JOURNAL OF
AGRICULTURAL AND
FOOD CHEMISTRY

Molecular Phylogeny and Species Identification of Sardines

Marc Jérôme,*,[†] Christophe Lemaire,[†] José M. Bautista,[‡]
Joël Fleurence,[†] and Monique Etienne[†]

Laboratoire Biochimie des Protéines & Qualité, IFREMER, Rue de l'île d'Yeu, BP 21105, F44037
Nantes Cedex 03, France, and Department Bioquimica y Biologia Molecular IV, Universidad
Complutense de Madrid, Ciudad Universitaria, E-28040 Madrid, Spain

The DNA sequence diversity of *Sardina pilchardus* (Walbaum, 1792) and some closely related species of *Clupeomorpha* was investigated using the mitochondrial DNA gene encoding cytochrome *b*. The nucleotide sequences of complete and partial mtDNA cytochrome *b* were determined in numerous specimens. Sequence divergence between species and genera was evenly distributed in the cytochrome *b* gene but rather high compared to reports for other fish species. Phylogenetic analyses on complete cytochrome *b* were used to study the relationships among the considered species. *S. pilchardus* was easily differentiated, showing a genetic distance of 0.25 with respect to *Clupeidae* species and 0.26 with respect to the other species. A species-specific short fragment (<150 bp) was isolated by polymerase chain reaction (PCR) using primers designed for *Clupeomorpha*. A rapid and reliable PCR method using restriction fragment length polymorphism (RFLP) with two restriction enzymes (MnlI/HinfI) was optimized for unambiguous differentiation of *S. pilchardus* from the other species tested (raw and canned products).

KEYWORDS: Species identification; sardine; *Clupeidae*; cytochrome *b* gene; genetic divergence; PCR−RFLP

INTRODUCTION

Species identification of sardine type products is of increasing importance in Europe. European Union regulations stipulate that canned sardines must be prepared exclusively from *Sardina pilchardus* (Walbaum, 1792), whereas the *Codex Alimentarius* standard provides a positive list of 21 small pelagic fishes (**Table 1**) that can be used in the preparation of canned sardine or sardine type products. The name of the product written on the label must be "sardines" when *S. pilchardus* is used exclusively. When other species are used, the name "sardines" must be completed by a distinctive designation, i.e. the name of a country, a geographic area, the species, or the common name of the fish in accordance with the laws and customs of the country in which the products are sold.

As most of the external features allowing morphological identification of whole fish are not apparent after processing, there is a need for analytical methods of authentication to detect mislabeling. The traditional methods of fish species identification, such as isoelectric focusing (IEF) of proteins (*1−3*) or HPLC (*4, 5*), are applicable only to raw fish. Other electrophoretic methods, such as SDS−PAGE or urea IEF (*6−11*), can be used to identify processed fish (cooked, smoked, prefried, breaded). However, these techniques cannot be used when processing highly denatures proteins, especially in canning. As

Table 1. List of the 21 Small Pelagic Fish Authorized by the *Codex Alimentarius* for Preparation of Canned Sardines or Sardine Type Products

	common name	scientific name	world production (tons)[a]
sardine	European pilchard	*Sardina pilchardus*	940 727
sardine type product	South American pilchard	*Sardinops sagax*[b]	937 269
	Californian pilchard	*Sardinops caeruleus*[b]	365 784
	Japanese pilchard	*Sardinops melanostictus*[b]	295 788
	Southern African pilchard	*Sardinops ocellatus*[b]	196 534
	Australian pilchard	*Sardinops neopilchardus*[b]	nonindexed
	round sardinella	*Sardinella aurita*	633 578
	Indian oil sardine	*Sardinella longiceps*	282 075
	goldstripe sardinella	*Sardinella gibbosa*	161 200
	Madeiran sardinella	*Sardinella maderensis*	123 674
	Brazilian sardinella	*Sardinella brasiliensis*	82 283
	European sprat	*Sprattus sprattus*	696 243
	sandy sprat	*Hyperlophus vittatus*	nonindexed
	anchoveta	*Engraulis ringens*	1 729 064
	Argentine anchoita	*Engraulis anchoita*	13 417
	Californian anchoita	*Engraulis mordax*	2335
	Atlantic herring	*Clupea harengus*	2 419 117
	round herring	*Etrumeus teres*	58 569
	Atlantic thread herring	*Opisthonema oglinum*	18 752
	Pacific menhaden	*Ethmidium maculatum*	40 845
	Western Australian gizzard shad	*Nematalosa vlaminghi*	nonindexed

[a] World production (FAO, 1998) is indicated in tons for each species. [b] The traditional species named *Sardinops* spp. indicated are grouped together under a single designation, *Sardinops sagax*.

an alternative to protein analysis, DNA-based techniques provide a new authentication tool (*12, 13*). As DNA is more heat stable

* Corresponding author. Phone: 33 2 40 37 40 00. Fax: 33 2 40 37 40 71. E-mail: Marc.Jerome@ifremer.fr.
† IFREMER.
‡ Universidad Complutense de Madrid.

