JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Toward Fish and Seafood Traceability: Anchovy Species Determination in Fish Products by Molecular Markers and Support through a Public Domain Database

MARC JÉRÔME,*,† JANN THORSTEN MARTINSOHN,‡ DELPHINE ORTEGA,‡ PHILIPPE CARREAU,‡ VÉRONIQUE VERREZ-BAGNIS,† AND OLIVIER MOUCHEL†

Département Sciences et Techniques Alimentaires Marines, IFREMER, Rue de l'Ile d'Yeu, B. P. 21105, F44037 Nantes Cedex 03, France, and Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), JRC.G.4 - Maritime Affairs, Via Enrico Fermi 2749 (TP 051), I-21027 Ispra (Va), Italy

Traceability in the fish food sector plays an increasingly important role for consumer protection and confidence building. This is reflected by the introduction of legislation and rules covering traceability on national and international levels. Although traceability through labeling is well established and supported by respective regulations, monitoring and enforcement of these rules are still hampered by the lack of efficient diagnostic tools. We describe protocols using a direct sequencing method based on 212−274-bp diagnostic sequences derived from species-specific mitochondria DNA cytochrome *b*, 16S rRNA, and cytochrome oxidase subunit I sequences which can efficiently be applied to unambiguously determine even closely related fish species in processed food products labeled "anchovy". Traceability of anchovy-labeled products is supported by the public online database AnchovyID (http://anchovyid.jrc.ec.europa.eu), which provided data obtained during our study and tools for analytical purposes.

KEYWORDS: Species identification; food product; anchovy; *Clupeomorpha*; *Engraulis*; cytochrome *b* gene; cytochrome oxidase subunit I gene; 16S rRNA gene; direct sequencing; FINS; public domain database

INTRODUCTION

International trade of fish and fish products has reached record highs, with a global capture fisheries production of 95 million tons in 2004 and an estimated first-sale value of US$84.9 billion (FAO). (*1*) These impressive numbers clearly illustrate the importance of the fishing industry, and its rise is paralleled by increasingly intricate global trading patterns. This, along with recent major food safety problems (e.g., BSE) and an elevated exigency with respect to quality distinction and product safety from the side of the consumer, stresses the necessity for the establishment of a comprehensive traceability system within the seafood business. Under EU law, traceability is part of a legal framework, with respect to food safety in general (*2*) and in the fisheries sector in particular by laying down labeling rules (European Commission (*3*)). However, although the legal framework for traceability and labeling of fish products is elaborate, monitoring and enforcement of the existing rules are complicated and insufficient because of the lack of efficient diagnostic tools. Additionally, in the case of ready-processed fish, substitution with cheaper products or unauthorized species is difficult to detect, especially if they are salted, ripened, or canned.

Anchovy is an important commercial species in the EU, with intensive fisheries in several European waters, (*4, 5*) and the incentive for fraud, consequently, is potentially high. Intensive fishing pressure led local stocks in some areas already at the edge of extinction. For this reason, following the advice delivered by the Scientific, Technical and Economic Committee on Fisheries, the fishing of anchovy in the Bay of Biscay in France is for the time being prohibited. (*6*) This clearly illustrates that monitoring and traceability in the anchovy fishery are a burning issue, because if further restrictions are applied, fraudulent substitutions of *Engraulis encrasicolus* by other species are likely to increase.

The European anchovy *E. encrasicolus*, prepared in the traditional manner (manual skinning and filleting), is protected by regional denomination legislation, but no common legislation exists in Europe. French law states than the label "anchovy" can be mentioned for products prepared from both *Engraulis* species, *E. encrasicolus* and *E. anchoita*, but Italian law permits

* Corresponding author. Telephone: +33 2 40 37 40 00. Fax: +33 2 40 37 40 71. E-mail: Marc.Jerome@ifremer.fr.
† IFREMER.
‡ IPSC.

10.1021/jf703704m CCC: $40.75
Published on Web 05/02/2008

only the species *E. encrasicolus*. Nevertheless, because of the similarity of flesh in both appearance and texture, it is not trivial to distinguish the five highly commercial *Engraulis* species *E. encrasicolus*, *E. anchoita*, *E. ringens*, *E. mordax*, and *E. japonicus*. Sebastio et al. (*7*) showed that semi-preserved anchovies are susceptible to substitution using closely related fish species as gilt sardine.

Another obstacle in the authentication of anchovy products is the genetic heterogeneity of *Engraulis* species, shown by several studies. Several authors showed this. Recently, Magoulas et al. (*8*) described 88 haplotypes of *E. encrasicolus* defining two phylogroups (A and B), which are separated by 3.2% sequence divergence. Yu et al. (*9*) examined sequence variation of a fragment of cytochrome *b* (cytb) and cytochrome oxidase C subunit I (COI) DNA among three populations of the Japanese anchovy, *E. japonicus*, from the Yellow Sea and East China Sea. In this work, variable nucleotide sites defined 29 and 32 haplotypes in cytb and COI data, respectively. Grant et al. (*10, 11*) showed from allozymes and mtDNA sequences a deep separation between New World anchovies (including *E. encrasicolus* and *E. japonicus*) and Old World anchovies (including *E. mordax*, *E. ringens*, *E. anchoita* and the genus *Anchoa*) and bipolarity in the anchovy genus *Engraulis*.

DNA-based authentication methods rely on the detection of species-specific sequences. Three molecular identification methods are currently widely used for processed food authentication applicable even if DNA is extensively damaged by manufacturing processes such as food-canned products. PCR-RFLP was used with success in thermally processed seafood to distinguish the European anchovy from gilt sardine (*Sardinella aurita*) (*7*), hake baby foods, and canned tuna (*12*−*15*). More recently, Takashima et al. (*16*) used PCR-RFLP to identify two commercially important species of Atlantic horse mackerel (*Trachurus trachurus*). Infante et al. (*17*) applied species-specific PCR and a mutiplex-PCR assay for the authentication of canned Atlantic mackerel (*Scomber scombrus*), as did Trotta et al. (*18*) for the identification of fish fillets from grouper (*Epinephelus* spp.). The third method was introduced by Bartlett and Davidson (*19*−*21*) when they reviewed problems encountered with species identification of raw and processed meat and underlined the fine interest of using PCR direct sequence analysis as the FINS method. The method was successfully used to authenticate processed cephalopods, commercial canned tuna, and sardine-type products. (*22*−*25*) Recently, Pepe et al. (*26*) proposed to use the direct sequencing method to identify various commercial surimi-based products, and Akasaki et al., (*27*) by proposing a method of authentication at the genus level, showed its value to control imported chirimen to Japan.

The aim of this project was to develop an effective operational diagnostic tool using valid pertinent molecular markers to determine fish species in processed food products labeled "anchovy". Three mitochondrial genes coding for 16S rRNA, cytochrome oxidase subunit I, and cytochrome *b* were tested for their ability to authenticate anchovy species.

In this study, we provide three sets of primers specifically designed to obtain three short diagnostic DNA fragments (212−274 bp) from these three mitochondrial genes. These diagnostic sequences can be used alone or jointly for reliable identification. This method can be applied for the detection of fraudulent or unintentional mislabeling of the five main commercial *Engraulis* species used for processed products. Using the direct sequencing method and calculating bootstrap values, we compared sequences from various anchovy products with reference sequences of 14 *Clupeomorpha* species, including the five main commercial *Engraulis* species. Several specimens of *E. encrasicolus* were analyzed to consider intraspecies variability.

