tudela (2009). Although introgression of

224    these two species is known to occur, it is an uncommon event, with only 2-3% of *T. thynnus* showing an

225    identical sequence to *T. alalunga* (Viñas and Tudela 2009). However, this could be problematic if a tuna

226    product labeled as containing *T. alalunga* showed a top sequence match to *T. thynnus*. In order to verify

227    species in this case, additional analysis would be recommended using a nuclear DNA target. In this study,

228    ITS1 was chosen as the supplemental nuclear DNA target, due to a previous study reporting the ability to

229    differentiate *T. alalunga* and albacore-like *T. thynnus* using a ~650 bp region of the ITS1 region (Viñas and

230    Tudela 2009). In order to allow for identification in a canned tuna product, primers targeting a 179-bp region

231    within ITS1 were designed in the current study. As shown in Figure 2, phylogenetic analysis of *T. alalunga*

232    and *T. thynnus* reference sequences for this region resulted in clear separation between the two species, with

233    all albacore-like *T. thynnus* sequences grouping within the *T. thynnus* cluster (100% bootstrap support).

234         Among species groups for which more than one CR reference sequence was available, the average

235    intraspecies genetic variation for the CR mini-barcode ranged from $0.60 \pm 0.39\%$ for *T. obesus* to $12.54 \pm$

236    $2.32\%$ for *K. pelamis*, not including introgressed sequences. Similarly, Cawthorn et al. (2011) previously

237    reported *T. obesus* to show the lowest average intraspecies variation ($0.46 \pm 0.08\%$) among *Thunnus* species

238    for a 450-bp fragment of the CR. In order for a DNA target to be used for species differentiation with DNA

239    barcoding, the maximum intraspecies variation must be less than the minimum interspecies variation. This is

240 commonly referred to as a DNA barcode gap and can be represented graphically by plotting the maximum

241 intraspecies divergence on the x-axis and the minimum interspecies divergence on the y-axis (Rasmussen et

242 al. 2009). Species that have a DNA barcode gap will then be represented by data points that fall above the 1:1

243 ratio line between these axes, while species with data points falling below the line cannot be differentiated

244 with the gene target. As shown in Figure 3, a barcode gap was present for the CR mini-barcode region

245 targeted in this study for four of the seven species that could be analyzed in this manner. Due to the inclusion

246 of introgressed sequences, *T. orientalis, T. thynnus,* and *T. alalunga* did not have barcode gaps for the CR

247 mini-barcode region. However, when introgressed sequences were excluded, all seven species showed a CR

248 mini-barcode gap (results not shown). DNA barcode gaps could not be determined for species with only one

249 reference sequence available, as intraspecies variation could not be calculated in these cases. Use of the ITS1

250 supplementary marker to differentiate *T. alalunga* from albacore-like *T. thynnus* revealed the presence of a

251 barcode gap (Fig. 3). These results indicate that the ITS1 fragment targeted in this study could be used for

252 species confirmation in instances where a sample labeled as albacore tuna shows a top species match to *T.*

253 *thynnus* based on the CR mini-barcode.

254 *Sequencing results*

255     Of the 53 samples tested in this study with the CR mini-barcode, 26 showed successful PCR

256 amplification based on the results of gel electrophoresis and 24 were successfully bi-directionally sequenced

257 (Table 2). This includes 23 samples extracted initially and 1 sample labeled as "white albacore tuna" that

258 underwent a repeat DNA extraction with the MP FastPrep-24 instrument. The sequences had an average

259 consensus length of $232 \pm 14$ bp, average HQ% of $82.6 \pm 22.2\%$, and an average ambiguity percentage of

260 $0.20 \pm 0.49\%$. The quality of these sequences was slightly lower than previous species identification studies

261 involving the full-length COI barcode, which have reported averages of 87.5-93.6%HQ and 0.05-0.14%

262 ambiguous bases (Kane and Hellberg 2016; Quinto et al. 2016). The difference in quality is likely due to the

263 highly processed nature of the canned tuna products, as compared to uncooked meat products examined in

264 the previous studies. Among the 24 canned tuna products for which sequences were obtained, 21 could be

265 identified at the species level based on the results of BLAST, with a query coverage of $\geq 98\%$ and at least a

266 95% identity match (Table 2).

