JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2002,** *50,* 963–969 **963**

Statistical Validation of the Identification of Tuna Species: Bootstrap Analysis of Mitochondrial DNA Sequences

JAVIER TEROL,*,† ROSARIO MASCARELL,‡ VICTORIA FERNANDEZ-PEDROSA,§ AND MANUEL PÉREZ-ALONSO‡

Departamento de Genética, Universidad de Valencia, Dr. Moliner 50, Burjasot E46100, Spain, and Sistemas Genómicos SL, C/Benjamin Franklin 12, CEEI, Parque Tecnológico de Valenicia, Paterna E46980, Spain

Sequencing of the mitochondrial cytochrome *b* gene has been used to differentiate three tuna species: *Thunnus albacares* (yellowfin tuna), *Thunnus obesus* (bigeye tuna), and *Katsuwonus pelamis* (skipjack). A PCR amplified 528 bp fragment from 30 frozen samples and a 171 bp fragment from 26 canned samples of the three species were analyzed to determine the intraspecific variation and the positions with diagnostic value. Polymorphic sites between the species that did not present intraspecific variation were given a diagnostic value. The genetic distance between the sequences was calculated, and a phylogenetic tree was constructed, showing that the sequences belonging to the same species clustered together. The bootstrap test of confidence was used to determine the statistical validation of the species assignation, allowing for the first time a quantification of the certainty of the species assignation. The bootstrap values obtained from these results indicate that the sequencing of the cytochrome *b* fragments allows a correct species assignation with a probability ≥95%.

KEYWORDS: Canned tuna; species identification; yellowfin tuna; bigeye tuna; skipjack; cytochrome *b*; PCR amplification; DNA sequencing; genetic distances; bootstrap test

INTRODUCTION

Tuna species are of a great commercial importance, and their economic value in the market depends on which species has been used in the elaboration of the canned fish. According to EU Regulation 1536/92, the term *tuna* includes species belonging to the genus *Thunnus* (*T. alalunga*, *T. albacore*, *T. thynnus*, *T. obesus*, and others) and the species *Katsuwonus pelamis*. Commercial labeling is used to differentiate the species used in the preparation of the food product; *white tuna* refers to *T. alalunga*; *light tuna* or yellowfin indicates *T. albacares*; and *tuna* can be *K. pelamis* or any other *Thunnus* species.

The labeling *light tuna* was a commercial denomination created by the Spanish producers that has been recently recognized by the Spanish legislation (RD 1193/2000). Sales of *light* tuna account for most of the canned tuna fish market in Spain. The use of cheaper and lower quality species to produce canned tuna labeled as *light* tuna has been detected, which constitutes a fraud that must be avoided. Therefore, the development of analysis tools to allow the enforcement of labeling regulations will benefit the canning industry, the food analysis laboratories, and, of course, the customers, who can be assured they are paying for the right product.

Although raw fish is easily identified with standard biochemical methods, the unequivocal classification of canned tuna presents important problems produced by the thermal treatment that is applied to the fish during the canning process. This thermal treatment involves at least one sterilization step that causes the denaturalization of the muscle proteins, which leads to subsequent loss of analytical specificity (*1*). DNA is a much more stable molecule and allows more precise classification of species, based on differences of the sequences of the genes used for the study. On the other hand, the Polymerase Chain Reaction (PCR) technique allows the amplification of DNA fragments from just a few original copies, making the analysis very sensitive.

Analyses of mitochondrial DNA have been shown to be very useful to resolve evolutionary relationships among closely related species of the *Thunnus* genus (*2*, *3*). These relationships had been unsuccessfully analyzed using traditional biochemical methods (*4*). Referring to species identification, the amplification and sequencing of a 307 bp segment of the mitochondrial cytochrome *b* gene was used to successfully differentiate four tuna species (*T. thynnus*, *T. obesus*, *T. albacares*, and *T. alalunga*) from raw specimens (*5*).

The identification of tuna species is more difficult when canned samples are analyzed. The thermal treatments applied to the muscle tissue during the elaboration of the canned tuna result in DNA degradation and fragmentation, which limit the amount of DNA obtained and the size of the fragment to be amplified, respectively.

Several works have used the sequencing of mitochondrial DNA for the identification of canned tuna species, but in all of them either the fragment analyzed was too short to allow the discrimination between very close species or the number of

* Author to whom correspondence should be addressed.
† Present address: Sistemas Genómicos SL.
‡ Universidad de Valencia.
§ Sistemas Genómicos SL.

10.1021/jf011032o CCC: $22.00
Published on Web 02/02/2002

specimens studied was too low to discard intraspecific variation that could cause the misidentification of a sample (*6*−*9*).

The aim of this work is to develop an analysis based on the sequencing of mitochondrial DNA to unambiguously differentiate *T. albacares* from other commercial species usually found in the market, especially *T. obesus* and *K. pelamis*. To achieve this, we have analyzed a number of frozen samples of these species to detect the intraspecific variation as well as the diagnostic positions in a 528 bp fragment of the mitochondrial cytochrome *b* gene amplified by PCR. We have also implemented, for the first time in this kind of work, the bootstrap test, which allows a quantitative estimate of the degree of confidence of the species assignation.