44 *J. Agric. Food Chem.,* Vol. 51, No. 1, 2003

Jérôme et al.

than many proteins, analyses involving nucleic acid are more likely to be used on processed fish. Moreover, DNA can provide more information than protein, due to the degeneracy of the genetic code and the presence of a majority of noncoding regions rich in useful polymorphism for species evolution studies.

Developments in molecular biology allow the gene sequence from a large number of individuals to be determined in a short period of time (*14*). The increased use of nucleotide sequences in polymerase chain reaction (PCR) and of direct sequencing for determining phylogenetic relationships in vertebrates has made databases available that are well suited for comparative studies. Segments of the mitochondrial DNA (mtDNA) genome have proved to be particularly useful for identification of the animal origin of various meat or fish preparations (*15*, *16*). Mitochondrial DNA has several advantages over nuclear DNA for diagnostic studies of fish products, particularly a greater abundance in nuclear acid extracts and a copy number several times higher (*17*). As mitochondrial DNA is maternally inherited, haploid, sequence ambiguities from heterozygous genotypes are theoretically avoided (*13*). The cytochrome *b* gene, which has been targeted for many different studies, has a relatively high mutation rate and sufficient point mutations to allow discrimination of even closely related species of small pelagic fishes (*18*, *19*). However, mitochondrial DNA also exhibits a degree of intraspecific variability useful for population identification in pelagic fishes (*20, 21*–*23*).

Current knowledge of the systematics and biogeography of the widely distributed species *Sardina pilchardus, Sardinops* spp., *Sardinella* spp., and *Engraulis* spp. relies mainly on a few works (*20, 21, 24*–*30*). Sardines (*Sardina, Sardinops*) and anchovies are distributed in temperate zones worldwide and have representative species or populations in most of the temperate boundary current systems (*21*). Sardinella are found in subtropical and tropical zones and have representative species or populations in most of the subtropical and tropical boundary current systems. A previous report concerning the genetic divergence between *S. pilchardus* and *Sardinops* spp. (*21*) found a Kimura's genetic distance (*31*) of 0.23 on a fragment of cytochrome *b*. Another study using allozymes also reported a great divergence (*32*). In all cases, divergence time between *Sardina* and *Sardinops* genera was greater than 11 million years, corresponding to the closure of the Tethys Sea in the Miocene.

The purpose of the present work was 2-fold. First, the genetic divergence between commercially important Clupeiformes species was studied on the entire cytochrome *b* gene sequence to determine whether the genetic distance of *S. pilchardus* from other *Clupeidae* (Sardinops, Sardinella, Sprat, and Herring) is great enough to warrant distinct labeling. Second, relevant tools were developed for discrimination of sardine type products by amplifying the DNA fragment of cytochrome *b* suitable for obtaining species-specific sequences. The selected region and size of the PCR product had to provide sufficient information for unambiguous diagnosis and allow the use of a PCR−RFLP approach for simple, rapid differentiation of *S. pilchardus* from the other species.

MATERIALS AND METHODS

Sampling. For the complete mitochondrial cytochrome *b* gene study and restriction analysis, all specimens obtained and identified by collectors (**Table 2**) were frozen, except *Sardinops caeruleus*, which was sampled in ethanol (80%). Two additional cytochrome *b* sequences derived from a data bank were used for consideration of nucleotide divergences on complete cytochrome *b: Sardinops melanostictus* (Genbank accession number AB032554) and *Engraulis japonicus* (AB040676). Canned sardine type products were purchased at local

Table 2. Sampling for Mitochondrial Cytochrome *b* Analysis

species	location	date collcd	collectors[a]
Sardinella aurita	Ivory Coast	Oct 2000	Xavier Bard, IRD, Ivory Coast
Sardinella maderensis	Ivory Coast	Oct 2000	Xavier Bard, IRD, Ivory Coast
Sardina pilchardus	Bay of Biscay	May 2000	Ifremer (Thalassa ship), France
Sardinops sagax	Chile coast	Sep 1996	IFOP, Chile
Sardinops caeruleus	Pacific	Mar 2000	Héctor Villalobos Ortiz, IPN, Mexico
Engraulis encrasicolus	Bay of Biscay	May 2000	Ifremer (Thalassa ship), France
Clupea harengus	North Sea	Feb 1997	Ifremer (Thalassa ship), France
Sprattus sprattus	Bay of Biscay	May 2000	Ifremer (Thalassa ship), France

[a] IRD: Institut de Recherche pour le Développement. IFOP: Instituto de Fomento Pesquero. IPN: Instituto Poltécnico Nacional.

