1   **Highlights**

- 2   • *cytb* targeted NGS allowed the identification of fish species in mixed
- 3   tuna samples.
- 4   • Sequence recovery from prepared tuna mixtures was very precise.
- 5   • Skipjack tuna was overrepresented compared to *Thunnus* species.
- 6   • Phylogenetic grouping of recovered sequences was suitable for tuna tin
- 7   analysis.
- 8   • This technology is a promising tool for tuna tin authentication in the
- 9   future.

1   **Title**

2   **Species identification in mixed tuna samples with next-generation**

3   **sequencing targeting two short cytochrome b gene fragments**

4

5   **Running title**

6   **NGS for mixed tuna samples**

7

8   **Authors**

9   Kristina Kappel^{a, 1, 2}, Ilka Haase^b, Christine Käppel^{b, 3}, Carmen G. Sotelo^c,
10   Ute Schröder^a

11

12   **Affiliations**

13   ^a Max Rubner-Institut (MRI), Department of Safety and Quality of Milk and
14   Fish Products, Palmaille 9, 22767 Hamburg, Germany
15   (kristina.kappel@mri.bund.de; ute.schroeder@mri.bund.de)

16   ^b Eurofins Genomics, Anzinger Str. 7a, 85560 Ebersberg, Germany
17   (ilkahaase@eurofins.com; christinekaeppel@eurofins.com)

18   ^c Instituto de Investigaciones Marinas, Superior de Investigaciones
19   Científicas (CSIC), Eduardo Cabello 6, 36208 Vigo, Spain
20   (carmen@iim.csic.es)

21

---

¹ Corresponding author. Tel.: +49 431 609 2253; Fax: +49 431 609 2300; E-mail: kristina.kappel@mri.bund.de

² Present address: Max Rubner-Institut (MRI), Hermann-Weigmann-Str. 1, 24103 Kiel, Germany

³ Present address: Uphoff GmbH, Im Gewerbegebiet 22, 83093 Bad Endorf, Germany

22   **Abstract**

23   Conventional Sanger sequencing of PCR products is the gold standard for
24   species authentication of seafood products. However, this method is
25   inappropriate for the analysis of products that might contain mixtures of
26   species, such as tinned tuna. The purpose of this study was to test whether
27   next-generation sequencing (NGS) can be a solution for the authentication
28   of mixed products. Nine tuna samples containing mixtures of up to four
29   species were prepared and subjected to an NGS approach targeting two
30   short cytochrome b gene (*cytb*) fragments on the Illumina MiSeq platform.
31   Sequence recovery was precise and admixtures of as low as 1% could be
32   identified, depending on the species composition of the mixtures. Duplicate
33   samples as well as two individual NGS runs produced very similar results.
34   A first test of three commercial tinned tuna samples indicated the presence
35   of different species in the same tin, although this is forbidden by EU law.

36

37   **Keywords**

- 38   •   Tuna mixtures
- 39   •   Tuna tins
- 40   •   Next-generation sequencing
- 41   •   Authentication
- 42   •   Cytochrome b
- 43

44   **1 Introduction**

45

46   With the advent of PCR based methods targeting specific DNA marker
47   regions, species authenticity testing of highly processed seafood like
48   cooked, hot smoked or even tinned products became possible, as DNA
49   remains stable during most manufacturing processes. With earlier
50   techniques the nature of DNA sequences was investigated indirectly, e.g. by
51   analysis of restriction fragment length polymorphisms (RFLPs) or single
52   strand conformation polymorphisms (SSCPs). Today the most frequently
53   used technology for species authentication is the determination of the
54   complete DNA sequence by PCR and conventional Sanger sequencing
55   aimed at DNA markers like the mitochondrial cytochrome b (*cytb*) or
56   cytochrome oxidase I (*coxI*) genes (Griffiths et al., 2014). These
57   “barcoding” methods have resulted in a variety of seafood mislabelling
58   studies (for review see Pardo, Jiménez, & Pérez-Villarreal (2015)) and new
59   surveys are constantly being published from all over the world (e.g.
60   Carvalho, Palhares, Drummond, & Frigo, 2015; Galal-Khallaf, Ardura,
61   Borrell, & Garcia-Vazquez, 2016; Mariani et al., 2015; Shokralla, Hellberg,
62   Handy, King, & Hajibabaei, 2015). However, Sanger sequencing
63   technology is inappropriate for the analysis of seafood products that might
64   contain mixtures of different species, like tuna tins or surimi products etc.,
65   because the presence of DNA templates featuring different sequences results
66   in overlapping peaks in the sequence chromatograms. These ambiguous
67   peaks can easily be overlooked or misinterpreted as background noise,
68   which leads to the incorrect assignment of product authenticity, although
69   undeclared species might have been added during manufacturing. In order to
70   obtain reliable sequence information from mixed samples, amplified gene
71   fragments have to be cloned prior to sequencing. This is a labour-intensive
72   approach, requiring sequencing of many clones in order to gain significant
73   results (Pompanon, Deagle, Symondson, Brown, Jarman, & Taberlet, 2012).

74 Next-generation sequencing (NGS) technologies might overcome this
75 problem through massively parallel sequencing of different template
76 molecules from one sample (Mardis, 2008), which makes NGS extremely
77 promising for the analysis of food products containing mixtures of seafood
78 species.

79 NGS analysis of barcode marker sequences (“metabarcoding”) is already
80 common for environmental monitoring of microbial communities (Luo,
81 Tsementzi, Kyrpides, Read, & Konstantinidis, 2012) or invertebrate species
82 (Carew, Pettigrove, Metzeling, & Hoffmann, 2013; Diaz-Real, Serrano,
83 Píriz, & Jovani, 2015) and was very recently proposed as the next-
84 generation tool for ecological studies and standardised biodiversity
85 monitoring of aquatic ecosystems (Valentini et al., 2016). Early studies
86 investigated the potential of this approach for the identification of animal or
87 plant ingredients in food products like meat samples (Bertolini, Ghionda,
88 D’Alessandro, Geraci, Chiofalo, & Fontanesi, 2015; Tillmar, Dell’Amico,
89 Welander, & Holmlund, 2013) or in Traditional Chinese Medicines (Cheng
90 et al., 2014; Coghlan et al., 2012). However, to our knowledge, there are
91 only very few NGS studies targeting seafood samples, so far. De Battisti,
92 Marciano, Magnabosco, Busato, Arcangeli and Cattoli (2014) used
93 pyrosequencing for market samples representing individual specimens as a
94 fast and cheap method for food control. In another study Park et al. (2012)
95 identified fish species in fish cakes using NGS and microarray analysis, but
96 the authors do not provide a lot of information about the details of the
97 methods as well as the NGS results.