MATERIALS AND METHODS

Samples. Thirty frozen samples of tuna fish, corresponding to yellowfin tuna (*Thunnus albacares*), skipjack (*Katsuwonus pelamis*), and bigeye tuna (*Thunnus obesus*), were obtained from a national producer of canned tuna fish. The specimens used in this study were 7 yellowfin tunas, 13 bigeye ones, and 10 skipjacks.

We also analyzed 26 samples of canned tuna fish, corresponding to the 3 above-mentioned species, that had been classified by the manufacturer as follows: 9 yellowfin tuna, 5 skipjacks, and 12 bigeye tunas.

DNA Extraction. DNA extraction from frozen samples was carried out as in Kocher et al. (*10*), with some modifications. The muscle tissue samples were homogenized in a Stomacher Lab-Blender, and 200 mg was used for the DNA extraction. The homogenates were incubated for 4 h in 100 mM Tris-HCl, pH 8.0, 30 mM ethylendiamintetraacetic acid (EDTA), 100 mM NaCl, 5% sodium docecyl sulfate (SDS), 50 mM dithiothreitol (DTT), and 300 μg/mL proteinase K (TENSD), at 65 °C, and centrifuged for 5 min. DNA was isolated from the supernatant with three extractions with phenol/chloroform/isoamyl alcohol (24:24:1), precipitated with 2-propanol at room temperature, and resuspended in 100 μL of H_2O.

For the canned tuna, oil and lipids were removed by hatching the muscle tissue overnight in chloroform/methanol/water (1:2:0.8). The defatted tissue was recovered and employed for DNA extraction as described above.

Mitochondrial Cytochrome *b* Gene Fragment Amplification. The sequences for the mitochondrial cytochrome *b* gene were obtained from GenBank for the three species of interest (L11557, L11559, and L11539) and aligned with the Clustal X program. Primers were designed to bind the three species cytochrome *b* genes, to ensure amplification.

These primers are

AD: 5′CCT TAC AGG ACT ATT CCT CGC3′

AR: 5′CCG ATG TTT CAT GTT TCT TTG3′

BD: 5′GGC CGA GGC CTT TAC TAC GGC3′

BR: 5′GAA GAA TCG GGT GAG GGT GGC3′

CD: 5′GTT GAA TGA ATC TGA GGA GGC3′

CR: 5′GAT CAC GAA ACC AAG GAG G3′

PCR reactions with the primer combination AD/CR amplify a 528 bp fragment of the cytochrome *b* gene, and the primers AR, BD, BR, and CD were used to sequence the amplified fragments, to ensure a double-strand reading for each sample. If necessary, they could be also used to PCR amplify three overlapping fragments that constitute the original 528 bp sequence. These primers were designed for the amplification of DNA from canned samples, where the size of the fragment is critical to obtain PCR products.

PCR Reactions. The amplifications were performed in 1× reaction buffer without $MgCl_2$, with 0.2 mM dNTP, 0.5 μM each primer, 1 unit of DNA polymerase (Biotools 10.012), and between 0.2 and 1 μg of DNA. The final volume of the reactions was 50 μL, and they were carried out in a Biometra UNO II system with 40 cycles (94 °C, -30 s; 50 °C, 30 s; and 72 °C, 1 min).

DNA Sequencing. The sequencing was performed directly on the amplified fragments at the laboratories of Sistemas Genómicos S.L. DNA sequence was determined with an ABI-377 automatic sequencer following the supplier's instructions.

Sequence Analysis. Multiple alignments were performed with the Clustal X program (*11*). Genetic distances were calculated by employing the Tamura and Nei model for protein coding DNA sequences (*12*); the phylogenetic trees were constructed with the UPGMA (*13*) and neighbor-joining (*14*) methods, and the bootstrap test was carried out with 1000 iterations. These analyses were performed with the MEGA platform (*15*).

Database Searches. The GenBank (GB) and European Molecular Biology (EMB) databases were searched for mitochondrial cytochrome *b* sequences from other related tuna species, using the Basic Local Alignment Search Tool (BLAST) program at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) (*16*).

RESULTS

Analysis of Frozen Samples. Mitochondrial DNA was obtained from the frozen samples, and a 528 bp fragment of the cytochrome *b* gene was PCR amplified as described under Materials and Methods. We obtained amplification from 7 *T. albacares* samples, 13 *T. obesus* ones, and 10 *K. pelamis* individuals. The amplified DNA was sequenced at the laboratories of Sistemas Genómicos, and the resulting sequences were further analyzed.

The 30 cytochrome *b* fragments were aligned with the Clustal X program (*11*) to the sequences obtained from the GenBank database, corresponding to *Thunnus thynnus thynnus* (no. L11560), *Thunnus maccoyii* (no. L11558), *Thunnus atlanticus* (no. AF239963), *Thunnus tonggol* (no. AF239964), and *Thunnus alalunga* (no. L11556). *Sarda sarda* (no. L11549) was used as an out group. The results are shown in **Figure 1**, where only the variable positions are indicated.