The original sequences (~600 bp) from 16S, COI, and cytb, along with the short diagnostic sequences inferred, were deposited in the public database AnchovyID, which also provides tools for comparison and authentication.

Table 1. Fifteen Species (Reference Specimens) and 11 Commercial Anchovy and Sardine-Type Products Sampled for the Study [b]

reference specimens	scientific name	common name
	Engraulis encrasicolus (10)	European anchovy
	Engraulis japonicus	Japanese anchovy
	Engraulis anchoita	Argentine anchovy
	Engraulis ringens (2)	anchoveta
	Engraulis mordax	Californian anchovy
	Sardina pilchardus (2)	European pilchard
	Sardinella aurita	round sardinella
	Sardinella brasiliensis[a]	Brazilian sardinella
	Sardinella maderensis	Madeiran sardinella
	Ethmalosa fimbriata	Bonga shad
	Sardinops sagax	South American pilchard
	Coilia nasus	Japanese grenadier anchovy
	Sprattus sprattus	European sprat
	Clupea harengus	Atlantic herring
	Anchoa nasus	longnose anchovy

anchovy-type products	abbrev	origin
	A1	Argentina
	A2	Japan
	A3	Italy
	A4	Italy
	A5	Italy: taken on a pizza pie
	A8	France
	A12	France
	A13	France labeled *E. encrasicolus*
	A14	France

sardine-type products	abbrev	origin
	A7	Asia
	A10	Asia

[a] *Sardinella brasiliensis* is considered as a synonym of *Sardinella janeiro* (Fishbase, ver April 2008). [b] The numeral in parentheses indicates the number of specimens considered in this study.

MATERIALS AND METHODS

Sampling: Fish and Products. The fish samples included in the study (**Table 1**) were collected either as whole specimens or as tissue samples (frozen or preserved in 70% (v/v) ethanol). The samples were chosen according to criteria assuring a widespread selection representing the major genera, various species in processed products, and gilt sardine of closely related fish species. Additionally, some species of less commercial importance were included that could be used for substitutions, intentionally or unintentionally.

Among the samples were specimens of the five main commercial species of anchovy belonging to the genus *Engraulis* (Engraulidae family): *E. encrasicolus* (European anchovy), *E. anchoita* (Argentine anchovy), *E. ringens* (Peruvian anchovy), *E. mordax* (Northern anchovy), and *E. japonicus* (Japanese anchovy).

Sampling was supplemented by other species to include (1) phylogenetically closely related species (specimens representative of each subfamily of Engraulidae: Coiliinae (genus *Coilia*) and Engraulinae (genus *Anchoa*)), (2) species whose commercial name or origin could lead to confusion or diagnostic errors especially with small pelagic canned products, and (3) specimens of Clupeidae belonging to the three commercially important genera *Sardina*, *Sprattus*, and *Sardinella*.

To include investigations on intraspecies variations of *E. encrasicolus*, 10 specimens were collected from the Bay of Biscay: two

Table 2. Primer Sequences for PCR Amplifications of Long and Short DNA Fragments of the 16S, COI, and Cytb Genes

gene	primer sequence 5′−3′		reference
cytb	CTAACCCGATTCTTTGCCTTCCACTTCCT	CytBI-7F	Bautista et al. (*24*)
	CCGACTTCCGGATTACAAGACCG	Trucytb-R	Bautista et al. (*24*)
cytb (short DNA fragment)	GAGTGGTACTTCCTCTTTGCCTACGC	Cytb-Fc	this study
	GGGCTTTGTTCTCCAATCACCCTGC	Cytb-Rc	this study
16S	GCCTGTTTAACAAAAACAT	MIY L2510	Miya and Nishida (*28*)
	CGGTCTGAACTCAGATCACGT	MIY H3059	Miya and Nishida (*28*)
16S (short DNA fragment)	TCTGTCCGTGCAGAAGCGG	Enc16S-Fc	this study
	TTGCGCTGTTATCCCTAGG	Enc16S-Rc	this study
COI	GGTCAACAAATCATAAAGATATTGG	LCO1490	Folmer et al. (*29*)
	TAAACTTCAGGGTGACCAAAAAATCA	HCO2198	Folmer et al. (*29*)
COI (short DNA fragment)	TCTAATCGGTGGGTTCGGGAATTGACTA	EncCOI-Fc	this study
	TGGCAGGGATTTCATCAATCCTAGG	EncCOI-Rc	this study

specimens (EeG1 and EeG2) were caught near the French river Gironde, two were from the south of the bay (EeS1 and EeS2), and five other specimens (Eng enc 1 to Eng enc 5) were collected in different areas of the Celtic Sea.

Eleven commercial anchovy products such as canned products and semi-preserved foodstuff (salted, marinated, and ripened) processed in European countries including France and Italy and in South American and Asian countries were examined (**Table 1**). These commercial products were purchased at local supermarkets or supplied by collaborators. Nine typical anchovy products and three sardine-type products were analyzed. Only one of the commercial anchovy products clearly mentions the Latin name of the species used. The samples consisted of traditional anchovy products such as fillets in olive or vegetable oil, sometimes with piquant sauce or capers or marinated with seasoning, semi-preserved or canned, and canned products labeled "sardines/sardinas" prepared in vegetable oil (groundnut or olive oil) sometimes also with seasoning such as lemon or red chilli.

DNA Isolation, Amplification, and Sequencing. Total genomic DNA extraction of references specimen was performed according to previously described procedures used for Clupeidae studies: a phenol/chloroform/isoamyl alcohol method (PCI) (*13, 24*) and a Chelex method (Bio-Rad). (*25*)

For commercially processed products, oil and lipids were removed from sample tissue by soaking them in water before the DNA extraction. Subsequently, they were washed using 70% ethanol followed by a second washing step with distilled water. DNA of "defatted" samples was isolated following the same procedures as described above.

PCR amplifications were carried out using Hybaid PCR Express (Hybaid) and Bio-Rad MyCycler (Bio-Rad). The reactions were set up in volumes of 50 μL (references species) or 100 μL (processed products) containing PCR buffer [75 mM Tris-HCl, pH 9.0; 50 mM KCI; 20 mM $(NH_4)_2SO_4$, 2 mM $MgCI_2$]; 400 μM dNTP mix; 0.2 μM each primer; 2.5 units of UptiTherm DNA polymerase (Uptima-Interchim); and 0.5−1 μL of template DNA (PCI extracts) or 1−3 μL of DNA solution (Chelex supernatant).

Cycling conditions were

- For the 16S gene: 94 °C for 5 min, 35 cycles [94 °C for 30 s, 55 °C (long and short DNA fragments) for 30 s, 72 °C for 60 s], followed by a final extension for 5 min at 72 °C;
- For the COI gene: 94 °C for 180 s, 35 cycles [94 °C for 30 s, 48 °C (long DNA fragment) or 48 to 55 °C (short DNA fragment depending of the primers) for 30 s, 72 °C for 30 s], followed by a final extension for 5 min at 72 °C;
- For the cytb gene: 95 °C for 5 min, 35 cycles (94 °C for 30 s, 50 °C for 45 s, 72 °C for 60 s), followed by a final extension for 5 min at 72 °C (long DNA fragment) and 94 °C for 180 s, 35 cycles (94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s), followed by a final extension for 5 min at 72 °C (short DNA fragment).