Table 3. Primer Sequences for PCR Amplification of Cytochrome *b* Gene

no.	name	sequence 5′3′	ref
1	Fishcytb-F	ACCACCGTTGTTATTCAACTACAAGAAC	Bautista (unpublished)
2	Cytb13-R	GGGGTAAAGTTGTCTGGGTCTCC	Bautista (unpublished)
3	Cytb11-F	CGATTCTTCGCATTCCACTTCCT	Bautista (unpublished)
4	Trucytb-R	CCGACTTCCGGATTACAAGACCG	Bautista (unpublished)
5	FishcytbMOF	TGACTCGAAAAGCCACGgTTGTTATTCAAC	this study
6	TrucytbMOR	CGGTCTTGTAATCCGAAGACCCG	this study

supermarkets or supplied by collaborators. Samples 1−11 were traditional French or Moroccan canned sardines (*Sardina pilchardus* labeled "sardines") prepared in vegetable oil (groundnut or olive oil) or tomato sauce, sometimes with seasoning such as lemon or red chili. Sample 12 (labeled "sardinen filets") was canned in Peru, sample 13 was a Japanese canned product, sample 14 was a Venezuelan product (labeled "sardinen filets"), sample 15 was a French can of anchovies (labeled "filets d'anchois avec olives"), and sample 16 was a can of Brisling sardines from Norway (probably prepared with *Sprattus sprattus*).

DNA Extraction. Total genomic DNA extraction was performed according to a previously described procedure (*33*). DNA was precipitated with absolute ethanol, and the pellet was resuspended in Tris-EDTA buffer and stored at 4 °C until PCR. The phenol/chloroform/isoamyl alcohol method (PCI) was applied for preparation of the DNA required for mitochondrial cytochrome *b* gene amplification. For our reference samples and sardine type products, a Chelex (Biorad, Hercules, CA) method was tested concurrently because of its simplicity and rapidity. A small muscle section was placed on filter paper to remove oil and liquid. The dried muscle was then vortexed in 300 μL of a 5% Chelex water solution (Chelex 100 resin) with 20 μL of proteinase K (Qiagen, 20 mg/mL) and 30 μL TE buffer (10 mM Tris HCL, pH 8.0; 1 mM EDTA). The mixture was incubated at 56 °C for 4 h to break down all tissue and then heated at 100 °C for at least 15 min to denature and precipitate resin-bound protein. The DNA suspension was stored at 4 °C until use for PCR amplification.

Mitochondrial Cytochrome *b* Gene Fragment Amplification. PCR amplifications were done using Hybaid PCR Express (Hybaid, Ashford, U.K.). An initial step (95 °C, 15 min) was performed to activate the HotStarTaq polymerase (Qiagen GmbH, Hilden, Germany). Cycling conditions (35 cycles) were 95 °C for 40 S, 52 °C for 40 S, and 72 °C for 40 S, followed by a final extension for 10 min at 72 °C. Reactions were carried out in 100 μL of reaction volume: 1 μL of template DNA was added to 99 μL of PCR mix consisting of 84.5 μL of sterile distilled water, 10 μL of 10× PCR buffer containing 1.5 mM $MgCl_2$ and 2 μL of dNTP mix (10 mM of each) plus 1 μL of each primer (0.5 μM) and 2.5 units of HotstarTaq DNA polymerase.

A primer set originally developed for *Oncorhynchus mykiss* (Bautista, unpublished) was used for amplification in two overlapping fragments of mitochondrial cytochrome *b* gene: 5′-half (**Table 3**, nos. 1 and 2) and 3′-half (**Table 3**, nos. 3 and 4). Modified primers from fishcytb and trucytb [fishcytbmo, trucytbmo (**Table 3**, nos. 5 and 6)] were designed for this study and used when necessary for some samples.

Cleanup, Electrophoresis, and Sequencing of PCR Products. The 1.5% agarose gels (Agarose Na, Amersham Pharmacia Biotech,

Molecular Phylogeny of Sardines

J. Agric. Food Chem., Vol. 51, No. 1, 2003 **45**

Freiburg, Germany) were employed to check DNA extractions and PCR products, using TAE buffer (2 mM EDTA, 40 mM Tris acetate, pH 8.5) with ethidium bromide for band characterization via ultraviolet transillumination (Image Master VDS-CL, Amersham Pharmacia Biotech, Freiburg, Germany).

Prior to sequencing, double-stranded PCR products were purified by filtration through an Qiagen QIAquick column according to the manufacturer's protocol. PCR fragments were used for direct cycle sequencing with the ABI Big Dye Terminator cycle sequencing kit. Sequencing reactions were performed (by ESGS France) on an ABI PRISM3100 DNA sequencer (Applied Biosystems, Foster City, CA) in both directions with the primers used for PCR amplification.