The analysis of the alignment showed that 124 positions, of 528, were variable ones. The remaining 404 positions were constant in all of the species, including the ones from different genera, which accounts for the high degree of conservation of the cytochrome *b* gene. Low intraspecific variation was found in the different individuals of the three species analyzed, with 9 polymorphic variations in the yellowfin tunas, 9 in the bigeye ones, and 15 in the skipjacks. This low intraspecific variation allowed us to define some polymorphic sites that could be used as diagnostic positions to differentiate *T. albacares* from the other species. *T. obesus* presented 9 diagnostic positions (67, 124, 175, 202, 232, 313, 322, 433, and 463) and *K. pelamis* 41 differential nucleotides with respect to *T. albacares*, the last number reflecting the higher variability that is expected for species from different genera. The diagnostic positions were chosen because they did not show any intraspecific variation in any of the analyzed sequences. We also found diagnostic positions for other species close to *T. albacares*: 3 for *T. atlanticus*, 5 for *T. tonggol*, 10 for *T. maccoyii*, and 9 for *T. thynnus thynnus*. However, as only one sequence from each species was analyzed, the final number of diagnostic positions could be subject to variation if intraspecific variation were found when more individuals from these species are analyzed.

We estimated the genetic distances between the cytochrome *b* gene sequences for a more accurate species identification. We used the Tamura and Nei method (*12*), as this model allows the analysis of the differences in the nucleotide sequence considering different values for transitions and transversions and taking into account each base frequency to determine the importance of any base substitution. These estimates show how the intraspecific distances are very low, ranging from 0 to 0.006 in *T. albacares* and *T. obesus* and from 0.006 to 0.013 in *K.*

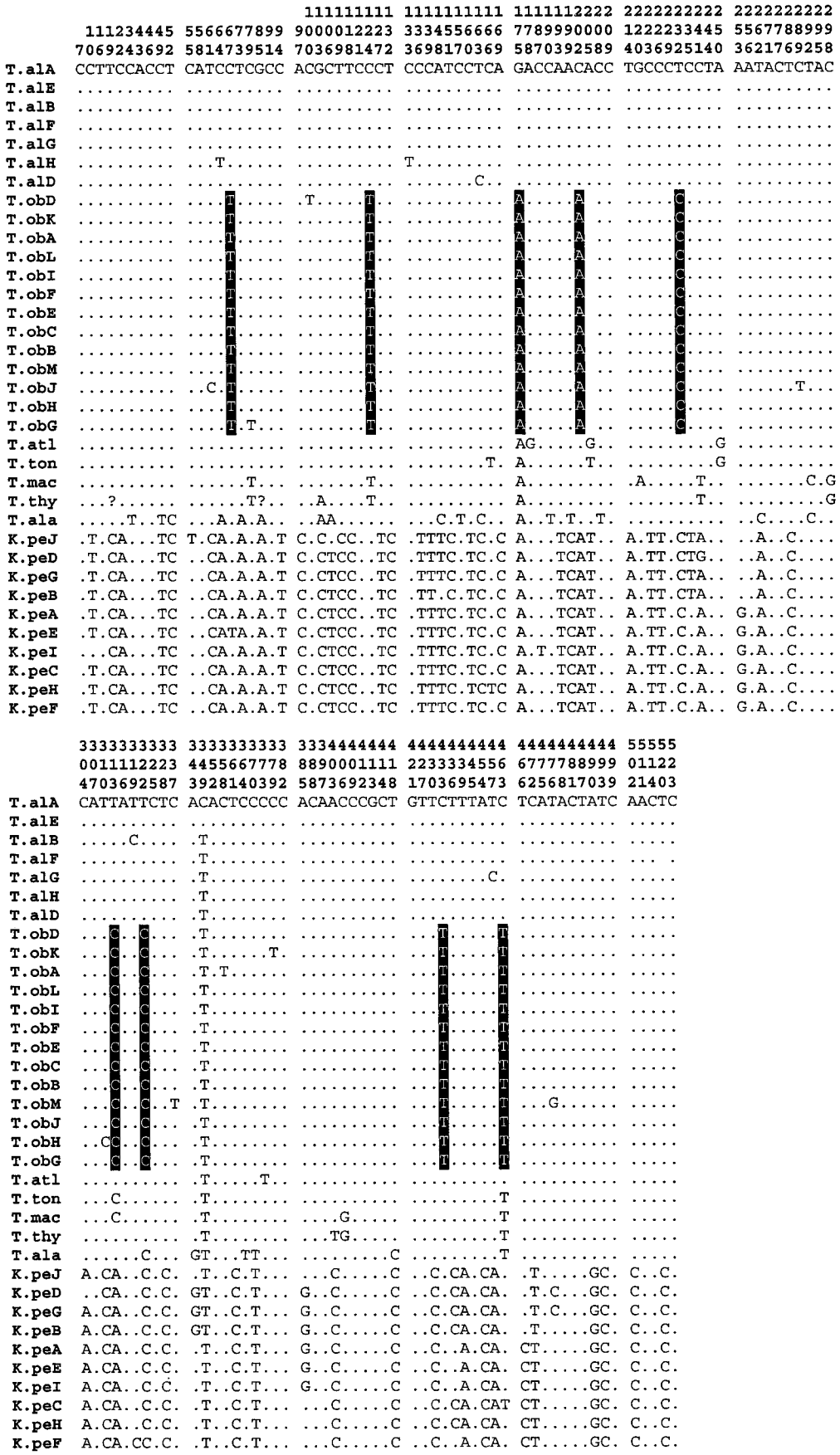

Figure 1. Multiple alignment of the 528 bp cytochrome *b* gene fragment from the analyzed frozen samples with the sequences obtained from the GenBank. The alignment was performed with the Clustal X program, and only variable sites are shown. The sequences are indicated as follows: *T. albacares*, *T.al*A–H; *T. obesus*, *T.ob*A–M; *K. pelamis*, *K.pe*A–J; *Thunnus thynnus thynnus*, *T.thy*; *Thunnus maccoyii*, *T.mac*; *Thunnus atlanticus*, *T.atl*; *Thunnus tonggol*, *T.ton*; *Thunnus alalunga*, *T.ala*. "?" indicates undetermined positions.