For optimal results in some cases, the cycling conditions were appropriately adapted in terms of cycling number (between 30 and 40) and annealing temperature.

The DNA amplification was controlled on 1.5−3% agarose gels (Agarose HR, Uptima-Interchim), using TAE buffer (2 mM EDTA, 40 mM Tris acetate, pH 8.5). Ethidium bromide was used for band visualization via ultraviolet transillumination (Image Master VDSCL, Amersham Pharmacia Biotech). The size of expected PCR products was estimated using the GeneRuler 100bp DNA ladder plus and the GeneRuler DNA ladder mix (MBI Fermentas). For sequencing reactions, the DNA template was quantified using the MassRuler DNA ladder, low range (MBI Fermentas) on agarose gels.

For the reference species, three sets of primer pairs (**Table 2**) were used in the PCR amplifications to obtain 600/700-bp DNA fragments of 16S rRNA, (*28*) COI, (*29*) and cytb. (*30*)

On the basis of the DNA sequences determined for these DNA fragments, we designed three primer sets (**Table 2**, short DNA fragment 200/300 bp) for the amplification of short species-distinctive fragments of the COI, 16S rRNA, and cytb genes. For each gene as primer anchor regions two conserved zones flanking the diagnostic target region of the examined species, especially the five main *Engraulis* species, were identified.

Cleanup and Sequencing of PCR Products. Before sequencing, double-stranded PCR products were purified by filtration through a Qiagen QIA quick column according to the manufacturer's protocol. PCR fragments were used for direct cycle sequencing with the dye terminator cycle sequencing kit (Beckman). Sequencing analysis was performed with a Beckman Coulter CEQ 8000 DNA sequencer in both directions with the primers used for PCR amplification.

Sequence Alignment, Genetic Distances, and Phylogenetic Analyses. The DNA sequences were edited with BioEdit software. (*31*) Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.1. (*32*) Nucleotide divergences were computed using the Tamura−Nei model, (*33*) which takes substitutional rate biases and the inequality of base frequencies into account. Phylogenetic trees were constructed using the neighbor-joining (NJ) method, (*34*) and the robustness of topology nodes was tested by the bootstrap method with 1500 iterations.

Database Development and Implementation. The AnchovyID database (AIDB) structure was implemented using an object relational database management system, PostGreSql (v8.1), and administrated by the Web-based tool phpPgAdmin. Queries to the database (to create, retrieve, update, and delete data) were carried out by using structured query language. It was hosted and managed by the JRC Ispra and was run on a Red Hat Enterprise Linux 5 open source platform.

The Web-based application allowing the end users to access the AIDB was developed using the latest available Web technologies, such as Java2 Enterprise and Java Server Pages (JSP) for the implementation and hypertext markup language and cascading style sheets for the graphical part of the interface. This Web interface was run on an Apache Tomcat (v5.5) Web server (servlet container that supports our JSPs). The Web site URL is http://anchovyid.jrc.ec.europa.eu.

Starting from the homepage, the end user can navigate through a set of menus and tools for database mining, to visualize and analyze the data inserted in AIDB's sets, such as the DNA sequences and origins of the studied samples. Moreover, documentation material, such as the protocols, is available and downloadable from the Web application.

Table 3. Specimens of Clupeidae and Engraulidae Species Considered for Alignments and Nucleotide Divergences Estimations Using the Tamura−Nei Model[a]

species/specimens	Abb	COI (594 bp)	cytb (574 bp)	16S (588 bp)
Engraulis encrasicolus	Ee	X (AM911166)	X (AM911183)	X (AM911209)
	EeG1	X (AM911180)	X (AM911184)	X (AM911212)
	EeG2	X	X	X
	EeS1	X (AM911181)	X (AM911185)	X (AM911211)
	EeS2	X (AM911182)	X	X
Engraulis japonicus	Ej	X (AM911168)	X (AM911190)	X (AM911214)
Engraulis japonicus (AB040676)	Ej-N	N	N	N
Engraulis anchoita	Ea	X (AM911169)	X (AM911194)	X (AM911215)
Engraulis ringens	Er	X (AM911170)	X (AM911192)	X (AM911210)
	Er2		X (AM911193)	X (AM911215)
Engraulis mordax	Em	X (AM911171)	X (AM911191)	X (AM911213)
Engraulis mordax (AB246182)	Em-N			N
Coilia nasus	Cn	X (AM911172)	X (AM911196)	X (AM911208)
Coilia mystus (DQ315693.1)	Cm-N			N
Anchoa nasus	An		X (AM911195)	
Anchoa compressa (AY846745)	Ac-N			N
Sardina pilchardus	Sp	X (AM911167)	X (AM911198)	X (AM911200)
Sardinella aurita	Saur	X (AM911173)	X (AM911197)	X (AM911207)
Sardinella brasiliensis	Sbra	X (AM911174)		X (AM911206)
Sardinella maderensis	Smad	X (AM911175)		X (AM911205)
Clupea harengus	Ch	X (AM911176)		X (AM911204)
Sprattus sprattus	Sspr	X (AM911177)	X (AM911199)	X (AM911201)
Sardinops sagax	Ssag	X (AM911178)		X (AM911202)
Ethmalosa fimbriata	Ef	X (AM911179)		X (AM911203)

[a] X indicates that the DNA sequences were obtained and aligned during this study, and their genetic distances were compared. X (EMBL accession numbers) indicates that the DNA sequences were submitted to the EMBL bank. N means references were taken from GenBank (NCBI). For cytb, five additional DNA sequences were obtained from *E. encrasicolus* specimens and considered for alignments Eng-enc1, Eng-enc2 (AM911187), Eng-enc3 (AM911186), Eng-enc4 (AM911189), and Eng-enc5 (AM911188).

RESULTS AND DISCUSSION

This study shows that the short diagnostic sequences derived from the three mitochondrial DNA genes, 16S rRNA, cytochrome *b*, and cytochrome oxidase I (**Table 2**), can conveniently be used, either separately or jointly, with a direct sequencing protocol, to authenticate processed anchovy products. Using this method, we obtained unambiguous identification for various heat-sterilized samples and ripened products. The method is therefore of practical value to all laboratories concerned with the authentication of this kind of seafood product and is made available as a toolkit, containing the DNA sequences, protocols, and sequence analysis algorithms, in the public AIDB.

Long DNA Sequences. Previous to the identification and characterization of the short diagnostic sequences, the three mitochondrial genes (16S, COI, and cytb) were tested for their suitability to differentiate and authenticate Engraulidae species. Partial COI, 16S, and cytb regions were amplified using, respectively, the primers of Folmer et al., (*29*) Miya and Nishida, (*28*) and Sevilla et al. (*30*) In total, 19 sequences of 594 bp for COI, 23 sequences of 588 bp for 16S rRNA (gaps included), and 16 sequences of 574 bp for cytb were aligned, and the Tamura−Nei distances (TN distances) (*33*) between sequences were calculated for each gene. The specimens analyzed are listed in **Table 3**.

Additionally, COI, cytb, and 16S rDNA sequences were taken from GenBank (NCBI) to further explore the overall nucleotide divergences.

The phylogenetic relationships between the reference species studied are illustrated by NJ trees in **Figure 1A−C**.