267     Two of the products that could not be identified at the species level with BLAST (T18 and T50) were

268     labeled as either albacore or white tuna and showed 99% sequence identity to multiple GenBank entries for

269     both *T. alalunga* and *T. thynnus*. As shown in Figure 1, the CR mini-barcode sequences for these two

270     products grouped within the *T. alalunga*/albacore-like *T. thynnus* clade. In order to verify species, these two

271     canned tuna samples were subsequently sequenced with the ITS1 primers designed in this study (Table 1).  A

272     successful sequence was obtained for one (T50) of the two samples and this result allowed for a positive

273     identification of *T. alalunga*, with 100% sequence identity in GenBank.  Furthermore, as shown in Figure 2,

274     the ITS1 sequence for this sample showed a clear phylogenetic grouping within the *T. alalunga* cluster.

275     The third sample that could not be identified at the species level based on the results of BLAST was

276     labeled as yellowfin tuna (T14). The CR mini-barcode for this sample showed a 99% identity match to

277     thirteen GenBank entries for *T. albacares* and to one entry for *T. thynnus* (Accession ID DQ087565). While

278     it is possible that this represents an instance of introgression, previous studies reporting introgression

279     between these two species could not be found. The *T. thynnus* entry was published as part of a study that

280     sequenced *T. thynnus* larvae collected in the Gulf of Mexico (Carlsson et al. 2007). *T. albacares* is also

281     known to inhabit the Gulf of Mexico (Collette and Nauen 1983) and it is possible that this sample was

282     morphologically misidentified in its larval form. Furthermore, when the CR mini-barcode for T14 was

283     analyzed alongside the reference sequences used in this study, it showed a clear grouping within the *T.*

284     *albacares* clade (Fig. 1). Therefore, based on the combination of BLAST and phylogenetic results, the

285     sample was determined to be *T. albacares*.

286     Among the two successfully sequenced samples labeled as "tuna" or "light tuna", one was identified as *K.*

287     *pelamis* and the other was identified as *T. alalunga,* respectively. Both of these species are listed in the CFR

288     for canned tuna, so these two products are considered to be properly labeled (21 CFR 161.190). However, the

289     use of *T. alalunga* in a product labeled as "light tuna" could be misleading to consumers that are intentionally

290     limiting their consumption of *T. alalunga* due to mercury concerns, as products containing *T. alalunga* are

291     typically associated with the designation of "white tuna" (Burger and Gochfeld 2004).

292     One instance of mislabeling was discovered in this study, with a sample labeled as tongol tuna identified

293     as striped bonito (*Sarda orientalis*) with 98% sequence identity. This result was confirmed upon re-

294     sequencing and the sample was determined to be mislabeled, as striped bonito is not listed in the CFR for

295     canned tuna. In this case, it is possible that *S. orientalis* was accidentally harvested alongside tongol tuna and

296     processed into cans labeled as tongol tuna, considering that striped bonito often school with small tunas in

297     the Indian and Pacific Oceans (Collette et al. 2011). Although this specific substitution was not reported

298     previously, *S. orientalis* has been known to be processed as canned tuna (Lin and Hwang 2007). The rate of

299     mislabeling found in this study (4.5%) is similar to a previous study on canned salmon, which tested 80

300     products and found 5% to be mislabeled (Rasmussen Hellberg et al. 2011). Another study investigating

301     canned tuna products in European countries reported a higher mislabeling rate of 15%, with mislabeling

302     found in products labeled as *T. albacares*, *A. rochei*, and *A. thazard* (Espiñeira et al. 2009).

303     In this study, the sequencing success rate was highest among products labeled as albacore or white tuna

304     (64.3%), followed by products labeled as yellowfin tuna (50%) and tongol tuna (50%). During primer

305     optimization, amplification success was obtained with albacore tuna (canned and fresh/frozen) as well as

306     skipjack tuna (canned). However, only one of ten commercial samples labeled as skipjack tuna was

307     successfully sequenced and only one of the five products labeled as light tuna was successfully sequenced.