Table 1. Genetic Distances between the Sequences from the Frozen Samples Estimated by the Tamura and Nei Method[a]

	T.alA	T.alB	T.alD	T.alE	T.alF	T.alG	T.alH	T.obA	T.obB	T.obC	T.obD	T.obE	T.obF	T.obG	T.obH
T.alA[b]															
T.alB	.004														
T.alD	.004	.004													
T.alE	.000	.004	.004												
T.alF	.002	.002	.002	.002											
T.alG	.004	.004	.004	.004	.002										
T.alH	.006	.006	.006	.006	.004	.006									
T.obA	.021	.021	.021	.021	.019	.021	.023								
T.obB	.019	.019	.019	.019	.017	.019	.021	.002							
T.obC	.019	.019	.019	.019	.017	.019	.021	.002	.000						
T.obD	.021	.021	.021	.021	.019	.021	.023	.004	.002	.002					
T.obE	.019	.019	.019	.019	.017	.019	.021	.002	.000	.000	.002				
T.obF	.019	.019	.019	.019	.017	.019	.021	.002	.000	.000	.002	.000			
T.obG	.021	.021	.021	.021	.019	.021	.023	.004	.002	.002	.004	.002	.002		
T.obH	.021	.021	.021	.021	.019	.021	.023	.004	.002	.002	.004	.002	.002	.004	
T.obI	.019	.019	.019	.019	.017	.019	.021	.002	.000	.000	.002	.000	.000	.002	.002
T.obJ	.023	.023	.023	.023	.021	.023	.025	.006	.004	.004	.006	.004	.004	.006	.006
T.obK	.021	.021	.021	.021	.019	.021	.023	.004	.002	.002	.004	.002	.002	.004	.004
T.obL	.019	.019	.019	.019	.017	.019	.021	.002	.000	.000	.002	.000	.000	.002	.002
T.obM	.023	.023	.023	.023	.021	.023	.025	.006	.004	.004	.006	.004	.004	.006	.006
K.peA	.122	.122	.117	.122	.119	.119	.122	.123	.120	.120	.123	.120	.120	.123	.118
K.peB	.122	.122	.117	.122	.120	.120	.122	.123	.121	.121	.123	.121	.121	.123	.118
K.peC	.125	.125	.120	.125	.122	.122	.125	.121	.118	.118	.121	.118	.118	.121	.116
K.peD	.125	.125	.120	.125	.122	.122	.125	.126	.123	.123	.126	.123	.123	.126	.120
K.peE	.124	.124	.119	.124	.122	.122	.124	.120	.118	.118	.120	.118	.118	.120	.115
K.peF	.122	.117	.117	.122	.120	.120	.122	.123	.121	.121	.123	.121	.121	.123	.118
K.peG	.127	.127	.122	.127	.124	.124	.127	.128	.125	.125	.128	.125	.125	.128	.123
K.peH	.125	.125	.120	.125	.122	.122	.125	.126	.123	.123	.126	.123	.123	.126	.121
K.peI	.122	.122	.117	.122	.119	.119	.122	.123	.120	.120	.123	.120	.120	.123	.118
K.peJ	.120	.120	.115	.120	.118	.118	.120	.121	.119	.119	.121	.119	.119	.121	.116

	T.obI	T.obJ	T.obK	T.obL	T.obM	K.peA	K.peB	K.peC	K.peD	K.peE	K.peF	K.peG	K.peH	K.peI	K.peJ
T.obI															
T.obJ	.004														
T.obK	.002	.006													
T.obL	.000	.004	.002												
T.obM	.004	.008	.006	.004											
K.peA	.120	.120	.123	.120	.125										
K.peB	.121	.121	.123	.121	.125	.011									
K.peC	.118	.118	.121	.118	.123	.006	.013								
K.peD	.123	.123	.126	.123	.128	.015	.008	.017							
K.peE	.118	.118	.120	.118	.122	.002	.013	.008	.017						
K.peF	.121	.121	.123	.121	.125	.004	.015	.006	.019	.006					
K.peG	.125	.125	.128	.125	.130	.012	.004	.013	.004	.013	.015				
K.peH	.123	.123	.126	.123	.128	.006	.013	.004	.017	.008	.006	.013			
K.peI	.120	.120	.123	.120	.125	.004	.015	.010	.019	.006	.008	.015	.010		
K.peJ	.119	.119	.121	.119	.123	.013	.010	.012	.013	.015	.013	.010	.012	.017	

[a] Gray triangles indicate the distances between individuals from the same species. [b] Names are indicated as in **Figure 1**.

pelamis, which would show a higher variability (**Table 1**). On the other hand, the genetic distances measured between the species of our interest is 1 order of magnitude higher than the intraspecific ones, as can be seen when we estimate the mean distances between the species groups: 0.021 ± 0.006 between *T. albacares* and *T. obesus* and 0.122 ± 0.015 between *T. albacares* and *K. pelamis*. To determine if these genetic distances were significant, we estimated the coefficient of differentiation (N_{ST}), obtaining $N_{ST} = 0.908 \pm 0.017$ (*17*). The statistical *D* was determined as described by Lynch and Crease (*17*) ($D = 48.49$), and the χ^2 test was applied, confirming that, with a 99.995% degree of confidence, the mean distances between the species groups were significant.