For COI (**Figure 1A**), the two families, Engraulidae and Clupeidae, appear well separated into clades as shown by the bootstrap value at the node that is maximal for Engraulidae. Inside this family, two subgroups, the "North anchovies" (*E. japonicus* (*E.j.*), *E. mordax*, and *E. encrasicolus* (*E.e.*)), the South American anchovies (*E. ringens* and *E. anchoita*), and the species *Coilia nasus* are distinguished with the maximum bootstrap value of 100%. Also, *E.e.* and *E.j.* are clearly separated from the Californian anchovy *E. mordax* (bootstrap (BT) = 99). All specimens of *E. encrasicolus* are clearly separated from the closely related species *E. japonicus* (two haplotypes and a bootstrap value equal to 100%). The five aligned COI sequences of *E. encrasicolus* differ in a few nucleotide positions (97.5% of identity). The sequences show that four haplotypes (A, B, C, D) can be defined: A for EeG2 and Ee, B for EeS1, C for EeS2 but highly similar to B, and D for EeG1.

For 16S (**Figure 1B**), the two families, Engraulidae and Clupeidae, are also well separated: the bootstrap value at the node is maximal for the Clupeidae clade. Inside the Engraulidae clade, two subclades, the "North anchovies" (*E. japonicus*, *E. mordax*, and *E. encrasicolus*) and the species *Anchoa compressa* and the South American anchovies (*E. ringens* and *E. anchoita*) are distinguished with the maximum bootstrap value of 100%. Inside the Engraulidae species, *C. nasus* and *Coilia mystus* are grouped and isolated. The two specimens of *E. ringens* are similar. Regarding the two close species *E. encrasicolus* and *E. japonicus*, all the specimens of *E. encrasicolus* are clearly separated from the two haplotypes of *E. japonicus* (bootstrap value of 99%).

The sequence alignment shows a high similarity in the sequences of *E. encrasicolus* (>99% of identity). The size of the 16S sequences is 568 bp except for the specimen EeG1, which is 569 bp. Three haplotypes (A, B, C) can be distinguished: A for Ee and EeG2, B for EeS1 and EeS2, and C for EeG1. This grouping based on the haplotypes is in accordance with the results obtained for the COI sequence. It also confirms the identification of specimen EeG1, which could be distinguished from the other *E. encrasicolus* specimens.

For cytb (**Figure 1C**), both families, Engraulidae and Clupeidae, are well separated. Inside the Engraulidae group, two subgroups (*E. japonicus* and *E. encrasicolus*) and the South American anchovies (*E. ringens* and *E. anchoita*) are distinguished with the maximum bootstrap value of 100%. *E. mordax*

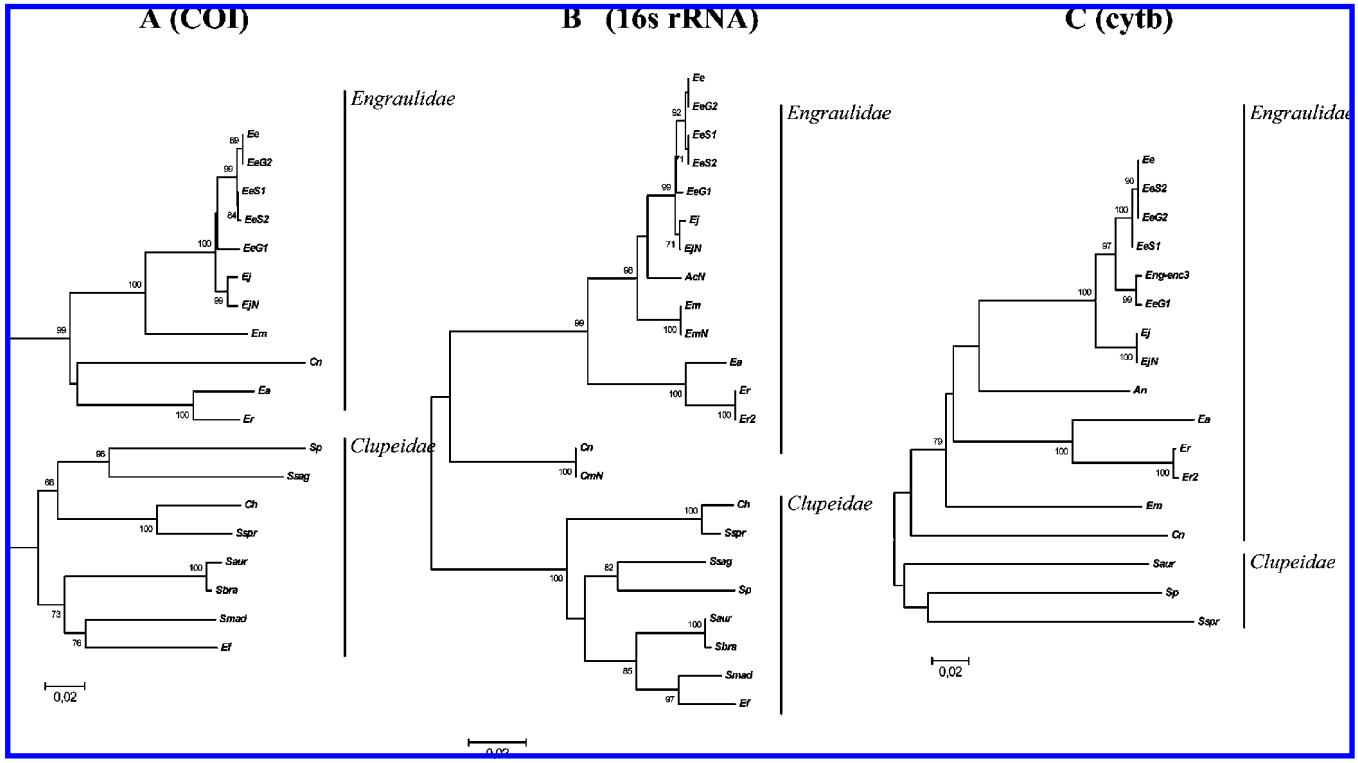

Figure 1. (**A**) NJ tree inferred from TN distances between 19 sequences of partial COI gene in 14 species of *Clupeiformes* (six Engraulidae species and eight Clupeidae species). Numbers denote bootstrap percentages based on 1500 replications. Only bootstrap scores greater than 60% are shown. (**B**) NJ tree inferred from TN distances between sequences of partial 16S gene (23) in 14 species of *Clupeiformes* (eight Engraulidae species (15 sequences) and eight Clupeidae species (eight sequences)). Numbers denote bootstrap percentages based on 1500 replications. Bootstrap scores greater than 70% are shown. (**C**) NJ tree inferred from TN distances between sequences of partial cytb gene (16) in 11 species of *Clupeiformes* (eight Engraulidae species and three Clupeidae species). Numbers denote bootstrap percentages based on 1500 replications. Bootstrap scores greater than 60% are shown.

is clustered with an 80% score. Inside the subgroup, *E. japonicus* and *E. encrasicolus*, the two species are strongly distinguished with the maximum bootstrap value of 100%. Inside the subgroup of the South American anchovies (*E. ringens* and *E. anchoita*), the two specimens of *E. ringens* are closely related and clustered with a 100% score.

The alignment shows that all sequences obtained from the *E. encrasicolus* specimens are highly similar (96% of identity). Three haplotypes (A, B, C) can be distinguished: A as Ee, EeG2 and EeS2, B as EeS1, and C as EeG1. As previously noted for both 16S and COI genes, EeG1 is a specimen of *E. encrasicolus* slightly different by its sequence from all the other *E. encrasicolus* specimens. Analysis of this sequence shows 13 variable sites at the third codon position.