308     Additional optimization based on sequencing success in both fresh/frozen and canned samples may help to

309     improve the performance of the assay. Interestingly, sequencing success in the current study was observed

310     more frequently in samples containing oil (57.9% success), including olive, canola, and soybean oil,

311     compared to samples canned in water with no oil (42.9% success) or samples canned in water and vegetable

312     broth with no oil (18.2% success). Similarly, Chapela et al. (2007) reported that higher quantities of DNA

313     were obtained from tuna samples canned in oil than from samples canned in vinegar, brine, or tomato sauce,

314     suggesting a protective effect of oil. Because most samples with sequencing failure also showed a negative

315     result following gel electrophoresis, there appears to have been a problem with PCR amplification of the

316     target region. Possible explanations for this include the presence of PCR inhibitors, lack of primer-binding

317     and DNA fragmentation from the canning process. Most samples that failed sequencing had short sequence

318     reads (~ 100 bp) that were poor quality and did not pass assembly. These reads showed a similar pattern of

319     chromatogram peaks as the first ~100 bp for the sequence reads which were successfully assembled,

320     suggesting that the primers were binding but that the target fragment was not completely elongated. Similarly,

321     previous studies have reported limited success in amplifying 200-400 bp fragments of DNA from canned

322     tuna (Lin and Hwang 2007; Quinteiro et al. 1998; Unseld et al. 1995). However, when these studies targeted

323 shorter regions of DNA ranging from 123 to 176 bp in length, amplification was successful, indicating that

324 DNA may be fragmented to less than 200 bp during the canning process. In the few cases where no sequence

325 reads were recovered, a possible explanation for sequencing failure could be the presence of a non-tuna

326 species that could not be amplified by the CR primers.

327 **Conclusions**

328 In the current study, a DNA mini-barcoding system was developed for the identification of fish species

329 commonly found in canned tuna. Phylogenetic analysis revealed that a 236-bp CR mini-barcode could

330 differentiate all 13 tuna species examined, except in rare cases of introgression involving *T. thynnus*/*T.*

331 *alalunga* and *T. thynnus*/*T. orientalis*. Use of a supplemental ITSI marker allowed for the differentiation of

332 introgressed individuals of *T. alalunga* and *T. thynnus.* Although the methodology developed in this study

333 does not allow for the differentiation of introgressed individuals of *T. thynnus* and *T. orientalis,* these species

334 are not typically found in canned tuna, and can be identified with a previously identified stretch of the CR.

335 Laboratory testing of the CR primers developed in this study demonstrated successful identification for *T.*

336 *alalunga, K. pelamis, T. tonggol,* and *T. albacares* in canned tuna, as well as *S. orientalis*, which is not listed

337 on the CFR for canned tuna and was detected in a mislabeled product. However, DNA sequencing was

338 unsuccessful in a number of products, likely due to factors such as DNA fragmentation and PCR inhibitors

339 present in canned tuna. Use of a shorter genetic region within the CR and/or further optimization of the assay

340 may help to improve PCR amplification and sequencing success.

341 **Acknowledgements**

345 **Compliance with Ethical Standards**

347 Conflict of Interest: Rosalee S. Hellberg declares that she has no conflict of interest. Jacquelyn K. Mitchell

348 declares that she has no conflict of interest.

349 Ethical approval: This article does not contain any studies with human participants or animals performed by

350 any of the authors.

351 Informed consent: Not applicable.

352 **References**

353 Alvarado Bremer JR, Viñasa J, Mejutoc J, Elyd B, Plae C (2005) Comparative phylogeography of Atlantic

354     bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression,

355     and population expansion on the regional phylogenies of two highly migratory pelagic fishes.

356     Molecular Phylogenetics and Evolution 36:169-187

357 Burger J, Gochfeld M (2004) Mercury in canned tuna: white versus light and temporal variation. Environ Res

358     96:239-249

359 Carlsson J, McDowell JR, Carlsson JE, Graves JE (2007) Genetic identity of YOY bluefin tuna from the

360     eastern and western Atlantic spawning areas. J Hered 98:23-28

361 Carlsson J, McDowell JR, Diaz-Jaimes P, Carlsson JE, Boles SB, Gold JR, Graves JE (2004) Microsatellite

362     and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population

363     structure in the Mediterranean Sea. Mol Ecol 13:3345-3356

364 Cawthorn D-M, Steinman HA, Witthuhn RC (2011) Establishment of a mitochondrial DNA sequence

365     database for the identification of fish species commercially available in South Africa. Mol Ecol