98 The tribe of tuna species (*Thunnini*, family *Scombridae*) is one of the most
99 valuable fish groups for consumption and consumers are willing to pay high
100 prices for particular species, especially for true tunas (*Thunnus* species) like
101 albacore (*T. alalunga*) or Atlantic bluefin tuna (*T. thynnus*). Indeed, the
102 worldwide demand for *Thunnus* species has led to a serious decrease of fish
103 population sizes with certain species being either classified as vulnerable,

104 threatened, endangered or even critically endangered, according to IUCN
105 criteria (Restrepo, Juan-Jordá, Collette, Frédou, & Rosenberg, 2016).
106 Tinned tuna is a highly appreciated product in a lot of countries like the
107 United States, the United Kingdom and many other European countries
108 (Leadbitter & Benguerel, 2014). There is a specific European regulation,
109 laying down common European marketing standards for preserved tuna and
110 bonito (Council Regulation (EEC) No 1536/92). This regulation clearly
111 states that preserved tuna products must be prepared either from *Thunnus*
112 species or from skipjack tuna (*Katsuwonus pelamis*) and that different
113 species may not be mixed in the same container. However, the latter is not
114 mandatory for products in which the muscular structure has disappeared.
115 Despite these legal requirements, analysis of tuna tins with PCR and Sanger
116 sequencing sometimes leads to the assumption, that different species have
117 been mixed in one tin because of the presence of overlapping chromatogram
118 peaks, even after repeated Sanger sequencing reactions.

119 The aim of this study was to test whether a *cytb* targeted NGS reaction on
120 the Illumina MiSeq platform is suitable for the identification of tuna species
121 within prepared mixtures of different fish species. In addition to that and
122 considered as an initial test, NGS of *cytb* segments followed by
123 phylogenetic analysis of *cytb* sequences was subsequently tested on three
124 tinned tuna products.

125   **2 Materials and methods**

126   2.1 Sample preparation

127   Nine samples were prepared, containing muscle tissue of albacore (*T.*
128   *alalunga*), yellowfin tuna (*T. albacares*), bigeye tuna (*T. obesus*) and
129   skipjack tuna (*K. pelamis*) as well as Atlantic bonito (*Sarda sarda*) in
130   varying compositions and proportions (see Table 1). Albacore (*T. alalunga*)
131   was purchased directly from the producer as frozen fillets. Bigeye tuna (*T.*
132   *obesus*) and yellowfin tuna (*T. albacares*) were purchased as fresh fillets
133   and skipjack tuna (*K. pelamis*) and Atlantic bonito (*S. sarda*) as fresh whole
134   specimens at local markets. Authenticity of all samples was verified by
135   conventional Sanger sequencing of a PCR-amplified 413 nt fragment of the
136   *cytb* gene as described in Kappel and Schröder (2015b).

137   Muscle tissue of each fish sample was homogenised separately with a
138   Grindomix GM 200 (Retsch, Haan, Germany) for larger quantities (70 to
139   180 g), or with the Dispomix system (Medic Tools AG, Zug, Switzerland)
140   for lower quantities (up to 60 g). Purées of different species were mixed
141   afterwards in defined weight proportions. These mixtures (35 g) were
142   homogenised again, by thoroughly grinding them in plastic mortars with
143   plastic pestles. Homogenous mixtures were split into duplicates of 13 g each
144   as well as an additional retention sample of 5 g. Samples were marked with
145   encrypted IDs and the presence of duplicates was not communicated to the
146   company, which was to perform the targeted NGS reaction. All steps were
147   performed on ice if possible and samples were also held on ice until further
148   processing (about 24 h).

149

150   2.2 DNA isolation, *cytb*-PCRs and conventional Sanger sequencing

151   In order to be able to verify the recovered sequences of the subsequent NGS
152   analysis, the *cytb* marker sequences were initially determined with
153   conventional Sanger sequencing for each of the utilised tuna specimens.

154 Total DNA of individual tuna specimens was isolated from 50 to 100 mg of
 155 fish muscle with CTAB as described by Rehbein (2005) and DNA
 156 concentrations were measured as fluorescence enhancement following
 157 binding of the dye Hoechst 33258 (Downs & Wilfinger, 1983) with calf
 158 thymus DNA as a DNA standard. Two short *cytb* fragments (“BDR” and
 159 “BMID”) were amplified in independent PCRs with the primers BDR-L
 160 (5'-GCMAACGGSGCNTCYTTCTTCT-3') and BDR-H-mod1
 161 (5'-TGACGGTAGCHCCTCAGRADGACATTTGTCC YCA-3')
 162 (unmodified and modified, respectively, according to González Sotelo,
 163 Medina Méndez, Pérez Martín, Quinteiro Vázquez, and Rey-Méndez
 164 (2002)), as well as BMID-L-mod1
 165 (5'-ATCYCATTCACCCATACTWCTC-3') and BMID-H-mod1
 166 (5'-AATAGGAARTATCATTCRGGTTTTRATG-3') (modified according to
 167 L15424 and H15573 from Meyer (1993)), respectively. PCR reactions were
 168 performed in volumes of 20 µl containing about 20 ng extracted DNA, 10 µl
 169 HotStarTaq *Plus* Master Mix (Qiagen, Hilden, Germany), 10 pmol of
 170 forward and reverse primer (Biometra, Göttingen, Germany) each and a
 171 total concentration of 2 mM MgCl₂. Temperature profiles were as follows:
 172 5 min 95 °C followed by 35 cycles of 20 s 94 °C, 30 s 55 °C and 20 s 72 °C
 173 for BDR-PCR and 5 min 95 °C followed by 35 cycles of 40 s 94 °C, 80 s 50
 174 °C and 80 s 72 °C for BMID-PCR. Reactions were finished with a final
 175 extension step of 7 min (BDR) or 10 min (BMID) at 72 °C. Fragment
 176 lengths without primer sequences were calculated to be 131 nt for BDR and
 177 126 nt for BMID respectively. Presence and lengths of PCR amplicons were
 178 checked on 2% (w/v) agarose gels. Subsequent Sanger sequencing of 1:10
 179 diluted amplicons was performed by LGC Genomics (Berlin, Germany)
 180 without prior purification. Chromatograms were visually checked and edited
 181 in Chromas Lite version 2.1 (Technelysium Pty Ltd, South Brisbane,
 182 Queensland, Australia), primer sequences were removed and consensus
 183 sequences were assembled from forward and reverse strands.