Diagnostic DNA Fragments. Encouraged by the resolution power obtained with the long DNA fragments, we explored whether short variable regions inside 16S, COI, and cytb sequences could be used for identification purposes, since this would make the analytical procedure more convenient. The candidate regions chosen display an interesting nucleotide polymorphism and are flanked by two conserved zones allowing the design of primers for amplification of the target *Engraulis* species. PCRs were performed on commercial anchovy or sardine-type products in the same way as with all the reference specimens collected. Five new specimens of *E. encrasicolus* (named Eng-enc 1−5 and collected in different areas of the Celtic sea) were added to the reference specimens collection to further enlarge the sample size for the investigation into intraspecies variation. Multiple sequence alignments of the short fragments from the three genes 16S, COI, and cytb were performed. On the basis of these alignments, the TN distances were estimated using diagnostic nucleotide positions (i.e., sequence differences) of the reference specimens and commercial processed samples for each gene. The TN distance estimates were used to infer NJ trees. To simplify the NJ trees, the group of Clupeidae references was sometimes condensed and only haplotypes of species were shown in particular for *E. encrasicolus*.

Diagnostic Cytb DNA Fragment. Primers for the amplification of a short DNA fragment of cytb were designed to preferentially amplify Engraulidae species. For sardine-type products such as A7 and A10, the PCR remained below the detection level. In this study, the short sequences for the three Clupeidae references were taken from the long DNA sequences of cytb. A total of 27 sequences of 274 bp were aligned, and nucleotide divergences were computed.

Inside the genus *Engraulis*, species divergence was 0.05 and 0.07 between the two closely related anchovy species *E. japonicus* and *E. encrasicolus*. Intradivergence for *E. encrasicolus* was 0−0.04. The *E. encrasicolus* references were clearly discriminated from the other anchovy species, *E. mordax* and the South American anchovies (0.21−0.23). Divergence between the two South American anchovy species (*E. ringens* and *E. anchoita*) was 0.08. The genetic distances between the *E. encrasicolus* references and *A. nasus* were 0.18−0.19. The genetic distances between the two families Engraulidae and Clupeidae exceeded 0.25 ($0.25 < D < 0.35$).

An NJ tree (**Figure 2C**) was inferred from the TN distances between the diagnostic cytb sequences of the reference specimens and the commercially processed samples. Bootstrap analysis provided strong support for clades associated with *E. japonicus* (BT = 99) and the South American *E. anchoita* and

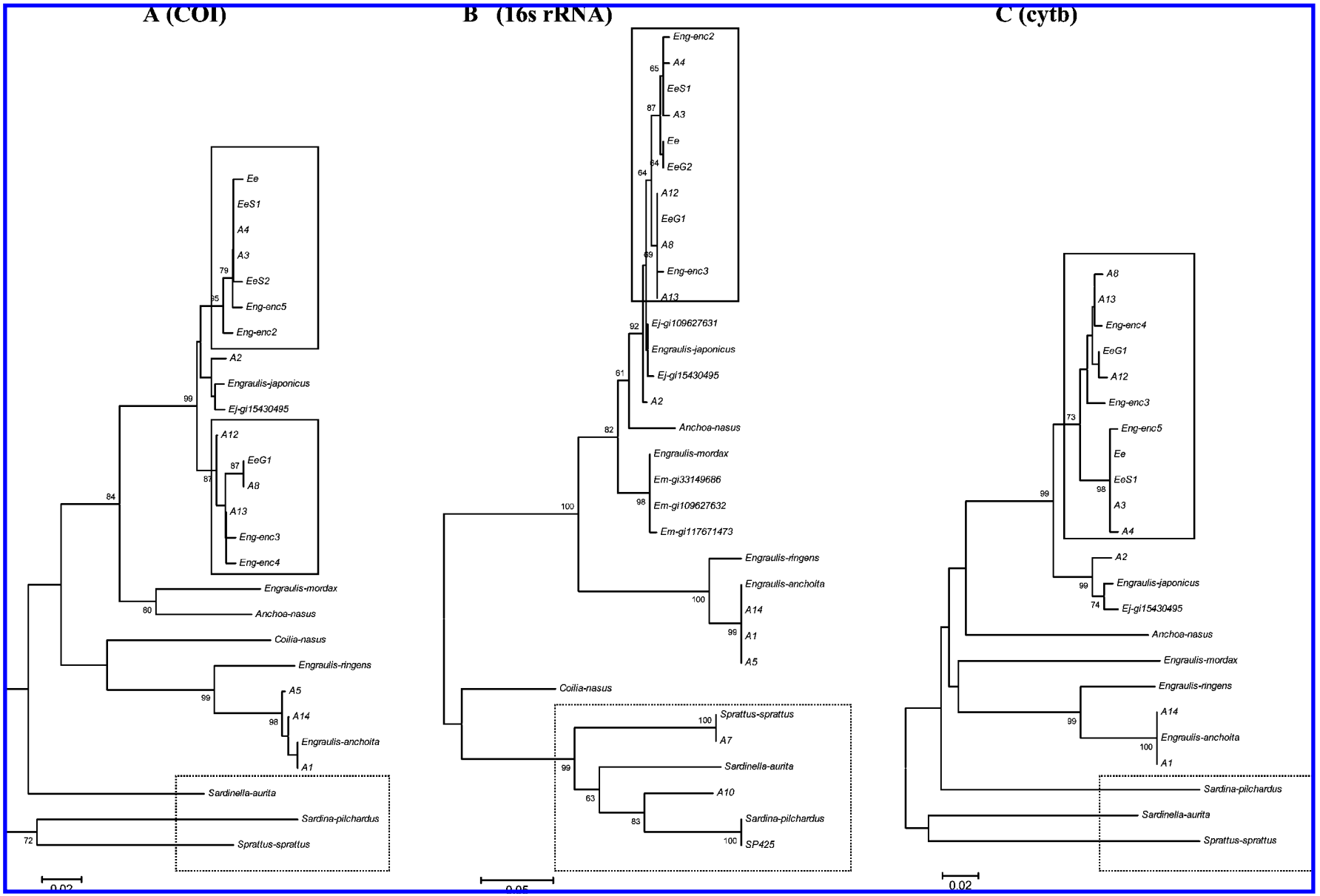

Figure 2. NJ trees inferred from TN distances between sequences of partial COI gene (212 bp) (**A**), between sequences of partial 16S gene (257 bp) (**B**), and between sequences of partial cytb gene (274 bp) (**C**). Numbers denote bootstrap percentages based on 1500 replications. Bootstrap scores greater than 60% for 16S and 70% for COI and cytb are shown. Haplotypes of *E. encrasicolus* are shown by solid squares. Dotted squares group the species belonging to the Clupeidae. Accession numbers are: for COI sequences of *E. encrasicolus*: AM911224 (Eng-enc2), AM911225 (Eng-enc3), AM9112226 (Eng-enc4), AM911227 (Eng-enc5), *A. nasus* AM911223; for 16S sequences of *E. encrasicolus*: AM911217 (Eng-enc2), AM911218 (Eng-enc3), AM9112219 (Eng-enc4), AM911220 (Eng-enc5), *A. nasus* AM911221, *Sardina pilchardus* 425 a.m.911222; and for cytb sequences of *E. encrasicolus*: AM911229 (Eng-enc2), AM911230 (Eng-enc3), AM9112231 (Eng-enc4), AM911232 (Eng-enc5), *A. nasus* AM911228.