366     Resour 11:979-991

367 Chow S, Nakagawa T, Suzuki N, Takeyama H, Matsunaga T (2006) Phylogenetic relationships among

368     *Thunnus* species inferred from rDNA ITS1 sequence. J Fish Biol 68:24-35

369 Chuang P-S, Chen M-I, Shiao J-C (2012) Identification of tuna species by a real-time polymerase chain

370     reaction technique. Food Chem 133:1055-1061

371 Collete BB (1999) Mackerels, molecules, and morphology. In: Seret B, Sire J-Y (eds) Proc. 5th Indo-Pac.

372     Fish Conf., Noumea, Paris. pp 149-164

373 Collette B et al. (2011) *Sarda orientalis*. The IUCN Red List of Threatened Species. v. 2015.2

374     http://www.iucnredlist.org/details/170313/0.

375 Collette BB, Nauen CE (1983) FAO Species Catalogue. Vol. 2. Scombrids of the world. An annotated and

376     illustrated catalogue of tunas, mackerels, bonitos and related species known to date. Rome: FAO.

377     FAO Fish. Synop. 125(2):137 p.

378  EFSA (2004) European Food Safety Authority. Opinion of the Scientific Panel on contaminants in the food

379     chain related to the toxicity of fishery products belonging to the family of Gempylidae. EFSA

380     Journal 92:1-5

381  Espiñeira M, Gonzalez-Lavin N, Vieites JM, Santaclara FJ (2009) Development of a method for the

382     identification of scombroid and common substitute species in seafood products by FINS. Food

383     Chem 117:698-704

384  FDA/EPA (2014) Fish: What Pregnant Women and Parents Should Know.

385     http://www.fda.gov/Food/FoodborneIllnessContaminants/Metals/ucm393070.htm. Accessed

386     8/2/2015

387  Handy SM et al. (2011a) A single laboratory validated method for the generation of DNA barcodes for the

388     identification of fish for regulatory compliance. J AOAC 94:201-210

389  Handy SM et al. (2011b) Single laboratory validated method for DNA-barcoding for the species

390     identification of fish for FDA regulatory compliance.  U.S. Food and Drug Administration Standard

391     Operating Procedure.  Updated September 2011.  Accessible at

392     http://www.fda.gov/food/foodscienceresearch/dnaseafoodidentification/ucm237391.htm.  Accessed

393     6/15/2015.

394  Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. P

395     Roy Soc Lond B Bio 270:313-321

396  ITIS (2015) Integrated Taxonomic Information System on-line database. http://www.itis.gov. Accessed

397     November 16, 2015

398  Jacquet JL, Pauly D (2008) Trade secrets: Renaming and mislabeling of seafood. Mar Policy 32:309-318

399  Kane DE, Hellberg RS (2016) Identification of species in ground meat products sold on the U.S. commercial

400     market using DNA-based methods. Food Control 59:158-163

401  Kearse M et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the

402     organization and analysis of sequence data. Bioinformatics 28:1647-1649

403  Kimura M (1980) A simple method of estimating evolutionary rate of base substitutions through comparative

404     studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120

405     Lin WF, Hwang DF (2007) Application of PCR-RFLP analysis on species identification of canned tuna.

406         Food Control 18:1050-1057

407     Lowenstein JH, Amato G, Kolokotronis SO (2009) The real maccoyii: identifying tuna sushi with DNA

408         barcodes – contrasting characteristic attributes and genetic distances. PLoS ONE 4:e7866

409     NFI (2014) National Fisheries Institute. Top 10 Consumed Seafoods. Raw data from the National Marine

410         Fisheries Service. https://www.aboutseafood.com/about/about-seafood/top-10-consumed-seafoods.

411         Accessed 11/17/2015

412     NMFS (2015) National Marine Fisheries Service. Fisheries of the United States, 2014. U.S. Department of

413         Commerce, NOAA Current Fishery Statistics No. 2014. https://www.st.nmfs.noaa.gov/commercial-

414         fisheries/fus/fus14/index. Accessed 11/17/15

415     Quinteiro J et al. (1998) Use of mtDNA direct polymerase chain reaction (PCR) sequencing and PCR-

416         restriction fragment length polymorphism methodologies in species identification of canned tuna. J