184

185 2.3 Tuna tins

186 Three tuna tin samples were purchased from local markets (see Table 1).
187 Two of the samples were retrieved from the same product and batch but
188 represented different tins and were labelled as “light tuna in vegetable oil”,
189 whereas the third sample was labelled with the scientific name *Thunnus*
190 *alalunga* in addition to the specification of containing “tuna fillets, white, in
191 olive oil”. Oil was removed by draining the total contents of the tins on
192 paper towels. About 35 g of the drained tissues were thoroughly grinded for
193 homogenisation and afterwards were split into duplicates and additional
194 retention samples, just the same as with the prepared fish mixtures.

195

196 2.4 *cytb*-targeted NGS on the Illumina MiSeq platform

197 DNA extraction, PCR amplification with Illumina adaptor introduction and
198 Illumina NGS reaction were performed by Eurofins Genomics (Ebersberg,
199 Germany). In brief, DNA was extracted from 5 g of prepared fish mixture
200 samples or tuna tin samples using the NucleoSpin Food kit (Macherey
201 Nagel, Düren, Germany) according to the manufacturer’s instructions with
202 an overnight proteinase K incubation step. BDR and BMID fragments were
203 amplified in separate PCRs with the primers BDR-L and BDR-H-mod1 as
204 well as BMID-L-mod1 and BMID-H-mod1 (see above) respectively. The
205 primers featured additional Illumina adapter sequences at the 5’- end
206 including binding sites for hybridisation of PCR products to the flow cell,
207 binding sites for sequencing primers, poly(N)-regions and also barcode
208 sequences for unique sample identification.

209 Quality of PCR amplicons was analysed with agarose gel electrophoresis
210 and / or spectrophotometric analysis (Nanodrop spectrophotometer, Thermo
211 Scientific). Additionally, amplicon lengths and amounts were estimated by
212 capillary electrophoresis (Caliper LabChip GX, Perkin Elmer) and when

213 unexpected short fragments were present, PCR products were purified with
214 gel electrophoresis prior to submission to NGS reaction. Amplicons were
215 applied in equimolar ratios to bridge amplification and sequencing by
216 synthesis on the Illumina MiSeq platform (Illumina, San Diego, California,
217 USA) by generating 300 nt paired end reads. As indicated in Table 1, most
218 of the samples were submitted to two independent NGS runs in order to
219 assess the reaction's repeatability.

220

221 2.5 NGS data evaluation

222 Sequencing data were processed by Eurofins Genomics. This included
223 standard quality checks based on Phred-scores, sorting of reads according to
224 primer and barcode sequences and clipping of primer and barcode sequences
225 from raw reads. Only those read-pairs for which primer and barcode
226 sequences were detected without mismatches in both forward and reverse
227 reads were considered for further processing in order to improve assignment
228 of reads to particular samples. Only forward reads were further processed as
229 sequences from reverse reads and stitched data did not differ from those of
230 forward reads (data not shown). Remaining forward reads were clustered
231 using CD-HIT-EST (<http://weizhongli-lab.org/cd-hit/>) with 100% sequence
232 identity threshold (parameter "-c 100"). Only clusters which included at
233 least 1% of the remaining reads were considered for evaluation. The cluster
234 representative sequences as marked by CD-HIT-EST were used for all
235 further evaluations.

236 As automated trimming of sequences did not yield satisfying results, cluster
237 master sequences were aligned to tuna reference sequences and were
238 trimmed again manually to the expected lengths of the fragments. The
239 resulting sequences were subsequently compared to the initially determined
240 sequences of the fish specimens used as starting material for the preparation
241 of the fish mixtures in order to check the accuracy of sequence recovery. For

242 the analysis of the three tuna tins, recovered sequences were
243 phylogenetically grouped with reference sequences from scombrotoxin
244 specimens (CSIC in-house tissue collection). In brief, tuna tin NGS
245 sequences were aligned together with scombrotoxin reference sequences with
246 ClustalX 2.1 (Larkin et al., 2007) and a neighbour joining tree was
247 calculated with MEGA5.2 (Tamura, Peterson, Peterson, Stecher, Nei, &
248 Kumar, 2011) using Kimura 2-parameter distances. Monophyletic groups
249 were collapsed for better illustration, if appropriate, and species assignment
250 was conducted if NGS generated sequences were clustered together with
251 reference sequences belonging to particular species.

252 For the preparation of summarising diagrams, percentages of reads of
253 particular sequence clusters from duplicate samples and, if available, two
254 individual NGS runs were averaged and are displayed with standard
255 deviations. For tuna tin analysis, percentages of reads of all sequences
256 assigned to particular species were totalled.

257

258   **3 Results and discussion**

259

260   This work aimed to analyse the feasibility of NGS based identification of
261   mixtures of different fish species, which can occur in highly processed
262   seafood products such as tuna tins. Prepared mixtures of tissues belonging to
263   different scombroid species served as model samples with defined species'
264   compositions. These mixed samples were investigated by an initial PCR
265   amplification of two short *cytb* fragments, followed by a NGS reaction on
266   the Illumina MiSeq platform, a benchtop NGS device of relatively low
267   operating costs (Shokralla et al., 2015). Recovered sequences were
268   compared to those of utilised specimens and read percentages were
269   calculated for sequence clusters consisting of sequences with 100%
270   identities, due to the high similarity of sequences within the *Thunnus* genus
271   (Viñas & Tudela, 2009). The same procedure was applied to three
272   commercial tuna tin samples, followed by phylogenetic grouping of
273   recovered sequences together with sequences from scombroid voucher
274   species for species assignment.

275

276   3.1 Mixed tuna samples and sequence variability

277   Nine fish mixtures were prepared as duplicates from tissues belonging to
278   four tuna species (albacore, yellowfin tuna, bigeye tuna and skipjack tuna)
279   and another scombroid species (Atlantic bonito) (see Table 1). Whereas
280   most of preserved tuna products are labelled as skipjack tuna (at least in
281   Germany), higher priced tinned tuna produced with true tuna species like
282   albacore or yellowfin tuna is also available. However, for profit reasons the
283   tuna in these products might be prone to be substituted with less valuable
284   skipjack tuna, as was shown in a former market study in Germany (Kappel
285   & Schröder, 2015a). Bigeye tuna was included in the present study as this
286   species was often found to be used as substitute species for yellowfin tuna

287 as fresh fish fillets in the German survey. Atlantic bonito, on the other hand,
 288 is not allowed to be marketed under the commercial designation “tuna”
 289 according to EU law. Thus this species was chosen as an example of a
 290 potential substitute for tinned tuna products without species specification.
 291 Special care was given to the homogenisation of fish mixtures, because only
 292 proportions of the samples were used for DNA extraction, thus the cells of
 293 the individual species were required to be evenly distributed within the
 294 samples. Species numbers varied from one to four species per mixture and
 295 the lowest admixture quantity was 1% (weight/weight).