PCR−RFLP of Short PCR Products. Strong degradations of the DNA due to canning or boiling processes do not permit one to amplify PCR fragments larger than 300 bp (*54*). Thus, the previously determined sequences of cytochrome *b* gene allowed us to design two primer sets solely for amplification of short species-distinctive sequences of the 5′ portion of cytochrome *b* gene. Two short overlapping variable regions were selected. Each region was flanked by two conserved zones allowing the design of primers for amplification of the target species. The first primer set (C-CB285dF, CGCCCACATTGGNCGAGG; C-CB431R, GTGGCCCCTCAGAAGGACATTTGGCC) produced a 147 bp fragment, and the second (C-CB284dF, AYGCNCACATTG-GNCGRGG; C-CB425dR, CCTCAGAADGACATTTGBCCTCA) generated a 142 bp PCR product. Restriction analysis was performed concurrently on these two DNA fragments.

Calculation of the corresponding RFLP patterns of investigated species with suitable restriction enzymes was done using the "Restriction" program (Infobiogen, www.infobiogen.fr).

The restriction endonucleases selected, MnlI and HinfI, were used to digest the short amplified product of sample from *Sardina pilchardus*, *Sardinella aurita*, *Sardinella maderensis*, *Sardinops sagax, Sardinops caeruleus*, *Engraulis encrasicolus*, and *Sprattus sprattus* (142 bp fragment) and from *Clupea harengus* (142 bp fragment). For each sample, 5−10 μL of the PCR reaction containing amplified DNA was digested overnight at 37 °C with 2 U of each endonuclease in a final volume of 20 μL.

Restriction fragments were separated by electrophoresis on Gene-GelClean15−24 (Amersham Pharmacia Biotech, Uppsala, Sweden) and visualized using the Pharmacia DNA silver-staining kit. Their size was determined in comparison with MassRuler low-range DNA ladders (Fermentas AB, Vilnius, Lithuania) and DNA marker 50−1000 bp (Bio-Whittaker, Vallensbaek, Denmark).

Phylogenetic Analysis. All alignments were performed on BioEdit software (*34*), and phylogenetic treatments were computed on MEGA 2.0 software (*35*). The scale of comparison between genera suggested long divergence times. As saturation effects due to transitional substitutions might affect the estimation of nucleotide distances (*36*), it was relevant to test whether saturation occurred or not for such divergences in our dataset. Thus, the mean transition/transversion ratio (TS:TV) was estimated over all the sequences and compared with the expected ratio at a saturation determined from observed frequencies by the formula of Holmquist (*37*): $R_{S:V} = (P_A P_G + P_C P_T):(P_A + P_G)(P_C + P_T)$, where P_A, P_G, P_C, and P_T are the observed base frequencies.

Nucleotide divergences were computed using the Tamura−Nei model (*38*), which takes substitutional rate biases and the inequality of base frequencies into account. Tree topologies were evaluated by three methods: neighbor-joining (*39*), as inferred from the distances matrix; parsimony (*40*), as implemented by MEGA 2.0; maximum likelihood, as implemented by DNAML, a PHYLIP program (*41*). Standard error of divergence estimates and the robustness of topology nodes were all tested by the bootstrap method.

RESULTS

Complete Cytochrome *b* Gene. Fragments of 1221 bp, including the complete cytochrome *b* gene of seven Clupeiformes species, were amplified and aligned against the sequences of *Sardinops melanostictus* and *Engraulis japonicus*, using BioEdit (*34*). The complete cytochrome *b* sequences of *Sardina pilchardus*, *Sardinops caeruleus, Sardinella maderensis, Sar-*

Table 4. Description of Haplotypes

species	authors	accession nos.
Sardinops melanostictus	Inoue et al., 2000	AB032554
Sardinops sagax	this paper	AF472586
Sardinops caeruleus	this paper	AF472585
Sardina pilchardus	this paper	AF472582
Sardinella maderensis	this paper	AF472583
Sardinella aurita	this paper	AF472584
Engraulis encrasicolus	this paper	AF472579
Engraulis japonicus	Inoue et al., 2001	AB040676
Clupea harengus	this paper	AF472580
Sprattus sprattus	this paper	AF472581

dinella aurita, Engraulis encrasicolus, Clupea harengus, and *Sprattus sprattus* were deposited in the GenBank (**Table 4**). Among the 1141 bp of the mitochondrial cytochrome *b* gene, 430 evenly distributed variable sites were recorded, 353 of which were phylogenetically informative. A bias in base composition was observed, as pyrimidines occurred more frequently than purine ($\chi^2 = 33.47$; ddl = 3, $p = 2.56 \times 10^{-7}$). The mean frequencies of thymine and cytosine were respectively 29 ± 0.63% and 29 ± 0.73% when adenosine and guanine occurred respectively at 23 ± 1.01% and 18.8 ± 0.94%. **Table 5** indicates the number of transition and transversion changes between species. The corrected TS:TV rates with the Tamura−Nei model ranged from 1.10 ± 0.19 (*E. japonicus*/*S. maderensis*) to 28.35 ± 12.92 (*S. melanostictus*/*S. caeruleus*), which constitutes an outlier of the distribution of TS:TV rates. Without this latter comparison, the maximum TS:TV rate dropped to 3.80 ± 0.86 (*C. harengus*/*S. sprattus*). Mean TS:TV for the 36 comparisons was 3.14 ± 0.54 (2.53 ± 0.51 without *S. caeruleus*), i.e., greater than the 0.52 value of the expected TS:TV rate at saturation.