417         Agric Food Chem 46:1662-1669

418     Quinto CA, Tinoco R, Hellberg RS (2016) DNA barcoding reveals mislabeling of game meat species on the

419         U.S. commercial market. Food Control 59:386-392

420     Rasmussen Hellberg RS, Morrissey MT (2011) Advances in DNA-based techniques for the detection of

421         seafood species substitution on the commercial market. JALA 16:308-321

422     Rasmussen Hellberg RS, Naaum AM, Handy SM, Hanner RH, Deeds JR, Yancy HF, Morrissey MT (2011)

423         Interlaboratory evaluation of a real-time multiplex polymerase chain reaction method for

424         identification of salmon and trout species in commercial products. J Agric Food Chem 59:876-884

425     Rasmussen RS, Morrissey MT (2008) DNA-based methods for the identification of commercial fish and

426         seafood species. Compr Rev Food Sci F 7:280-295

427     Rasmussen RS, Morrissey MT (2009) Application of DNA-based methods to identify fish and seafood

428         substitution on the commercial market. Compr Rev Food Sci F 8:118-154

429     Rasmussen RS, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and

430         trout species (*Oncorhynchus* and *Salmo*) from North America. J Agric Food Chem 57:8379-8385

431     Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees.

432         Mol Biol Evol 4:406-425

433 SeafoodHealthFacts (2015) Seafood Health Facts: Making Smart Choices. Canned Tuna.

434   http://seafoodhealthfacts.org/seafood_choices/canned_tuna.php. Accessed 11/16/15

435 Shokralla S, Hellberg RS, Handy SM, King I, Hajibabaei M (2015) A DNA mini-barcoding system for

436   authentication of processed fish products. Scientific Reports 5:e15894

437 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary

438   genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony

439   methods. Mol Biol Evol 28:2731-2739

440 Unseld M, Beyermann B, Brandt P, Hiesel R (1995) Identification of the species origin of highly processed

441   meat products by mitochondrial DNA sequences. PCR Methods and Applications 4:241-243

442 Viñas J, Tudela S (2009) A validated methodology for genetic identification of tuna species (Genus *Thunnus*).

443   PLoS ONE 4:e7606

444 Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol

445   74:329-356

446 Warner K, Timme W, Lowell B, Hirshfield M (2013) Oceana study reveals seafood fraud nationwide.

447   Washington, DC: Oceana.

448 Wong EH-K, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood.

449   Food Res Int 41:828-837

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466 **Table 1.** Primers used in this study. All primers include M13 tails described previously (Handy et al. 2011a).

| Primer set | Primer name | Primer sequence (5'-3') | Described | Ta | Target fragment length[a] |
|---|---|---|---|---|---|
| CR | Tuna CR_F | CAC GAC GTT GTA AAA CGA CGC AYG TAC ATA TAT GTA AYT ACA CC | In this work | 49ºC | 236 bp |
| | Tuna CR_R1 | GGA TAA CAA TTT CAC ACA GGC TGG TTG GTR GKC TCT TAC TRC A | In this work | | |
| | Tuna CR_R2 | GGA TAA CAA TTT CAC ACA GGC TGG ATG GTA GGY TCT TAC TGC G | In this work | | |
| ITS1 | ITS1_F | CAC GAC GTT GTA AAA CGA CTC TCC TGG TCA GGA CCT CGT | In this work | 61ºC | 179 bp |
| | ITS1_R | GGA TAA CAA TTT CAC ACA GGA AGC CTC CGC TKC CGC GCT T | In this work | | |
| M13 | M13F (-29) | CAC GAC GTT GTA AAA CGA C | Handy et al. (2011a) | N/A | N/A |
| | M13R | GGA TAA CAA TTT CAC ACA GG | Handy et al. (2011a) | N/A | N/A |

467 [a]Not including primers

468

469

19

470    **Table 2.** Summary of the 53 canned tuna products analyzed in this study with the CR mini-barcode. Top species matches were determined using BLAST.