296 The *cytb* marker segments BDR (131 bp) and BMID (126 bp), which had
 297 been shown to be suitable for tuna tin analysis before (Mariani et al., 2015),
 298 were determined by Sanger sequencing for all utilised specimens and are
 299 shown in the supplementary material. The short lengths of the chosen gene
 300 markers were essential, as the DNA in tinned tuna products is highly
 301 degraded because of the heat-sterilisation process; and fragment lengths are
 302 expected not to exceed 200 bp (Quinteiro et al., 1998). In pairwise
 303 comparisons the DNA sequences from the species used for mixture
 304 preparation varied by four to 20 nucleotides (BDR) or two to 25 nucleotides
 305 (BMID), respectively (see supplementary material for nucleotide identity
 306 matrices and sequence alignments). The lowest BDR difference was found
 307 between albacore and yellowfin tuna (4 SNPs), whereas BMID sequences
 308 displayed highest identity between yellowfin tuna and bigeye tuna (2 SNPs).
 309 Sequence identity was significantly lower between *Thunnus* species and
 310 skipjack tuna or Atlantic bonito (19-20 SNPs for BDR and 14-18 SNPs for
 311 BMID, respectively). The high sequence similarity between *Thunnus*
 312 species was expected due to the close relations within that genus (Viñas &
 313 Tudela, 2009). However, considering the use of a combination of two
 314 fragments, the observed sequence differences should be sufficient to identify
 315 the species with the NGS approach. It is worth to mention, that the BMID
 316 fragment corresponds nearly exactly with the *cytb* fragment, which has been

317 suggested to be the most promising minibarcode for NGS based
318 identification of species of the order *Rodentia* (Galan, Pagès, & Cosson,
319 2012).

320

321 3.2 NGS of prepared tuna samples

322 Average read numbers per sample were 372,363 for BDR and 359,103 for
323 BMID in NGS run 1, and 358,560 for BDR and 390,188 for BMID in run 2,
324 respectively. Sequences from clusters (representing sequences with 100%
325 sequence identity; see supplementary material for cluster details) with $\geq 1\%$
326 of all reads after quality control were compared to BDR and BMID
327 sequences from specimens that had been used for mixture preparation. The
328 read percentages are shown in Fig. 1 as mean values from duplicate samples
329 as well as two NGS runs.

330 Apart from a few exceptions, all *cytb* sequences from the specimens used
331 for mixture preparation were exactly recovered when clusters with a
332 threshold of 1% of total reads were considered, and recovery for BDR and
333 BMID was quite consistent. Even the sample with only 1% admixture of
334 skipjack tuna was correctly identified with read percentages of 1.6% (BDR)
335 and 2.7% (BMID), respectively. In two samples (*T. alalunga*/*K. pelamis*
336 50:50 and *K. pelamis*/*S. sarda* 90:10) a second *K. pelamis*-BMID sequence
337 was identified in addition to the original skipjack BMID sequence, which
338 differed from the original sequence by one nucleotide. However, this signal
339 was only slightly above the threshold and was only generated in one of the
340 duplicate samples and also one out of two individual NGS runs and might
341 likely be a PCR or sequencing error. In the latter sample, in addition to that,
342 the BMID signal for Atlantic bonito was missing, although the BDR signal
343 could be identified. After a closer inspection of the raw data, it was obvious
344 that the BMID *S. sarda* signal was present, too, but was below the threshold
345 of 1% of sequence reads. One false positive BMID signal with 100%

346 identity to the original skipjack sequence was detected in the sample
 347 containing only albacore and yellowfin tuna in equal proportions
 348 (*T. alalunga*/*T. albacares* 50:50). This signal was also slightly above the
 349 threshold and was only detected in one sample and one read and might very
 350 likely be a minor cross-contamination from the mixture preparation process.

351 Despite the precise sequence recovery, it is important to note that the
 352 different species were not recovered uniformly with the targeted NGS
 353 reaction, a fact which was not unexpected as it is well known, that PCR
 354 often leads to biased multiplex amplifications (Kanagawa, 2003; Pawluczyk,
 355 Weiss, Links, Aranguren, Wilkinson, & Egea-Cortines, 2015). In NGS
 356 applications there are even more sources for bias introduction (Pompanon,
 357 Deagle, Symondson, Brown, Jarman, & Taberlet, 2012). Despite this fact,
 358 other authors reported a high precision of abundance estimation of species
 359 with a targeted NGS approach (Diaz-Real et al., 2015). In the present study
 360 skipjack tuna was always considerably overrepresented compared to the
 361 *Thunnus* species. Of all investigated *Thunnus* species recovery of albacore
 362 and bigeye tuna was quite similar but exceeded that of yellowfin tuna.
 363 Whether this bias was introduced by differences in template melting
 364 behaviour, primer annealing to target sequences or strand elongation during
 365 the initial PCR step, by processes during the NGS reaction or was caused by
 366 differences inherent to the nature of the muscular structure of the different
 367 species (e.g. absolute number of mitochondrial genomes per cell) remains to
 368 be clarified.

369 Further tests with prepared tuna mixtures of all occurring “tuna” species and
 370 additional possible substitute species could help to define normalisation
 371 strategies for the quantification of species in tuna tins. A similar approach
 372 has also been suggested for other metabarcoding studies (Diaz-Real et al.,
 373 2015; Pawluczyk, Weiss, Links, Aranguren, Wilkinson, & Egea-Cortines,
 374 2015). An alternative approach to evade bias based on PCR amplification is
 375 the untargeted high-throughput sequencing of total genomic DNA, which

376 has recently been shown to be a very promising tool for foodstuff analysis
377 (Ripp et al., 2014), albeit requiring a considerable amount of additional
378 bioinformatics knowledge and efforts.