Nucleotide distances were computed according to the Tamura−Nei model (**Table 5**). Distances over genera exceeded 0.22, except for the genetic divergence between *C. harengus* and *S. sprattus* ($d = 0.13$). The genetic distance measured between *S. pilchardus* and the two *Sardinops* species was 0.22. Surprisingly, nucleotide divergence between the two *Sardinella* species ($d = 0.21$) was of the same order as that between genera ($\bar{d} = 0.25 ± 0.006$) when divergence was 0.04 between *Engraulis* species and 0.02 between *Sardinops* species. With a mean genetic distance of 0.26 ± 0.018, *S. pilchardus* was easily differentiated from the *Engraulidae* species studied as well as from the other C*lupeidae* species (0.25 ± 0.02). The nucleotide divergence between *Clupeidae* and *Engraulidae* families was 0.25 ± 0.01.
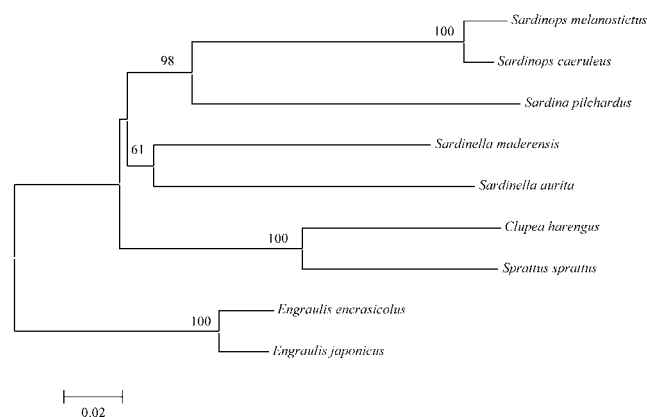
The neighbor-joining tree in **Figure 1** clearly shows the phylogenetic relations between the nine species, and the results were the same for parsimonious and maximum likelihood trees (data not shown). *E. encrasicolus* and *E. japonicus* are considered to be a sister group. Bootstrap values at the nodes are around 100%, except for the *Sardinella* genus in which both species are clustered with a 61% score. Although *S. pilchardus* and *Sardinops* belong to the same group, the former was clearly differentiated.

Diagnosis by Restriction on 147/142bp Fragments. In some cases unclear restriction patterns related to poor-quality PCR products was observed. With the C-CB285dF/C-CB431R primer set, 147 bp fragments with a high amplification yield were obtained for species of the genera *Sardina*, *Sardinops*, *Sardinella*, and *Engraulis*. On the contrary the yield was low, in particular from some canned sample where the DNA was probably highly degraded or when the PCR failed. This was especially true for species of the genera *Clupea* and *Sprattus*.

Table 5. Nucleotide Divergences between Pairs of Haplotypes (Complete Cytochrome *b* Gene) for Clupeiformes Fishes[a]

	S. pilchardus	S. melanostictus	S. caeruleus	C. harengus	S. sprattus	S. maderensis	S. aurita	E. encrasicolus	E. japonicus
S. pilchardus	–	146/65	144/66	167/88	152/96	133/92	149/94	133/115	134/115
S. melanostictus	0.22	–	27/1	147/95	158/ 95	139/79	144/ 99	127/114	134/114
S. caeruleus	0.22	0.02	–	138/96	151/96	140/78	141/98	129/113	130/113
C. harengus	0.28	0.26	0.25	–	104/34	138/94	144/98	139/105	131/105
S. sprattus	0.27	0.28	0.27	0.13	–	137/100	132/94	137/109	133/109
S. maderensis	0.24	0.23	0.23	0.25	0.26	–	112/90	110/115	105/113
S. aurita	0.26	0.27	0.26	0.26	0.24	0.21	–	124/113	123/115
E. encrasicolus	0.28	0.27	0.27	0.27	0.27	0.25	0.26	–	37/2
E. japonicus	0.28	0.28	0.27	0.26	0.27	0.24	0.26	0.04	–

[a] Above diagonal: number of transition/number of transversion. Below diagonal: Tamura–Nei distances.


Figure 1. Neighbor-joining tree inferred from Tamura–Nei distances between sequences of cytochrome *b* gene in nine species of *Clupeiformes.* Bootstrap scores greater than 50% are shown.

Table 6. Expected Size of the Short 147/142 Bp Fragments after Digestion by the Enzyme Pair MnlI–HinfI[a]

species	fragments
Sardina pilchardus	8*/16*/18/33/72
Sardinella aurita	8*/15/16*/18/18/33/39
Sardinella maderensis	8*/16*/18/105
Sardinops sagax	8*/10/16*/18/29/33/33
Sardinops caruleus	8*/10/16*/18/29/33/33
Sardinops melanostictus	8*/10/16*/18/33/62
Engraulis encrasicolus	8*/16*/18/105
Sprattus sprattus	8*/10/16*/18/27/68
Clupea harengus	8*/16*/18/105

[a] When digestion was applied to the 142 bp fragment, the two bands shown by an asterisk were slightly modified: the 8 bp band changed to a 9 bp band, and the 16 bp band changed to a 10 bp band.