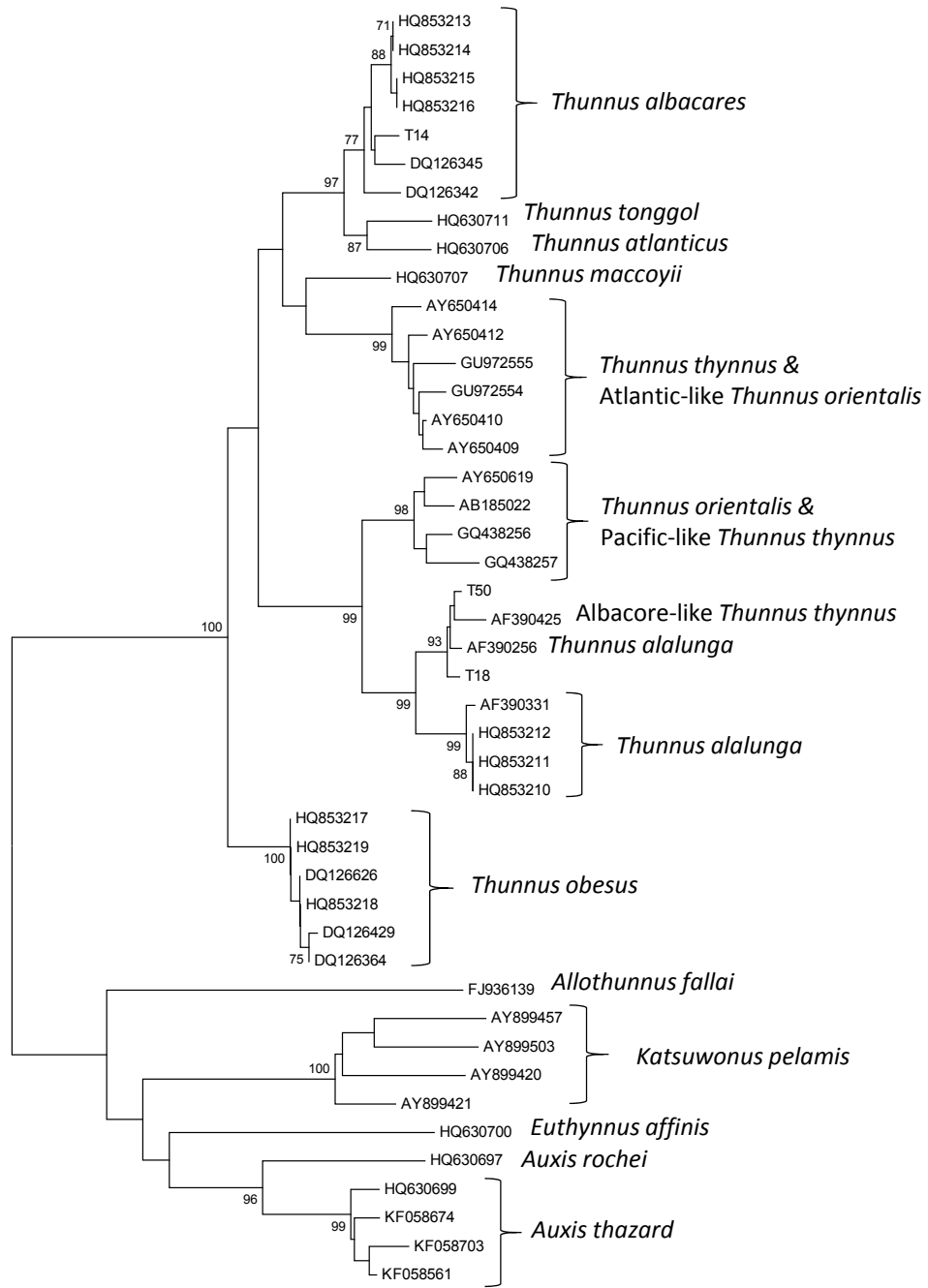| Product label | n | Expected species | Top species match in GenBank | % Sequence identity |
|---|---|---|---|---|
| Albacore/white tuna | 7 | *T. alalunga* | *T. alalunga* | 97-99% |
| Albacore/white tuna | 2 | *T. alalunga* | *T. alalunga/T. thynnus* | 99% |
| Albacore/white tuna | 5 | *T. alalunga* | Failed sequencing | N/A |
| Light tuna | 1 | Variety of potential species | *T. alalunga* | 95% |
| Light tuna | 4 | Variety of potential species | Failed sequencing | N/A |
| Skipjack tuna | 1 | *K. pelamis* | *K. pelamis* | 97% |
| Skipjack tuna | 9 | *K. pelamis* | Failed sequencing | N/A |
| Tongol tuna | 2 | *T. tonggol* | *T. tonggol* | 97-98% |
| Tongol tuna[a] | 1 | *T. tonggol* | *S. orientalis* | 98% |
| Tongol tuna | 3 | *T. tonggol* | Failed sequencing | N/A |
| Tuna | 1 | Variety of potential species | *K. pelamis* | 97% |
| Tuna | 3 | Variety of potential species | Failed sequencing | N/A |
| Yellowfin tuna | 8 | *T. albacares* | *T. albacares* | 99-100% |
| Yellowfin tuna | 1 | *T. albacares* | *T. albacares/T. thynnus* | 99% |
| Yellowfin tuna | 5 | *T. albacares* | Failed sequencing | N/A |