379

380

381 3.3 NGS of tuna tins

382 As an initial test of market samples, three purchased tuna tins were
383 subjected to the same experimental approach as the prepared tuna samples.
384 These samples had been selected based on results of a previous market
385 survey in which authentication of different seafood products was performed
386 by means of conventional Sanger sequencing (Kappel & Schröder, 2015a).
387 For the present study, two tins from the same product (“light tuna in
388 vegetable oil”) and batch were selected, because one of these showed
389 ambiguous Sanger sequencing results with five unreadable bases, even after
390 repeated sequencing reactions. The sequences recovered from the other tin
391 seemed to be clear, apart from one ambiguous position in the BMID
392 fragment, but species assignment based on phylogenetic grouping was
393 inconsistent for BDR (yellowfin tuna) and BMID (albacore or Pacific
394 bluefin tuna). The third sample was chosen because of clearly unambiguous
395 sequences (“tuna fillets, white, in olive oil”) and was considered to be
396 prepared from albacore only.

397 All tins were analysed in duplicates but only one was submitted to two NGS
398 reactions (see Table 1), whereas the other tins were only analysed once (see
399 supplementary material for cluster details). As the exact contents of the tins
400 were considered to be unknown (apart from indications from the Sanger
401 sequencing results), recovered sequences were phylogenetically grouped
402 with sequences from scombroid reference species. Species assignments
403 were performed based on grouping of recovered sequences in clusters with
404 reference sequences belonging to particular species (see Fig. 2 for neighbour

405 joining trees). With one exception, all sequences from the NGS reaction
406 clustered with sequences belonging to tuna species (*Thunnus* species or *K.*
407 *pelamis*). However, one BDR signal was identified in one sample and one
408 run, which did not cluster with any of the scombroid reference sequences. A
409 GenBank BLAST search resulted in a 100% match with Atlantic salmon
410 (*Salmo salar*, accession number LC012541.1). The total read percentages of
411 sequences assigned to particular species are shown in Fig. 3. Interestingly,
412 although the first two samples represent two tins of the same product and
413 batch, the picture of NGS signals is quite different. In tin 1 the highest
414 signals represent longtail tuna and yellowfin tuna *cytb* sequences, and
415 additional signals were identified for Pacific bluefin tuna, skipjack tuna,
416 albacore and bigeye tuna plus one very low signal for salmon (see above). In
417 contrast, only sequences from four *Thunnus* species were recovered from tin
418 2 (yellowfin tuna, Pacific bluefin tuna, albacore and bigeye tuna), with
419 bigeye tuna sequences being only recovered with BMID but not BDR. NGS
420 of the third tuna tin gave only sequences which clustered with albacore
421 reference sequences. This was consistent with the labelling of the tin, which
422 explicitly stated the scientific name “*Thunnus alalunga*”.

423 The occurrence of different nucleotide sequences in tin 1 was expected
424 because of previous Sanger sequencing results which showed overlaying
425 chromatogram peaks. It is remarkable that different sequences were also
426 recovered for BDR in tin 2, although no ambiguous sequence positions were
427 found with Sanger sequencing in the preliminary tests. However, the
428 sequences from tin 2 represent exactly those species which have been
429 identified with Sanger sequencing, where BDR and BMID gave inconsistent
430 results (see above). Even slightly different sequences (albeit clustering only
431 with albacore sequences) were found in the third tin, from which clear
432 sequences for both marker fragments had been produced with the
433 conventional sequencing approach.

Standard deviations for the second and third tuna tin samples were comparably low like those of the prepared tuna mixtures, but were larger for the first sample. The reasons for these high standard deviations in tin 1 are not clear. Although tin 1 was analysed in two runs in contrast to the other tuna tins, this might not be the true reason, as the standard deviations for the tuna mixtures were quite low despite being calculated from four single values (duplicate samples plus two NGS runs). Whereas clear signals were produced for longtail tuna, yellowfin tuna and Pacific bluefin tuna for both *cytb* markers in both of the duplicate samples and both runs, the signals for albacore, bigeye tuna and salmon were identified only in some samples or runs and might be considered as traces derived from the production process. The very high standard deviation of skipjack BMID sequence, on the other hand, is caused by one very high signal in one of the duplicate samples and one run, and might rather reflect an artefact. Nevertheless, it is obvious that in the tins of the product “light tuna in vegetable oil” different species were mixed, which does not comply with EU law.

The presented NGS approach followed by phylogenetic grouping of the recovered sequences with sequences from reference specimens is especially suited for authenticity testing of tinned tuna products because for products with still intact muscular structure the European Regulation (No 1536/92) clearly states that “different species may not be mixed in the same container”. Despite this, results from the present study indicate the occurrence of tissues belonging to different species in one tin, an incident which can be easily overlooked with conventional Sanger sequencing. However, further research has to be undertaken with artificially produced tuna tins to define thresholds for individual species and different species’ compositions in order to decide whether signals are true admixtures of species or are only traces which can be neglected.

Another issue important to note is that a species assignment to Pacific bluefin tuna, Atlantic bluefin tuna or albacore based on mitochondrial DNA

464 alone is not absolutely sound due to potential introgression between Atlantic
465 bluefin tuna and Pacific bluefin tuna, as well as between albacore and
466 Atlantic bluefin tuna. Introgressed mitochondrial DNA has been described
467 to occur in about 2-3% of the described individuals. To address this
468 problem, additional nuclear markers, like rDNA first internal transcribed
469 spacer DNA (*its1*) (Viñas & Tudela, 2009), should be targeted in the
470 described NGS approach. A recent publication describes the development of
471 *its1* minibarcode primers for tuna tin authentication (Mitchell & Hellberg,
472 2016). This *its1* minibarcode sequence allowed the differentiation of
473 introgressed Atlantic bluefin tuna and albacore, but unfortunately the
474 variability of this short *its1* sequence was not sufficient for differentiation of
475 introgressed Atlantic bluefin tuna and Pacific bluefin tuna.

476 As efforts and costs of NGS applications are still rather high, this technique
477 is not yet applied in official investigation offices for routine analyses of food
478 surveillance. But costs for NGS technologies are constantly dropping
479 (Meier, Wong, Srivathsan, & Foo, 2015; Sboner, Mu, Greenbaum,
480 Auerbach, & Gerstein, 2011) and a high capacity for sample multiplexing
481 further helps in reducing costs (Tillmar, Dell'Amico, Welanders, &
482 Holmlund, 2013). Private suppliers already provide NGS-based species
483 authentication of mixed samples at costs comparable to those of
484 conventional authentication based on Sanger sequencing. In addition to that
485 novel third generation technologies, e.g. hand-held devices, are being
486 developed (van Dijk, Auger, Jaszczyszyn, & Thermes, 2014), which would
487 be extremely helpful for food inspectors since they could be applied on-site.
488 Consequently, this approach will likely be appropriate for official food
489 surveillance in the near future and can easily be applied for products
490 containing other fish or seafood species.