Thus, C-CB285dF/C-CB431R was then slightly modified (degenerating into C-CB284dF/C-CB425dR) to allow extension of PCR amplification of a 142 bp sequence to all tested species, improve PCR amplification of some samples, and provide similar yields. Nevertheless, because of its high specificity and yield on *S. pilchardus*, the first primer set was retained, and restriction analysis was performed concurrently on the two DNA fragments of 147 and 142 bp.

Table 6 shows the predictablerestriction fragments for the eight reference species analyzed by restriction from a 147 bp or a 142 bp PCR product. The profiles generated, from 147 bp or 142 bp amplicon, show no significant differences but only slight modifications for minor fragments. In fact, the 16 bp and 8 bp restriction products common to all species tested on the 147 bp fragment were modified respectively to 10 and 9 bp bands with the 142 bp fragment.
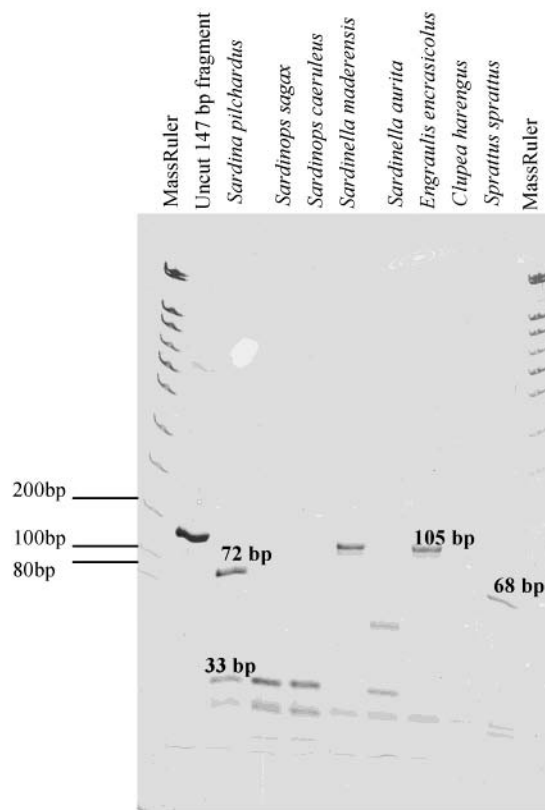
Figure 2. RFLP patterns of eight species belonging to the Clupeidae and Engraulidae families. The 142 bp PCR fragments were digested by MnLI/HinfI, separated by electrophoresis on GenegelClean 15/24, and silver-stained. The size (bp) of the main restriction fragment is indicated above the band.

The RFLP profiles obtained from reference samples are shown in **Figure 2**. As the smallest fragments (<18 bp) gave poor intensity or were not well discriminated in the gel, they were not used for the interpretation of results. Predictably, restriction by the MnlI−HinfI pair on *S. pilchardus* produced a typical profile with two main bands at 72 bp and 33 pb. Restriction profiles of *S. sagax* and *S. caeruleus* were identical, consisting of a main band composed of three fragments (two 33 bp and one 29 bp not discriminated in the gel) and minor bands composed of the smallest fragments. Theoretical restriction suggests that *S. melanostictus* would give products at 62 bp and 33 bp. *S. aurita* showed a characteristic pattern, with the largest products at 39 bp and 33 bp plus a minor band 18 bp. *S. maderensis*, *E. encrasicolus*, and *C. harengus* displayedthe same restriction pattern: a main band at 105 bp and a minor band at 18 bp. *S. sprattus* also had a unique profile, with a main band at 68 bp and no band of 33 bp.
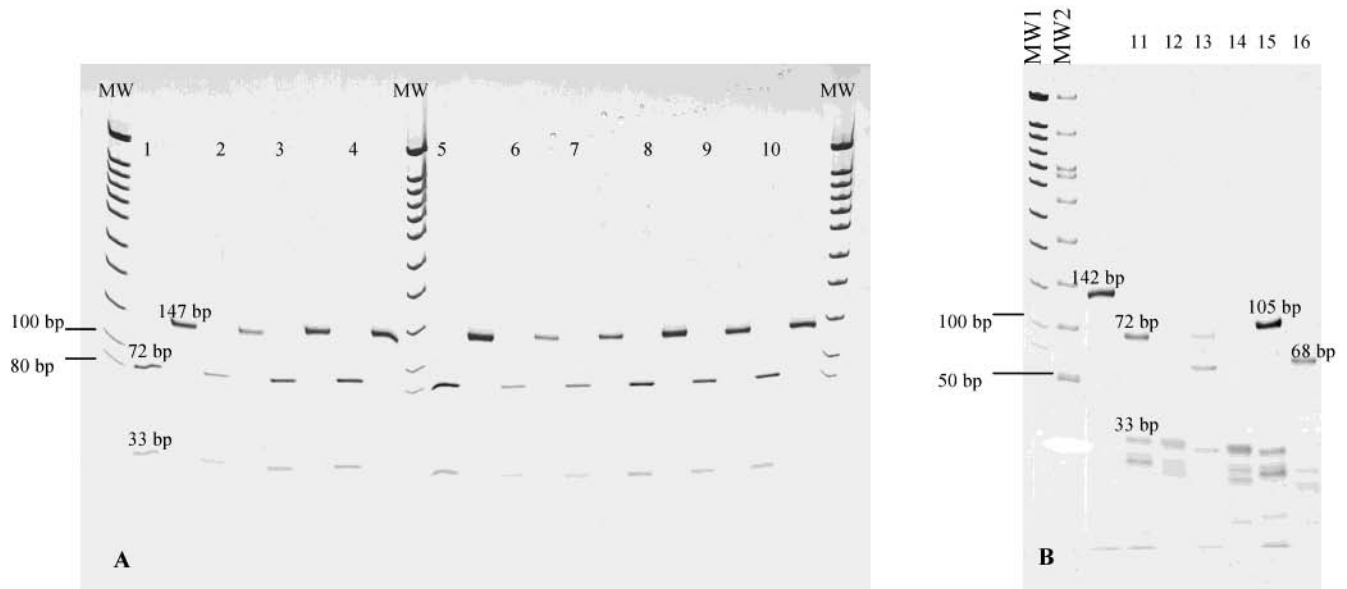
Molecular Phylogeny of Sardines