471    [a] Mislabeled sample.

472

473 **Figure Captions**

474 **Figure. 1** Neighbor-joining tree of the 236-bp CR mini-barcode targeted in this study. GenBank accession

475 numbers are shown for all reference sequences. The Kimura 2-parameter method was used to calculate

476 genetic distances and bootstrap values greater than 70% are shown. The tree includes three of the canned

477 commercial samples tested in the current study (T14, T18, T50)

478 **Figure. 2**. Neighbor-joining tree of the 179-bp ITS1 fragment targeted in this study. GenBank accession

479 numbers are shown for all reference sequences and *Katsuwonus pelamis* was used as an outgroup. The

480 Kimura 2-parameter method was used to calculate genetic distances and bootstrap values greater than 70%

481 are shown. The tree includes one of the canned commercial samples tested in the current study (T50)

482 **Figure. 3** Mini-barcode gaps for the tuna species targeted in this study, including introgressed sequences.

483 Data points are representative of the CR mini-barcode unless otherwise noted. Genetic distance was

484 calculated using the Kimura 2-parameter method. Species with only one reference sequence are not shown

71 HQ853213
88 HQ853214
HQ853215
HQ853216
77 T14
DQ126345
97 DQ126342

*Thunnus albacares*

HQ630711 *Thunnus tonggol*
87 HQ630706 *Thunnus atlanticus*

HQ630707 *Thunnus maccoyii*

AY650414
99 AY650412
GU972555
GU972554
AY650410
AY650409

*Thunnus thynnus* &
Atlantic-like *Thunnus orientalis*

AY650619
98 AB185022
GQ438256
GQ438257

*Thunnus orientalis* &
Pacific-like *Thunnus thynnus*

T50
AF390425   Albacore-like *Thunnus thynnus*
93 AF390256   *Thunnus alalunga*
T18

99 AF390331
HQ853212
99 HQ853211
88 HQ853210

*Thunnus alalunga*

HQ853217
HQ853219
DQ126626
HQ853218
DQ126429
75 DQ126364

*Thunnus obesus*

FJ936139   *Allothunnus fallai*

AY899457
AY899503
100 AY899420
AY899421

*Katsuwonus pelamis*

HQ630700   *Euthynnus affinis*
HQ630697   *Auxis rochei*

96 HQ630699
KF058674
99 KF058703
KF058561

*Auxis thazard*

100

99

100

0.05

GQ414555    Albacore-like *Thunnus thynnus*

AB212012    *Thunnus thynnus*

GQ414554

GQ414561    *Thunnus thynnus*

GQ414560    Albacore-like *Thunnus thynnus*

AB212008

AB127401

AB212009    *Thunnus thynnus*

AB212010

AB212011

100

AB212003

AB212006

AB211999

AB212000

AB212004

AB212005

T50

AB212001

AB127402

AB212002

GQ414557

88

*Thunnus alalunga*

AB212043    *Katsuwonus pelamis*

0.02