491

492   **4 Conclusions**

493   This study showed that species authentication of tuna samples containing
494   species mixtures is feasible with an Illumina MiSeq NGS approach targeting
495   two short fragments of the mitochondrial *cytb* gene. Admixtures of as low as
496   one percent tissue can be identified. However, the limit of detection strongly
497   depends on the species' composition. Due to amplification biases,
498   quantification of species in mixtures will not likely be possible with
499   considerable accuracy. Nevertheless, determination of high and low contents
500   is certainly feasible, when using knowledge about sequence representation
501   of particular species derived from prior experiments, like this study. NGS of
502   barcode markers followed by phylogenetic grouping with reference
503   sequences is a promising tool for species determination in tinned tuna
504   products, in which species mixtures might be present although this is
505   forbidden by European regulation. Since costs for NGS devices and services
506   are constantly dropping, the presented approach will likely be appropriate
507   for the official food surveillance in the near future and can easily be adapted
508   to products containing other fish or seafood species.

509 **Tables**

510 **Table 1**

Sample number	Sample number of duplicate	Species components in prepared mixture (in % weight/weight)/Tuna tin product	Analysed in NGS run
<i>Prepared tuna mixtures</i>			
01	14	<i>T. alalunga</i> (100)	1+2
02	17	<i>T. alalunga/K. pelamis</i> (50:50)	1+2
03	11	<i>T. alalunga/K. pelamis</i> (90:10)	1+2
04	13	<i>T. alalunga/K. pelamis</i> (99:1)	1+2
05	20	<i>T. alalunga/T. albacares</i> (50:50)	1+2
06	19	<i>T. alalunga/T. albacares/K. pelamis</i> (50:40:10)	1+2
07	12	<i>T. alalunga/T. albacares/T. obesus</i> (50:40:10)	1+2
08	15	<i>T. alalunga/T. albacares/T. obesus/K. pelamis</i> (25:25:25:25)	1+2
09	18	<i>K. pelamis/S. sarda</i> (90:10)	1+2
<i>Tuna can products</i>			
10	16	Light tuna in vegetable oil (tin 1)	1+2
21	24	Light tuna in vegetable oil (tin 2) ^a	2
22	25	Tuna fillets, white, in olive oil (<i>Thunnus alalunga</i>)	2

^a Same product and lot as sample n° 10, but different tin

511

512 Prepared mixed tuna samples and tuna tin products analysed by NGS.

513 Proportions of species in mixtures are given in brackets.

514   **Conflict of interest**

515   The authors declare no conflict of interest.

516

517   **Acknowledgements**

518   We kindly thank Roswitha Koch for sample preparation. We also thank

519   Lukas Windhager for proof reading of the article and the anonymous

520   reviewers for their helpful comments.

521   This work was funded by the European Union INTERREG Atlantic Area

522   Program (“LabelFish”, project 2011-1/163).

523

524   **Appendix: supplementary material**

525   Excel file with 1) BDR and BMID sequences from tuna specimens used for

526   preparation of mixed samples as well as nucleotide identity matrices and

527   sequence alignments, 2) NGS cluster details for prepared mixed tuna

528   samples, and 3) NGS cluster details for tuna tins

529

530   **References**

- 531   Bertolini, F., Ghionda, M. C., D'Alessandro, E., Geraci, C., Chiofalo, V., &
532   Fontanesi, L. (2015). A next generation semiconductor based
533   sequencing approach for the identification of meat species in DNA
534   mixtures. *Plos One*, 10(4), e0121701.
- 535   Carew, M. E., Pettigrove, V. J., Metzeling, L., & Hoffmann, A. A. (2013).
536   Environmental monitoring using next generation sequencing: rapid
537   identification of macroinvertebrate bioindicator species. *Frontiers*
538   *in Zoology*, 10(45).
- 539   Carvalho, D. C., Palhares, R. M., Drummond, M. G., & Frigo, T. B. (2015).
540   DNA Barcoding identification of commercialized seafood in South
541   Brazil: A governmental regulatory forensic program. *Food*
542   *Control*, 50(0), 784-788.
- 543   Cheng, X., Su, X., Chen, X., Zhao, H., Bo, C., Xu, J., Bai, H., & Ning, K.
544   (2014). Biological ingredient analysis of traditional Chinese
545   medicine preparation based on high-throughput sequencing: the
546   story for Liuwei Dihuang Wan. *Scientific Reports*, 4(5147).
- 547   Coghlan, M. L., Haile, J., Houston, J., Murray, D. C., White, N. E.,
548   Moolhuijzen, P., Bellgard, M. I., & Bunce, M. (2012). Deep
549   sequencing of plant and animal DNA contained within traditional
550   Chinese medicines reveals legality issues and health safety
551   concerns. *PLoS Genet*, 8(4), e1002657.
- 552   De Battisti, C., Marciano, S., Magnabosco, C., Busato, S., Arcangeli, G., &
553   Cattoli, G. (2014). Pyrosequencing as a Tool for Rapid Fish
554   Species Identification and Commercial Fraud Detection. *Journal of*
555   *Agricultural and Food Chemistry*, 62(1), 198-205.
- 556   Diaz-Real, J., Serrano, D., Píriz, A., & Jovani, R. (2015). NGS
557   metabarcoding proves successful for quantitative assessment of
558   symbiont abundance: the case of feather mites on birds.
559   *Experimental and Applied Acarology*, 67(2), 209-218.

560 Downs, T. R., & Wilfinger, W. W. (1983). Fluorometric quantification of
561 DNA in cells and tissue. *Anal Biochem*, 131(2), 538-547.

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

566 Galan, M., Pagès, M., & Cosson, J.-F. (2012). Next-generation sequencing
567 for rodent barcoding: species identification from fresh, degraded
568 and environmental samples. *Plos One*, 7(11), e48374.

569 González Sotelo, C. G., Medina Méndez, I., Quintero Vázquez, J., Pérez
570 Martín, R. I., & Rey-Méndez, M. (2002). Procedimiento para la
571 identificación de Albacora (*Thunnus alalunga*) en conservas de
572 atún blanco, Albacora o Bonito del Norte. Publication Number
573 2161137.