We used the neighbor-joining method to construct a phylogenetic tree based on the genetic distances previously calculated and tested the reliability of the inferred tree using the bootstrap test (BP), with 2000 replicas (*18*). The tree shows that all of the sequences belonging to individuals of the same species cluster together and that those clusters are strongly supported by the high BP values obtained: 99% for *K. pelamis*, 95% for *T. obesus*, and 79% for *T. albacares*. It is remarkable that all of the sequences of the species of our interest, *T. albacares*, cluster together, separate with respect to its closest species, *T. atlanticus* and *T. tonggol*. These results are in agreement with the ones obtained by Chow and Kishimo (*2*) with the analysis of the ATPase gene sequences of 26 individuals of the *Thunnus* genus and 3 individuals of *K. pelamis*. Our results show that the sequencing of the 528 bp cytochrome *b* fragment allows the differentiation of *T. albacares* from all of the other tuna fish species. On the other hand, the bootstrap values obtained validate statistically the analysis and give a quantitative measurement of the certainty of the assignment of a sample to one species.

Analysis of the Canned Samples. We assessed the validity of our method on canned tuna fish, and we analyzed a total of 29 samples, 9 yellowfin tuna (CYF1−9), 14 bigeye tuna

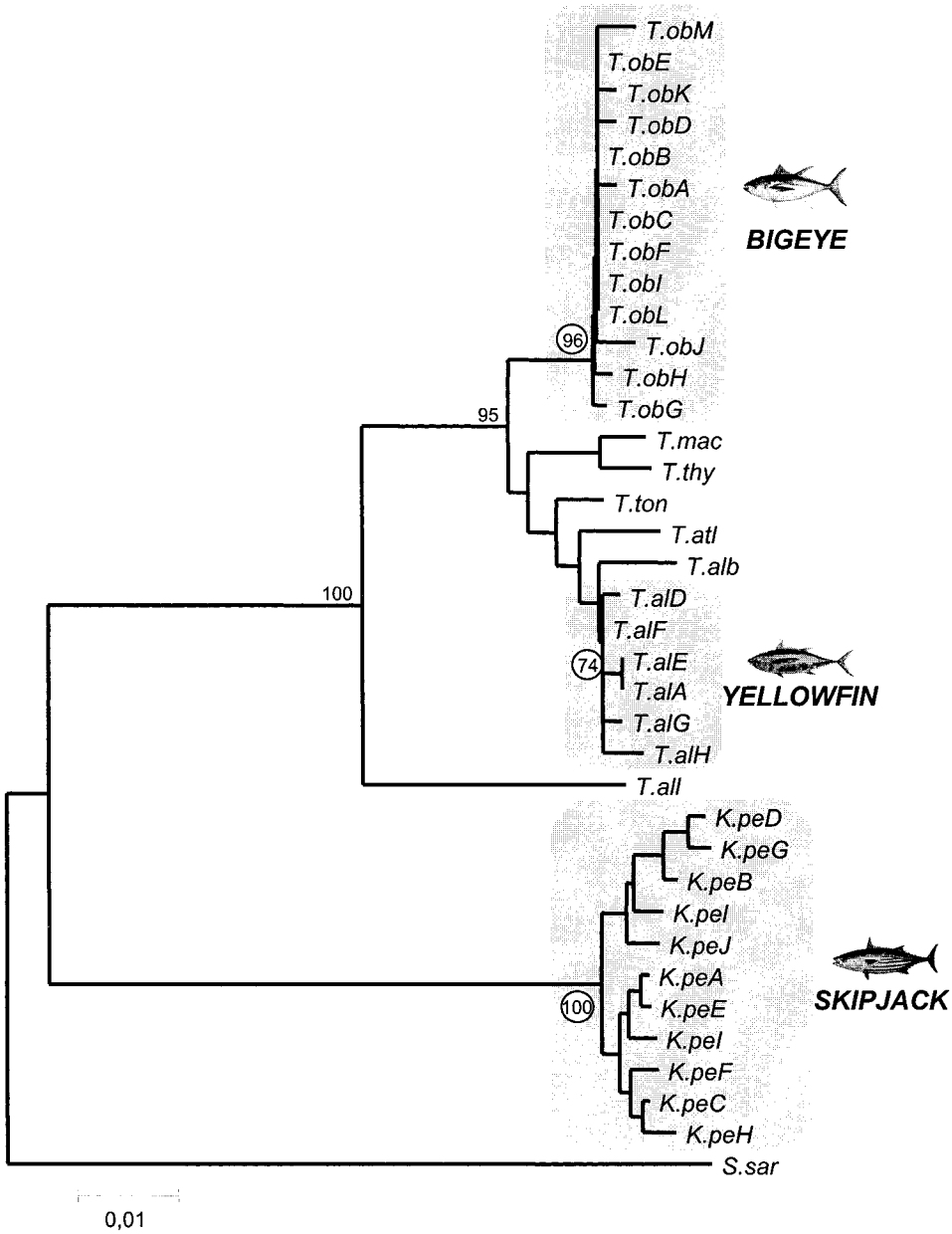

Figure 2. Phylogenetic tree for the sequences obtained from the frozen samples, constructed with the neighbor-joining method, based on the genetic distances that were calculated with the Tamura and Nei method from the multiple alignment in **Figure 1**. The clusters formed by individuals from the three species of interest are indicated with gray boxes; significant bootstrap values are encircled. *Sarda sarda* (*S.sar*) was used as an out group. The nomenclature is as in Figure 1.