E. ringens (BT = 99). Inside the cluster (formed by *E. encrasicolus* and *E. japonicus*), strongly supported by bootstrapping (BT = 99), the two groups of reference species *E. japonicus* (BT = 99) and *E. encrasicolus* (BT = 73) are clearly separated. Inside the cluster, *E. encrasicolus* haplotypes define two subgroups of *E. encrasicolus*.

A clear identification is possible for all anchovy-type products analyzed: A8, A12, and A13 are identified as *E. encrasicolus* and appear as closely related. A3 and A4 from Italy are grouped together and authenticate as *E. encrasicolus*. Products from Argentina (A1) and France (A14) are clearly identified as *E. anchoita*. The Japanese anchovy product (A2) is located near the two references of *E. japonicus*, and the bootstrap value of this group is 99.

No misidentification of anchovy products compared with Clupeidae references or misidentification of canned sardine products compared with Engraulidae references occurred.

Diagnostic COI DNA Fragment. In total, 29 sequences of 212 bp were aligned and analyzed. Inside the genus *Engraulis*, species divergence was 0.03 between the two closely related anchovy species *E. japonicus* and *E. encrasicolus*. Intraspecific divergence for *E. encrasicolus* was 0.00−0.05. Distances between *E. encrasicolus* and the *E. mordax* or *A. nasus* specimens were 0.12−0.14. Distances between *E. encrasicolus* and the South American anchovies were 0.19−0.23. Divergence between the two South American anchovy species (*E. ringens* and *E. anchoita*) was 0.08. The genetic distances between the five *Engraulis* species and Clupeidae were high (0.20−0.29). Both *E. mordax* and *A. nasus* group together, but the divergence between these two species is comparably high (0.22).

The NJ tree inferred from TN distances between the diagnostic COI sequences of the reference specimens and commercially processed samples is shown in **Figure 2A**. Inside the cluster *E. encrasicolus* and *E. japonicus*, haplotypes of *E. encrasicolus* define two subgroups. One of these subgroups (EeG1, Eng-enc 3 and 4) is clearly separated (BT = 99) from the second subgroup of *E. encrasicolus* specimens (Ee, EeG2, EeS1, EeS2, Eng-enc 2 and 5) while apparently being closely related to the subgroup formed by the haplotypes of *E. japonicus*.

Inside the cluster (*E. encrasicolus* and *E. japonicus*) A3, A4, A8, A12, A13 can be identified as *E. encrasicolus*, and A2, the Japanese anchovy product, as *E. japonicus* (BT = 68). Anchovy-labeled products from Argentina (A1), France (A14), or Italy (A5) are grouped together in the same cluster with the reference species *E. anchoita*.

When compared with Clupeidae references, no misidentification of anchovy products occurred.

Diagnostic 16S rRNA DNA Fragment. Primers for the amplification of a short fragment of 16S rRNA for both Engraulidae and Clupeidae species were designed. A new specimen of *Sardina pilchardus*, designated as SP425, was sequenced and included in the multiple alignments. Thirty sequences were obtained, and with the inclusion of five GenBank sequences, in total 35 sequences were aligned, spanning 257 nucleotide sites with gaps.

Inside the genus *Engraulis*, species divergence was 0.00/0.02 between the two closely related anchovy species *E. japonicus* and *E. encrasicolus*. Intraspecific divergence for *E. encrasicolus* was 0.01−0.02. The *E. encrasicolus* references were well discriminated from the other anchovy species except for *E. japonicus*. Distances between *E. encrasicolus* and the group of Californian *E. mordax* specimens were 0.05/0.06, 0.16 to 0.17 with the South American anchovies, and relatively close to *A. nasus* (0.05/0.06). Divergence between the two South American anchovy species (*E. ringens* and *E. anchoita*) was 0.04. The genetic distances between the two families Engraulidae and Clupeidae were high and exceeded 0.30. Overall, the short DNA sequence 16S clearly separated the two families Engraulidae and Clupeidae. However, the 16S rRNA sequence is by far less powerful than both the COI and cytB sequences to discriminate the species belonging to the Engraulidae, especially *E. encrasicolus* and *E. japonicus*.

An NJ tree (**Figure 2B**) was inferred from TN distances between the diagnostic 16S sequences of the reference specimens and commercially processed samples. The Clupeidae and Engraulidae families are clearly separated. Bootstrap analysis provided strong support for clades formed by *E. mordax* (BT = 98), the South American anchovies *E. anchoita* and *E. ringens* (BT = 100), and the Clupeidae species (BT = 99). Haplotypes of *E. encrasicolus* define two subgroups identical to those previously observed with cytb. The discrimination between the two closely related species *E. encrasicolus* and *E. japonicus* is weak. Identification was possible for all the anchovies and sardine-type products analyzed: A8, A12, and A13 are clustered together and identified as *E. encrasicolus*, A3 and A4 from Italy are grouped together and authenticated as *E. encrasicolus*, and A1, A5, and A14 are clearly authenticated as *E. anchoita*. The Japanese anchovy product A2 is located near the two references of *E. japonicus*.

An Asian sardine-type product (A7) was clustered with *Sprattus sprattus*. Several species belonging to the genus *Sprattus* are present in the Pacific Ocean, and some of them are commercially fished and processed. The Asian sardine-type product (A10) appears to be closely related to the two species *S. aurita* and *S. pilchardus*: the available reference species of Clupeidae are not sufficient to authenticate this product probably derived from a species of *Sardinella*.

Concatenation. The set of the 16 rRNA primers allowed us to reliably amplify short sequences of both anchovy- and sardine-type products. Regarding *Engraulis* species, a correct identification was possible except for the closely related species *E. encrasicolus* and *E. japonicus*. Concurrently, both sets of COI and cytb primers (more specific Engraulidae) were able to amplify short sequences from all the anchovy-type products but not from sardine-type products. In both cases, "*E. encrasicolus*" anchovy-type products were unambiguously identified as belonging to one of the two haplotype groups formed by specimens of *E. encrasicolus* in particular with respect to cytb.

We propose here to combine the specificity and resolution of each of the three markers and to use this approach (concatenated diagnostic sequences of the three mitochondrial genes: COI, 16s, cytb) as a means to increase the diagnostic power. Analysis of the combined data sets using the NJ method resulted in a highly resolving "diagnostic tree" (**Figure 3**) where most of the clades are supported by moderate to high bootstraps values. Both the Clupeidae and Engraulidae families are clearly separated from each other. Bootstrap analysis provided strong support for clades associated with Engraulinae genera (BT = 99). Inside these, each species of the genus *Engraulis* is clustered together (BT = 99/100). Haplotypes of *E. encrasicolus* define two subgroups of *E. encrasicolus* (BT = 75). The group of *E. encrasicolus* is well separated from the clade including both *E. japonicus* references (BT = 100). Each of the subgroups of *E. encrasicolus* consist of the same haplotypes as previously described.