574 Griffiths, A. M., Sotelo, C. G., Mendes, R., Pérez-Martín, R. I., Schröder,
575 U., Shorten, M., Silva, H. A., Verrez-Bagnis, V., & Mariani, S.
576 (2014). Current methods for seafood authenticity testing in Europe:
577 Is there a need for harmonisation? *Food Control*, 45, 95-100.

578 Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain
579 reactions (PCR). *Journal of Bioscience and Bioengineering*, 96(4),
580 317-323.

581 Kappel, K., & Schröder, U. (2015a). Species identification of fishery
582 products in Germany. *Journal für Verbraucherschutz und*
583 *Lebensmittelsicherheit*, 10(Suppl 1), S31-S34.

584 Kappel, K., & Schröder, U. (2015b). Substitution of high-priced fish with
585 low-priced species: Adulteration of common sole in German
586 restaurants. *Food Control*, 59, 478-486.

587 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P.
588 A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez,
589 R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007).

590   Clustal W and clustal X version 2.0. *Bioinformatics*, 23(21), 2947-
591   2948.

592   Leadbitter, D., & Benguerel, R. (2014). Sustainable Tuna – Can the
593   Marketplace Improve Fishery Management? *Business Strategy and*
594   *the Environment*, 23(6), 417-432.

595   Luo, C., Tsementzi, D., Kyrpides, N., Read, T., & Konstantinidis, K. T.
596   (2012). Direct comparisons of Illumina vs. Roche 454 sequencing
597   technologies on the same microbial community DNA sample. *Plos*
598   *One*, 7(2), e30087.

599   Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu.*
600   *Rev. Genomics Hum. Genet.*, 9, 387-402.

601   Mariani, S., Griffiths, A. M., Velasco, A., Kappel, K., Jérôme, M., Pérez-
602   Martín, R. I., Schröder, U., Verrez-Bagnis, V., Silva, H., &
603   Vandamme, S. G. (2015). Low mislabeling rates indicate marked
604   improvements in European seafood market operations. *Frontiers in*
605   *Ecology and the Environment*, 13(10), 536-540.

606   Meier, R., Wong, W., Srivathsan, A., & Foo, M. (2015). \$1 DNA barcodes
607   for reconstructing complex phenomes and finding rare species in
608   specimen-rich samples. *Cladistics*, 32, 100-110.

609   Meyer, A. (1993). Evolution of mitochondrial DNA in fishes. *Biochemistry*
610   *and Molecular Biology of Fishes*, 2, 1-38.

611   Mitchell, J. K., & Hellberg, R. S. (2016). Use of the Mitochondrial Control
612   Region as a Potential DNA Mini-Barcoding Target for the
613   Identification of Canned Tuna Species. *Food Analytical Methods*,
614   doi: 10.1007/s12161-016-0460-3.

615   Pardo, M. Á., Jiménez, E., & Pérez-Villarreal, B. (2015). Misdescription
616   incidents in seafood sector. *Food Control*, 62, 277-283.

617   Park, J. Y., Lee, S. Y., An, C. M., Kang, J. H., Kim, J. H., Chai, J. C., Chen,
618   J., Kang, J. S., Ahn, J. J., Lee, Y. S., & Hwang, S. Y. (2012).
619   Comparative study between Next Generation Sequencing

620   Technique and identification of microarray for Species

621   Identification within blended food products. *Biochip Journal*, 6(4),

622   354-361.

623   Pawluczyk, M., Weiss, J., Links, M. G., Aranguren, M. E., Wilkinson, M.

624   D., & Egea-Cortines, M. (2015). Quantitative evaluation of bias in

625   PCR amplification and next-generation sequencing derived from

626   metabarcoding samples. *Analytical and Bioanalytical Chemistry*,

627   407(7), 1841-1848.

628   Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman,

629   S. N., & Taberlet, P. (2012). Who is eating what: diet assessment

630   using next generation sequencing. *Molecular ecology*, 21(8), 1931-

631   1950.

632   Quinteiro, J., Sotelo, C., Rehbein, H., Pryde, S. E., Medina, I., Pérez-Martín,

633   R., Rey-Méndez, M., & Mackie, I. (1998). Use of mtDNA direct

634   polymerase chain reaction (PCR) sequencing and PCR-restriction

635   fragment length polymorphism methodologies in species

636   identification of canned tuna. *Journal of Agricultural and Food*

637   *Chemistry*, 46(4), 1662-1669.

638   Rehbein, H. (2005). Identification of the fish species of raw or cold-smoked

639   salmon and salmon caviar by single-strand conformation

640   polymorphism (SSCP) analysis. *European Food Research and*

641   *Technology*, 220(5-6), 625-632.

642   Restrepo, V., Juan-Jordá, M. J., Collette, B. B., Frédou, F. L., & Rosenberg,

643   A. (2016). Tunas and Billfishes.

644   [http://www.un.org/Depts/los/global_reporting/WOA_RPROC/Cha](http://www.un.org/Depts/los/global_reporting/WOA_RPROC/Chapter_41.pdf)

645   [pter_41.pdf](http://www.un.org/Depts/los/global_reporting/WOA_RPROC/Chapter_41.pdf), last accessed: 23 February 2016.

646   Ripp, F., Krombholz, C. F., Liu, Y., Weber, M., Schäfer, A., Schmidt, B.,

647   Köppel, R., & Hankeln, T. (2014). All-Food-Seq (AFS): a

648   quantifiable screen for species in biological samples by deep DNA

649   sequencing. *Bmc Genomics*, 15(639).

650 Sboner, A., Mu, X. J., Greenbaum, D., Auerbach, R. K., & Gerstein, M. B.
 651 (2011). The real cost of sequencing: higher than you think!
 652 *Genome Biology*, 12(8).
 653 Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., & Hajibabaei, M.
 654 (2015). A DNA Mini-Barcoding System for Authentication of
 655 Processed Fish Products. *Scientific Reports*, 5(15894).
 656 Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H.,
 657 Hallwachs, W., Golding, G. B., & Hajibabaei, M. (2015).
 658 Massively parallel multiplex DNA sequencing for specimen
 659 identification using an Illumina MiSeq platform. *Scientific Reports*,
 660 5(9687).
 661 Staley, C., Unno, T., Gould, T. J., Jarvis, B., Phillips, J., Cotner, J. B., &
 662 Sadowsky, M. J. (2013). Application of Illumina next-generation
 663 sequencing to characterize the bacterial community of the Upper
 664 Mississippi River. *Journal of Applied Microbiology*, 115(5), 1147-
 665 1158.
 666 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S.
 667 (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using
 668 Maximum Likelihood, Evolutionary Distance, and Maximum
 669 Parsimony Methods. *Molecular Biology and Evolution*, 28(10),
 670 2731-2739.
 671 Tillmar, A. O., Dell'Amico, B., Welander, J., & Holmlund, G. (2013). A
 672 universal method for species identification of mammals utilizing
 673 next generation sequencing for the analysis of DNA mixtures. *Plos*
 674 *One*, 8(12), e83761.
 675 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.
 676 F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud,
 677 C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont,
 678 D., Argillier, C., Baudoin, J.-M., Peroux, T., Crivelli, A. J., Olivier,
 679 A., Acqueberge, M., Le Brun, M., Møller, P. R., Willerslev, E., &