(CBE1−14), and 5 skipjack (CSJ1−5). It had previously been described that, due to the thermal treatment applied to the fish meat, the maximum size of the DNA fragments obtained from canned samples was ~176 bp (*8*). We decided to use the primers AD/AR to obtain a 171 bp fragment of the cytochrome *b* gene, to guarantee that the amplification would take place. The samples were treated as described under Materials and Methods, and up to 1 μg of the DNA extracted from the canned tuna was used for PCR reactions. We obtained amplification of the 171 bp fragment in all of the samples analyzed, which is the biggest one obtained so far from canned tuna.

All of the fragments were double-strand sequenced, and the resulting sequences were aligned with the Clustal X program. No intraspecific variation could be observed, as all of the sequences belonging to the same species were identical (data not shown), so we decided to use one sequence per species in the next studies (CYF1, CBE1, and SJ1). We compared them to the ones obtained from the frozen samples representing all of the polymorphic variations found in the three species (**Figure 3A**) and calculated the genetic distances with the Tamura and Nei method (**Table 2**). As expected, the genetic distances estimated for CYF1, CBE1, and SJ1, with respect to the other sequences, corresponded to the ones expected for specimens from *T. albacares*, *T. obesus*, and *K. pelamis*, respectively.

A phylogenetic tree was constructed with the UPGMA method, showing that CYF1, CBE1, and SJ1 clustered clearly with the sequences from the same species and that the clustering was confirmed by the high BP values obtained (**Figure 3B**). These results show the validity of our method for the identification of yellowfin tuna from canned samples.

DISCUSSION

DNA sequencing has become a powerful tool in phylogenetic studies, as it allows the establishment of the phylogenetic

A

```
                      111111 1111111111 1
          1124556667 7899000012 2333345666 6
          0929251473 9547036984 7236981036 9
CYF1      CTCCTCTCCT CGCACGCTTC CTCCCATCTC A
T.alA     .......... .......... .......... .
T.alD     .......... .......... ........C. .
T.alH     .......T.. .......... ..T....... .
CBE1      ........T. .........T .......... .
T.obA     ........T. .........T .......... .
T.obD     ........T. ....T....T .......... .
T.obG     ........T. T........T .......... .
T.obJ     ......C.T. .........T .......... .
CSJ1      TCATC.CA.A .ATC.CTCC. TC.TTTCTC. C
K.peA     TCATC.CA.A .ATC.CTCC. TC.TTTCTC. C
K.peB     TCATC.CA.A .ATC.CTCC. TC.TT.CTC. C
K.peE     TCATC.CATA .ATC.CTCC. TC.TTTCTC. C
K.peH     TCATC.CA.A .ATC.CTCC. TC.TTTCTCT C
K.peI     .CATC.CA.A .ATC.CTCC. TC.TTTCTC. C
K.peJ     TCATCTCA.A .ATC.C.CC. TC.TTTCTC. C
```

B

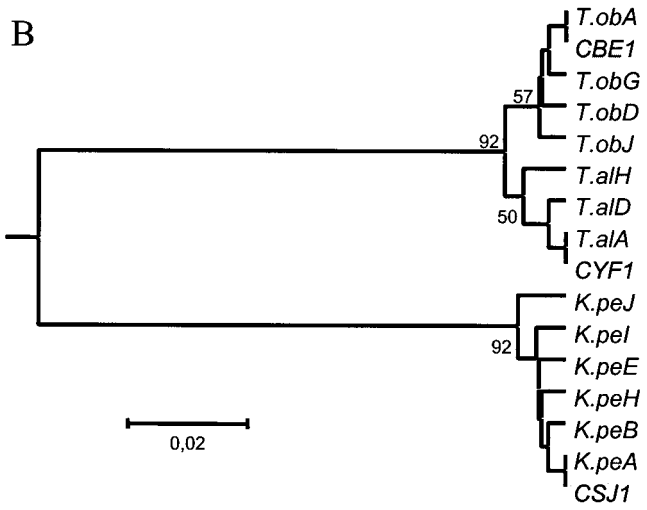

Figure 3. (A) Multiple alignment of the 271 bp fragment obtained for the canned samples (CYF1, CBE1, and CSJ1) and the frozen samples that presented polymorphic variations (the names are indicated as in **Figure 1**). (B) Phylogenetic tree constructed with the UPGMA method with the genetic distances calculated with the Tamura and Nei method from the alignment shown above. The sequences from the canned samples cluster separately with the sequences obtained from individuals of the same species. The bootstrap values (BP = 92) for the crucial branches reflect the reliability of the species assignation.

relationships between very close species. The sequence to be analyzed has to be conserved enough to permit amplification in different species with the same primers and, at the same time, has to present a degree of variability that allows differentiation of the species studied. The mitochondrial cytochrome *b* gene has been shown to be very useful for this purpose, as it shows considerable interspecies variation in nucleotide sequence, with a much lower intraspecific variation (*19*). This way, the analysis of different regions of the cytochrome *b* gene has been performed to investigate phylogenetic relationships among the species of genus as *Thunnus* (*2*), *Gammarus* (*20*), and *Leuscius* (*21*), subfamilies as *Cyprinidae* (*22*) and families as *Percidae* (*23*).