Tests of the "concatenated tree" as a tool of authentication were realized, and the results for A2, A4, and A14 COI, 16S,

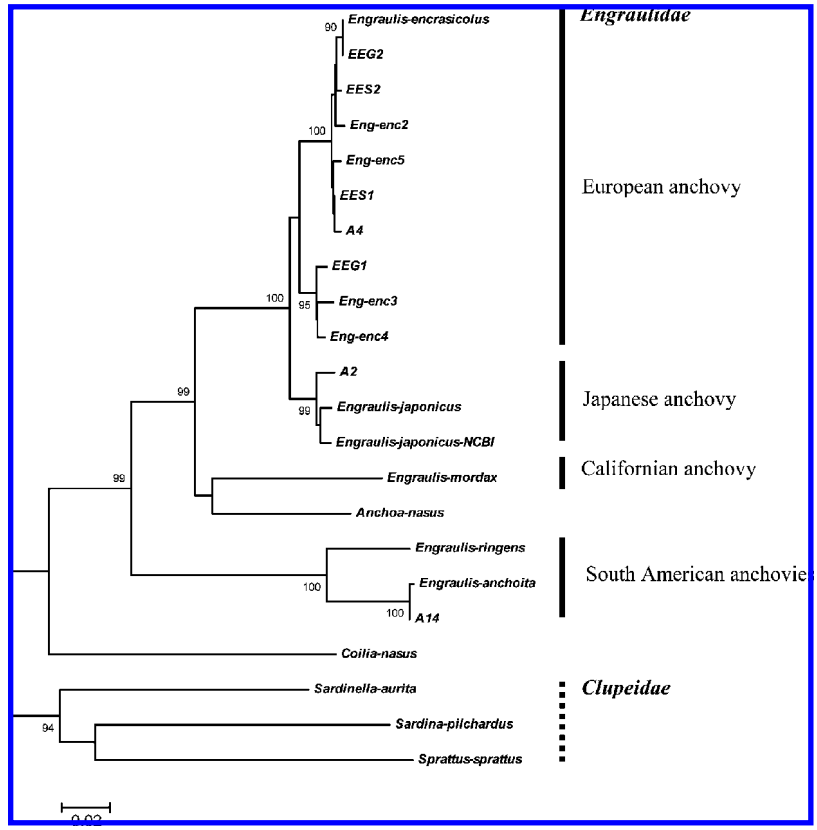

Figure 3. NJ tree inferred from TN distances between 19 concatenated diagnostic sequences of COI, 16S, and cytb gene (10 species of *Clupeiformes* composed of seven Engraulidae species and three Clupeidae species. A2, A4, and A14 are COI, 16S, and cytb concatenated sequences from anchovy-type products. Numbers denote bootstrap percentages based on 1500 replications. Bootstrap scores greater than 90% are shown.

cytb concatenated sequences are shown in **Figure 3**. They were unequivocally identified, respectively, as *E. japonicus*, *E. encrasicolus*, and *E. anchoita* anchovy-type products with a high bootstrap value of 99/100.

In summary, this study allowed us to obtain original sequences (600 bp on average) with high discrimination power for authentication purposes from the 16S rRNA, cytochrome oxidase subunit I, and cytochrome *b* genes for 15 *Clupeomorpha* species including the five main commercial *Engraulis* species. Sampling was carried out in such a way that each main genus of Engraulidae (subfamily Coilliinae, genus *Coilia* and Engraulinae genera *Anchoa* and *Engraulis*) and Clupeidae (subfamily Clupeinae, genera *Clupea*, *Sardinops*, *Ethmalosa*, *Sardina*, *Sardinella*, *Sprattus*) contained representatives of highly commercial species. Ten various specimens of *E. encrasicolus* were collected and analyzed to improve the authentication's method and to consider the intraspecies variation of this species.

For *E. encrasicolus*, we observed two groups of haplotypes, and phylogenetic trees showed that sequences of *E. encrasicolus* are grouped into two different clusters. This was consistent with the results found by Magoulas et al. (*8*) It would have been interesting to proceed in the same manner with the closely related species *E. japonicus* to increase the diagnostic resolution power inside the clade *E. japonicus*. Sequences for the tree genes of both species *E. japonicus* and *E. encrasicolus* showed that they are very closely related and clearly distinct from the North and South American species. These results were coherent with those reported by Grant et al. (*10*), Grant (*11*).

Three short mitochondrial diagnostic DNA fragments of COI (212 bp), 16S rRNA (257 bp with gaps), and cytb (274 bp) were identified within the long sequences previously described. These allowed us to classify most of the commercial samples collected for this study and to test a direct sequencing method with *Clupeomorpha* products such as commercial anchovies and some canned sardine-type products.

The 16S rRNA diagnostic sequence was particularly powerful with respect to identifying all the specimens of both families and all the anchovy- and sardine-type products at the genus level.

The COI and cytb markers proved to be similar with respect to their resolution power. Distance as measured by nucleotide diversity was more important between the closely related species *E. encrasicolus* and *E. japonicus*, and particularly with the cytb marker. Cytochrome *b* was the most powerful in discriminating the species *E. encrasicolus* and *E. japonicus* and was able to put together both haplotype subgroups of *E. encrasicolus* in the same cluster with a high level of reliability.

For both the closely related species *E. encrasicolus* and *E. japonicus*, no confused relationship was observed with the clades of the South American species of *E. ringens* and *E. anchoita* nor *E. mordax* or with both genera *Coilia* and *Anchoa*. No misidentification was reported between anchovy products and sardine-type products.

We propose that each diagnostic sequence could be used separately or pooled (concatenated sequences) in case of doubt or difficulties regarding the use of one of these three markers separately, or depending on the level of identification (species/genus) required for a diagnosis. This approach seems very valuable to us since it is sometimes difficult to get clear results with only one marker (because of problems with PCR and/or sequencing) and also a double or a triple check is a good way to back up the diagnosis.

AnchovyID Database. Traceability along the food production chain and product authentication plays an ever-increasing role in the food industry. While molecular methods have undoubtedly

a high potential in this respect, their effectiveness and support to the research realm as well as control authorities can only be guaranteed as long as data and tools are easily accessible and the resources are known. AnchovyID is implemented as a strategic tool in line with these requirements.

The AIDB contains all information acquired during this study. It was designed to allow for high flexibility and, as such, is capable of not only archiving already-collected data but also accommodating new data sets. AIDB is conceptually divided into two main parts, representing the main sources of data stored within it. The first set of data is derived from the specimens of the examined species, whereas the second contains reference sequences obtained from products labeled anchovy. It also contains information concerning the protocols used on the samples for data collection.

Upon entering AIDB, it is possible to directly access the complete data set, including DNA sequences and the origin of the specimen. The databases are accessible to end users through an online application (http://anchovyid.jrc.ec.europa.eu), which provides tools to perform basic homology searches with query sequences and to extract data. Additional links to phylogenetic analysis programs are supplied for more complex investigations. The content of AnchovyID is fully documented by downloadable reports and protocols.

ACKNOWLEDGMENT

We are grateful to Sébastien Lavoué (University of Tokyo) who provided us with precious advice and some samples; Patrice Pruvost, Philippe Béarez, and Samuel Iglésias (MNHN Paris) for their great help in acquiring standard fish specimens and Sandrine Vaz, Marc Taquet, Christian Dintheer, and Justine Dumay (IFREMER) for collecting samples; Véronique Mullack (CEVPM) and Jean Pascal Bergé (IFREMER) for processed samples; and Laetitia Kolypczuk and Monique Etienne (IFREMER) for their commercial legislation comments.

LITERATURE CITED

(1) FAO Fisheries and Aquaculture Department. The State of World Fisheries and Aquaculture, 2007.