J. Agric. Food Chem., Vol. 51, No. 1, 2003 **47**

Figure 3. Restriction patterns of 16 commercial canned products after digestion of PCR product with restriction enzymes MnlI and Hinfl. (A) 147 bp PCR product obtained with 10 European and Moroccan sardine type products labeled "sardines": (from left to right) MassRuler low-range ladder (MW), samples 1–4, MW; samples 5–10, MW. (B) 142 bp PCR product obtained with six sardine type products of different origin: (from left to right) MassRuler low-range ladder (MW1), DNA marker 50–1000 bp (MW2); 11, sardine type products labeled "sardines"; 12, Peruvian can; 13, Japanese can; 14, Venezuelan can; 15, French canned product labeled as anchovy filets; 16, canned product from Norway.

Figure 3A,B shows the restriction profiles obtained from 16 commercial sardine type products. A total of 10 of the 16 DNA extractions were performed successfully with the Chelex method. The 10 commercial canned sardines of different origin (samples 1–10) labeled "sardines" were analyzed by the PCR– RFLP method using 147 bp sequences (**Figure 3A**). The same characteristic 72/33 bp profile of *S. pilchardus* was observed unambiguously for each sample, regardless of origin (Mediterranean Sea and Atlantic Ocean from Morocco to France). **Figure 3B** shows the restriction patterns obtained using the MnLI/HinfI pair on 142 bp PCR products of six other commercial sardine type products. Sample 11, a French sardine type product labeled "sardines", gave the typical restriction pattern of *S. pilchardus* and was deposited on gel for comparison with the other five samples (12–16) not labeled as canned sardine. Samples 12 and 14, Peruvian and Venezuelan canned products. respectively, showed bands similar to those for the *S. sagax* or *S. caeruleus* references. Sample 13, a Japanese canned product, displayed the theoretical characteristic profile of *S. melanostictus*, but an unexpected small 72 bp fragment appeared that may have been an intermediate digestion product. MnLI can generate a 72 bp fragment that would be cut into two 62 and 10 bp fragments by HinfI. Sample 15, a French canned product labeled "anchovy filets", displayed the same profile as the *E. encrasicolus* reference, with a strong band at 105 bp. As expected, sample 16, Brisling sardines from Norway, showed a profile similar to that of the *S. sprattus* reference.

DISCUSSION

The purpose of this work was to study the genetic divergence between Clupeiformes species labeled as sardine type products and to define relevant tools to discriminate these pelagic fishes. Accordingly, cytochrome *b* was chosen for its high potentiality to discriminate species and intrafamilial genera (*15*). The results show good distinctions between the main genera on different scales of cytochrome *b* length as well as the power of the PCR– RFLP method to differentiate *Sardina pilchardus* from some

other *Clupeomorpha* genera, particularly *Sardinops*, even when these fishes are processed.

Biases in Base Composition and Homoplasy within *Clupeomorpha*. The study of base frequencies among the nine sequences of the complete cytochrome *b* gene showed a significant bias in favor of thymine and cytosine and a guanine deficit. This kind of disequilibrium has been reported in many studies on cytchrome *b* in fishes or other vertebrates (*42*–*45*). The excess of pyrimidines at the second codon position is usually attributed to the hydrophobic nature of the protein (*42*, *43*). Anti-G bias and excess A and C at the third position are also common features of the cytochrome *b* gene, in association with codon usage bias caused by selection for translation efficiency (*46*).