DNA sequencing has been used successfully in the identification of species of economic interest, and the analyses performed on the tuna species have been relevant. The first study on the genus *Thunnus* was performed by Barlett and Davidson (*5*), who analyzed a 307 bp fragment of the cytochrome *b* gene obtained from fresh samples. They found differences among the four

Table 2. Genetic Distance between the Canned Samples and the Control Sequences Estimated by the Tamura and Nei Method[a]

	CYF1	CBE1	CSJ1
CBE1	0,012		
CSJ1	0,165	0,184	
T.alA	0,000[a]	0,012	0,165
T.alD	0,006	0,018	0,157
T.alH	0,012	0,024	0,174
T.obA	0,012	0,000	0,184
T.obD	0,018	0,006	0,193
T.obG	0,018	0,006	0,193
T.obJ	0,018	0,006	0,174
K.peA	0,165	0,184	0,000
K.peB	0,158	0,177	0,006
K.peE	0,174	0,174	0,006
K.peH	0,174	0,193	0,006
K.peI	0,157	0,174	0,006
K.peJ	0,165	0,184	0,012

[a] Gray boxes indicate the distances between the sequences from the canned samples and the same species individuals.

species analyzed (*T. thynnus*, *T. obesus*, *T. albacares*, and *T. alalunga*) that could be used as genetic markers to identify them with a high degree of confidence. Unseld et al. (*6*) performed the amplification and sequencing of a rather short mitochondrial cytochrome *b* of 59 bp, which did not allow the differentiation of all the species analyzed. Ram et al. (*7*) carried out the first work on canned tuna fish, amplifying a 119 bp fragment of the cytochrome *b* gene in a small number of samples. These works based the species identification on just the presence of specific nucleotides among the different species and no further analysis was performed. Quinteiro et al. (*8*) established that the measurement of the genetic distances could be used to evaluate the similarity of one unknown sequence with a pool of reference samples. The distance matrix calculated was used to construct a phylogenetic tree where each tuna species analyzed was well separated in a different cluster. Referring to the species of our interest, the authors found only one polymorphic site with diagnostic value between *T. albacares* and *T. obesus*, with five individuals analyzed (two yellowfin and three bigeye tunas). Due to the low number of sequences analyzed, intraspecific variation cannot be discarded for the single diagnostic position in one of the species, and, if that was the case, a wrong identification of the sample would be produced.

We have analyzed a 528 bp sequence in a total of 20 fresh samples of yellowfin and bigeye tunas and have found 9 diagnostic positions that presented no intraspecific variation, allowing a very precise identification of the analyzed samples. As shown in **Figure 2**, the sequencing of the 528 bp fragment might be useful to differentiate the 39 species included in this study, even the closest ones, as *T. albacares*, *T. atlanticus*, and *T. tonggol*, although more individuals of those species have to be analyzed to ensure the absence of intraspecific variation in the diagnostic positions.

We have also analyzed the biggest fragment obtained so far from canned tuna fish (171 bp) in 26 samples belonging to the three species of our interest. The sequences obtained from these samples did not show any intraspecific variation, and *T. albacares* presented two diagnostic positions, 67 and 124, with respect to *T. obesus*, and 21 with respect to *K. pelamis*. If we consider both the frozen and canned samples, we have analyzed 16 yellowfin and 25 bigeye tunas, a total of 41 different individuals. The fact that in all them the two differential

nucleotides did not present any intraspecific variation supports strongly their use as diagnostic positions.

To develop a quantitative approximation to the species determination, we have calculated a genetic distance matrix for the sequences analyzed. When an unknown sample is analyzed, the measurement of the genetic distances provides an objective criterion to assign one sample to one determined species. Simply by sequencing a DNA fragment and comparing it to a pool of previous sequences from different species, we can obtain a measurement of the distance, and it is expected that the analyzed sample will show the lowest genetic distances with respect to the phylogenetic group to which it belongs, allowing the species identification.

With the calculated distance matrix we constructed a phylogenetic tree, which showed how the sequences originated from individuals of the same species clustered together. We applied the bootstrap confidence test, which is one of the most commonly used tests of reliability of an inferred tree. In fact, bootstrap results are usually interpreted as a measure of the probability that a phylogenetic estimate represents the true phylogeny (*24*). It has been calculated that bootstrap proportions $\geq$70% usually correspond to a probability $\geq$95% that the corresponding cluster is real (*25*). In the phylogenetic trees we constructed for both the frozen and the canned samples, the BP values obtained for the clusters of the sequences from the species of our interest were always >70% (see **Figures 2** and **3**), indicating that the clustering is real and, consequently, that the species assignation is correct. Thus, this work provides for the first time an analysis that produces a quantitative measurement of the certainty of the assignation of an unknown sample to a determined species.