(2) Commission Regulation (EC) No. 2065/2001. Off. J. Eur. Communities 2001, 278.

(3) European Parliament and Council of 28 January 2002, Regulation (EC) No. 178/2002. Off. J. Eur. Communities 2002, *31*, 1.

(4) Economic Assessment of European Fisheries: Economic Performance of Selected European Fishing Fleets, Annual Report 2005. European Commission, 2006.

(5) Klanjsčěk, J.; Legović, T. Is anchovy (*Engraulis encrasicolus*, L.) overfished in the Adriatic Sea. *Ecol. Modell.* **2007**, *201*, 312–316.

(6) Commission Regulation (EC) No. 1116/2006 of 20 July 2006 prohibiting fishing for anchovy in ICES sub-area VIII. Off. J. Eur. Union Ser. L: Legis. 2006, 199.

(7) Sebastio, P.; Zanelli, P.; Neri, T. M. Identification of anchovy (*Engraulis encrasicholus*, L.) and gilt sardine (*Sardinella aurita*) by polymerase chain reaction, sequence of their mitochondrial cytochrome *b* gene, and restriction analysis of polymerase chain reaction products in semipreserves. *J. Agric. Food Chem.* **2001**, *49*, 1194–1199.

(8) Magoulas, A.; Castilho, R.; Caetano, S.; Marcato, S.; Patarnello, T. Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Mol. Biol. Evol.* **2006**, *39*, 734–746.

(9) Yu, Z.; Kong, X.; Guo, T.; Jiang, Y.; Zhuang, Z.; Jin, X. Mitochondrial DNA sequence variation of Japanese anchovy *Engraulis japonicus* from the Yellow Sea and East China Sea. *Fish. Sci.* **2005**, *71*, 299–307.

(10) Grant, W. S.; Leslie, R. W.; Bowen, B. W. Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *J. Fish Biol.* **2005**, *67*, 1242–1265.

(11) Grant, W. S. A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): Assessing models of population structure and the Black Sea isolation hypothesis. *Genetica* **2005**, *125*, 293–309.

(12) Quinteiro, J.; Vidal, R.; Izquierdo, M.; Sotelo, C. G.; Chapela, M. J.; Perez-Martin, R.; Rehbein, H.; Hold, G. L.; Russell, V. J.; Pryde, S. E.; Rosa, C.; Santos, A. T.; Rey-Meñdez, M. Identification of hake species (*Merluccius* genus) using sequencing and PCR-RFLP analysis of mitochondrial DNA control region sequences. *J. Agric. Food Chem.* **2001**, *49*, 5108–5114.

(13) Quinteiro, J.; Sotelo, C. G.; Rehbein, H.; Pryde, S. E.; Medina, I.; Pérez-Martin, R. I.; Rey-Méndez, 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.

(14) Carrera, E.; Garcia, T.; Céspedes, A.; González, I.; Frenández, A.; Hernández, P. E.; Martin, R. Salmon and trout analysis by PCR-RFLP for identity authentication. *J. Food Sci.* **1999**, *64*, 410–413.

(15) Ram, J. L.; Ram, M. L.; Baidoun, F. 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.

(16) Takashima, Y.; Morita, T.; Yamashita, M. Complete mitochondrial DNA sequence of Atlantic horse mackerel *Trachurus trachurus* and molecular identification of two commercially important species *T. trachurus* and *T. japonicus* using PCR-RFLP. *Fish. Sci.* **2006**, *72*, 1054–1065.

(17) Infante, C.; Crespo, A.; Zuasti, E.; Ponce, M.; Perez, L.; Funes, V.; Catanese, G.; Manchado, M. PCR-based methodology for the authentication of the Atlantic mackerel *Scomber scombrus* in commercial canned products. *Food Res. Int.* **2006**, *39*, 1023–1028.

(18) Trotta, M.; Scho-Nhuth, S.; Pepe, T.; Cortesi, M. L.; Puyet, A.; Bautista, J. M. Multiplex PCR method for use in real-time PCR for identification of fish fillets from grouper (*Epinephelus* and *Mycteroperca* species) and common substitute species. *J. Agric. Food Chem.* **2005**, *53*, 2039–2045.

(19) Bartlett, S. E.; Davidson, W. S. Identification of *Thunnus* 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.

(20) Bartlett, S. E.; Davidson, W. S. FINS (forensically informative nucleotide sequencing): A procedure for identifying the animal origin of biological specimens. *BioTechniques* **1992**, *12*, 408–411.

(21) Davidson, W. S. *Analytical Methods of Food Authentication;* Blackie Academic & Professional: Glasgow, Scotland, 1998; pp 182−203.

(22) Chapela, M. J.; Sotelo, C. G.; Calo-Mata, P.; Perez-Martin, R. I.; Rehbein, H.; Hold, G. L.; Quinteiro, J.; Rey-Mendez, M.; Rosa, C.; Santos, A. T. Identification of cephalopod species (Ommastrephidae and Loliginidae) in seafood products by forensically informative nucleotide sequencing (FINS). *J. Food Sci.* **2002**, *67*, 1672–1676.

(23) Terol, J.; Mascarell, R.; Fernandez-Pedrosa, V.; Pérez-Alonso, M. Statistical validation of the identification of tuna species: Bootstrap analysis of mitochondrial DNA sequences. *J. Agric. Food Chem.* **2002**, *50*, 963–969.

(24) Jérôme, M.; Lemaire, C.; Bautista, J. M.; Fleurence, J.; Etienne, M. Molecular phylogeny and species identification of sardines. *J. Agric. Food Chem.* **2003**, *51*, 43–50.

(25) Jérôme, M.; Lemaire, C.; Verrez-Bagnis, V.; Etienne, M. Direct sequencing method for species identification of canned sardine and sardine-type products. *J. Agric. Food Chem.* **2003**, *51*, 7326–7332.

(26) Pepe, T.; Trotta, M.; Di Marco, I.; Anastasio, A.; Bautista, J. M.; Cortesi, M. L. Fish species identification in surimi-based products. *J. Agric. Food Chem.* **2007**, *55*, 3681–3685.

(27) Akasaki, T.; Saruwatari, T.; Tomonaga, H.; Sato, S.; Watanabe, Y. Identification of imported chirimen at the genus level by a direct sequencing method using mitochondrial partial 16S rDNA region. *Fish. Sci.* **2006**, *72*, 686–692.

(28) Miya, M.; Nishida, M. Use of mitogenomic information in teleostean molecular phylogenetics: A tree-based exploration under the maximum-parsimony optimality criterion. *Mol. Phylogenet. Evol.* **2000**, *17*, 437–455.

(29) Folmer, O.; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse matazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **1994**, *3*, 294–299.

(30) Sevilla, R. G.; Diez, A.; Norén, M.; Mouchel, O.; Jérôme, M.; Verrez-Bagnis, V.; van Pelt, H.; Favre-rey, L.; Krey, G.; Bautista, J. M. Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome *b* and nuclear rhodopsin genes. *Mol. Ecol. Notes* **2007**, *7*, 730–734.

(31) Hall, T. A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 9598

(32) Kumar, S.; Tamura, K.; Nei, M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinf.* **2004**, *5*, 150–163.

(33) 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.

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

Received for review December 19, 2007. Revised manuscript received March 5, 2008. Accepted March 14, 2008.

JF703704M