680 Dejean, T. (2016). Next-generation monitoring of aquatic
 681 biodiversity using environmental DNA metabarcoding. *Molecular*
 682 *ecology*, 25, 929-942.
 683 van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten
 684 years of next-generation sequencing technology. *Trends in*
 685 *genetics*, 30(9), 418-426.
 686 Viñas, J., & Tudela, S. (2009). A Validated Methodology for Genetic
 687 Identification of Tuna Species (Genus Thunnus). *Plos One*, 4(10),
 688 e7606.
 689

690   **Figure Captions**

691

692   **Fig. 1.** Recovered BDR and BMID sequences from NGS analysis of
 693   prepared tuna samples. Bars represent mean values of percentages of reads
 694   in clusters of duplicate samples, analysed in two runs each (resulting
 695   altogether in four values used for calculation of mean values and standard
 696   deviations). NGS cluster sequences were compared to sequences of applied
 697   specimens and were shown to be 100% identical unless stated otherwise. ^{*)}
 698   BMID sequence with one nucleotide difference to sequence of applied *K.*
 699   *pelamis* specimen. ^{a)} Percentage of reads in false positive *K. pelamis* BMID
 700   cluster was above the threshold of 1% only in sample 17, run 2. ^{b)}
 701   Percentage of reads in false positive *K. pelamis* BMID cluster was above the
 702   threshold only in sample 5, run 2. ^{c)} Percentage of reads in false positive *K.*
 703   *pelamis* BMID cluster was above the threshold only in sample 9, run 2. ^{d)}
 704   False negative result of *S. sarda* BMID sequence in both runs of both
 705   samples.

706

707   **Fig. 2.** Phylogenetic grouping of NGS cluster sequences of tuna tin analysis.
 708   Sequences from NGS clusters representing more than 1% of reads were
 709   used to construct a BDR (A) and a BMID (B) alignment respectively,
 710   together with sequences from scombroid reference specimens. Kimura 2-
 711   parameter distances were calculated and neighbour-joining trees were
 712   produced with Mega5.2. For better illustrations monophyletic groups were
 713   collapsed when appropriate. Samples are depicted in bold characters, and for
 714   sample 10 and 16 also numbers of NGS runs are given (sample 10.1 =
 715   sample 10 run 1). If more than one recovered sequence belonging to a
 716   particular sample was present in a clade, the number of different sequences
 717   is given in brackets. ^{*)} Recovered sequence displayed 100% identity to
 718   partial *cytb* GenBank sequence of Atlantic salmon (LC012541.1). ^{**)}

719 Sequences in cluster have highest similarity to *T. albacares* but are not
720 100% identical to sequences from *T. albacares* reference specimens nor do
721 they form a monophyletic group with these sequences. Nevertheless, these
722 sequences were assigned to *T. albacares* for calculation of sum of read
723 percentages for Fig. 3.

724

725 **Fig. 3.** Summary of species assignments to NGS cluster sequences of tuna
726 tin analysis. Bars represent totalised read percentages of all clusters having
727 been identified to belong to particular species based on neighbour-joining
728 tree grouping. Mean values and standard deviations are calculated based on
729 two samples, analysed in two runs each (sample 10 and 16), or two samples,
730 analysed in one run each (sample 21 and 24; 22 and 25), respectively. ^{a)}
731 High standard deviation because of very high read percentage in BMID
732 cluster of sample 10 (run 2). ^{b)} Read percentage above threshold only for
733 BDR sequence of run 1 of sample 10. GenBank BLAST search indicates
734 Atlantic salmon (*Salmo salar*, LC012541.1).

735

Table 1

Sample number	Sample number of duplicate	Species components in prepared mixture (in % weight/weight)/tuna tin product	Analysed in NGS run
<i>Prepared tuna mixtures</i>			
01	14	<i>T. alalunga</i> (100)	1+2
02	17	<i>T. alalunga</i> / <i>K. pelamis</i> (50:50)	1+2
03	11	<i>T. alalunga</i> / <i>K. pelamis</i> (90:10)	1+2
04	13	<i>T. alalunga</i> / <i>K. pelamis</i> (99:1)	1+2
05	20	<i>T. alalunga</i> / <i>T. albacares</i> (50:50)	1+2
06	19	<i>T. alalunga</i> / <i>T. albacares</i> / <i>K. pelamis</i> (50:40:10)	1+2
07	12	<i>T. alalunga</i> / <i>T. albacares</i> / <i>T. obesus</i> (50:40:10)	1+2
08	15	<i>T. alalunga</i> / <i>T. albacares</i> / <i>T. obesus</i> / <i>K. pelamis</i> (25:25:25:25)	1+2
09	18	<i>K. pelamis</i> / <i>S. sarda</i> (90:10)	1+2
<i>Tuna can products</i>			
10	16	Light tuna in vegetable oil (tin 1)	1+2
21	24	Light tuna in vegetable oil (tin 2) ^a	2
22	25	Tuna fillets, white, in olive oil (<i>Thunnus alalunga</i>)	2

^a Same product and lot as sample n° 10, but different tin

Prepared mixed tuna samples and tuna can products analysed by NGS. Proportions of species in mixtures are given in brackets.

Figure 1
[Click here to download Figure\(s\): Fig 1_Kappel.docx](#)

Figure 2
[Click here to download Figure\(s\): Fig 2_Kappel.docx](#)

Figure 3
[Click here to download Figure\(s\): Fig 3_Kappel.docx](#)

Supplementary Material

[Click here to download Supplementary Material: Supplementary material_Kappel.xlsx](#)

Supplementary Material caption

[Click here to download Supplementary Material: Supplementary material caption_Kappel_170426.docx](#)