On the other hand, the application of the BP test to our data shows the reliability of our results, indicating that the sequencing of the 271 bp fragment of the cytochrome *b* gene from unknown canned samples and their comparison to a pool of control sequences from different species is a proper method to assign the species to which the different samples belong.

LITERATURE CITED

(1) Mackie, I. M. Identifying species of fish. *Anal. Proc.* **1990**, *27*, 89−92.

(2) Chow, S.; Kishino, H. Phylogenetic relationships between tuna species of the genus *Thunnus* (*Scombridae, Teleostei*): inconsistent implications from morphology, nuclear and mitochondrial genomes. *J. Mol. Evol.* **1995**, *41*, 741−748.

(3) Alvarado-Bremer, J. R.; Naseri, I.; Ely, B. Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the motochondrial DNA control regions. *J. Fish Biol.* **1997**, *50*, 540−554.

(4) Ward, R. D. Population genetics of tunas. *J. Fish Biol.* **1995**, *47*, 259−280.

(5) Bartlett, S. E.; Davidson, W. S. Identification of tuna species by the Polymerase Chain Reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Can. J. Fish. Aquat. Sci*. **1991**, *48*, 309−317.

(6) Unseld, M.; Beyermann, B.; Brandt, P.; Hiesel, R. Identification of the species origin of highly processed meat products by mitochondrial DNA sequences. *PCR Methods Appl.* **1995**, *4* (4), 241−243.

(7) Ram, J. L.; Ram, M. L.; Baidoun, F. Authentication of Canned Tuna and Bonito by Sequence and Restriction Site Analysis of Polymerase Chain Reaction Products of Mitochondrial DNA. *J. Agric. Food Chem.* **1996**, *44*, 2460−2467.

(8) Quinteiro, J.; Sotelo, C. G.; Rehbein, H.; Pryde, S.; Perez-Martin, R. I.; Rey-Mendez, M.; Mackie, I. M. Use of mtDNA Direct Polymerase Chain Reaction (PCR) Sequencing and PCR-Restriction Fragment Length Polymorphism Methodologies in Species Identification of Canned Tuna. *J. Agric. Food Chem.* **1998**, *46*, 1662−1669.

(9) Lostanlen, A.; Le Roux, M. G.; Etienne, M.; Loreal, H.; Moisan, J. P.; Le Vacon, F. Adaptation métodologique de lídentification génétique des poissons en conserve: thons, sardines et maquereaux. *Sci. Aliments* **1999**, *19*, 601−606.

(10) Kocher, T. D.; Thomas, W. K.; Meyer, A.; Edwards, S. V.; Paabo, S.; Villablanca, F. X.; Wilson, A. C. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci U.S.A.* **1989**, *86*, 6196−6200.

(11) Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *24*, 4876−4882.

(12) Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol*. **1993**, *10*, 512−526.

(13) Swofford, D. L.; Selander, R. R. *BIOSYS-1. A Computer Program for the Analysis of Alelic Variation in Genetics*; University of Illinois: Urbana, IL, 1981.

(14) Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406−425.

(15) Kumar, S.; Tamura, K.; Nei, M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci*. **1994**, *10*, 189−191.

(16) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

(17) Lynch, M.; Crease, T. J. The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* **1990**, *7*, 377−394.

(18) Hedges, B. The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. *Mol. Biol. Evol.* **1992**, *9*, 366−369.

(19) Mackie, I. M.; Pryde S.; Gonzales-Sotelo, C.; Medina, I.; Perez-Martin, R.; Quinteiro, J.; Rey-Mendez, M.; Rehbein, H. Challenges in the identification of species of canned fish. *Trends Food Sci. Technol.* **1999**, *10*, 9−14.

(20) Meyran, J. C.; Monnerot, M.; Taberlet, P. Taxonomic status and phylogenetic relationships of some species of the genus *Gammarus* (*Crustacea, Amphipoda*) deduced from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **1997**, *8*, 1−10.

(21) Brito, R. M.; Briolay, J.; Galtier, N.; Bouvet, Y.; Coelho, M. M. Phylogenetic relationships within genus *Leuciscus* (*Pisces, Cyprinidae*) in Portuguese fresh waters, based on mitochondrial DNA cytochrome *b* sequences. *Mol. Phylogenet. Evol.* **1997**, *8*, 435−442.

(22) Briolay, J.; Galtier, N.; Brito, R. M.; Bouvet, Y. Molecular phylogeny of Cyprinidae inferred from cytochrome b DNA sequences. *Mol. Phylogenet. Evol.* **1998**, *9*, 100−108.

(23) Song, C. B.; Near, T. J.; Page, L. M. Phylogenetic relations among percid fishes as inferred from mitochondrial cytochrome b DNA sequence data. *Mol. Phylogenet. Evol.* **1998**, *10*, 343−353.

(24) Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **1985**, *39*, 783−791.

(25) Hillis, D. M.; Bull, J. B. An empirical test of bootstrapping as a method for assesing confidence in phylogenetic analysis. *Syst. Biol.* **1993**, *42*, 182−192.

Received for review August 3, 2001. Revised manuscript received November 16, 2001. Accepted November 16, 2001. This work was supported by Conservas Garavilla S.A.

JF011032O