In phylogenetic studies, homoplasy can obscure topologies. However, no evidence of homoplasy was found in our comparisons of species for the complete cytochrome *b* gene. Even if homoplasy existed, this would not have changed the topology, as the same results were observed for different methods of phylogenetic reconstruction and for any codon position.

Phylogenetic Relations within *Clupeomorpha*. As almost all of the species compared in this study were from different genera, divergences on the cytochrome *b* gene would conceivably allow easy discrimination. In fact, nucleotide divergences among the different sets of sequences generally exceeded 20%. In a previous study comparing many divergences on the cytochrome *b* gene (*47*), distances between fish confamilial genera rarely exceeded 20%. As Clupeiformes are regarded as an ancient taxon giving rise to the *Teleostei*, high divergences between genera are likely. Grant et al. (*21*) observed a Kimura distance of 23.3% between *Sardina* and *Sardinops*, which is consistent with the mean distance of 22% found between haplotypes of both genera on the complete cytochrome *b* gene. These two genera are considered to constitute a phylogenetic unit (*24*, *32*, *48*, *49*). The observations of Grant et al. (*24*) tend to confirm the hypothesis of Svetovidov (*50*) and Parrish (*26*) that *Sardina* was isolated from *Sardinops* after the closure of the Tethys Sea about 18 millions years ago. According to the

48 *J. Agric. Food Chem.,* Vol. 51, No. 1, 2003

Jérôme et al.

Table 7. Nucleotide Divergences (Tamura–Nei Distances) between Pairs of Haplotypes (Short Fragment) for the Eight Species of Clupeiformes Tested with the RFLP Method

	E. encrasicolus	C. harengus	S. sprattus	S. pilchardus	S. maderensis	S. aurita	S. caeruleus	S. sagax
E. encrasicolus		0.21	0.27	0.24	0.25	0.23	0.21	0.2
C. harengus			0.17	0.23	0.22	0.23	0.21	0.22
S. sprattus				0.31	0.23	0.18	0.28	0.27
S. pilchardus					0.24	0.18	0.21	0.22
S. maderensis						0.16	0.25	0.26
S. aurita							0.20	0.21
S. caeruleus								0.01

molecular clock (which records 2% of divergence per million years), our estimates (ca. 25%) suggest a divergence time of 12.5 million years.

Among the different genera studied, *Sprattus* and *Clupea* are particularly close (≈13%), which suggests a divergence time of 6.5 million years. Conversely, divergences among species of the genus *Sardinella* are particularly great. *Sardinella maderensis* and *Sardinella aurita* are genetically distant from each other by more than 20%, which corresponds to the order of divergence between genera of the *Clupeidae* family. On this basis, it seems likely that one of these two taxa should be assigned to one other genus. However, as the history of genes does not exactly match that of species (*51*), this level of divergence needs to be checked in other genes. In any event, confusions could exist for small divergences (*51*, *52*), which is not exactly the case described here. If the effect of gene history is not excluded, our data tend to show a discrepancy between morphological and molecular taxonomy. In this respect, other replacements in taxonomy have been reported for fish from molecular data (*53*).

Phylogenetic data indicate that the order of Clupeiformes is composed of genera clearly differentiated by cytochrome *b* sequences. Thus, the diagnosis of morphologically close species belonging to different genera, such as *Sardina* and *Sardinops*, could easily be performed from sequence data.

Diagnosis of Species by the PCR−RFLP Method. Sequencing, though a powerful tool, is not used routinely in food control laboratories, which suggests that PCR−RFLP could be a relevant alternative. For applications on canned fish, it is necessary to amplify short DNA fragments (∼150 bp), because treatments such as boiling and canning can severely damage DNA (*54*). Moreover, PCR−RFLP is a common technique used in the diagnosis of commercial species (*33*, *55*). It can be inexpensive, easy to use, and repeatable (*56*), if the number of restriction enzymes is limited and if rare enzymes are not selected.

The results for the PCR−RFLP method indicate that the primers defined were quite suitable for amplification, particularly the adapted form providing a 142 bp fragment usable for the eight Clupeiformes species tested. In all cases, the specific restriction profile of *S. pilchardus* was obtained, regardless of the origin and form (frozen or canned) of the samples. No ambiguities were observed between *S. pilchardus* and the other species, especially *Sardinops*. The purpose of this RFLP method was to differentiate species of *S. pilchardus* from the other eight species tested (sardine type products) but not to distinguish each species tested from the others. After restriction analysis, the specific profiles of *S. aurita*, *S. sprattus*, and *S. melasnoctictus* were discriminated but not those of the other three species, namely *S. maderensis, E. encrasicolus*, and *C. harengus*. *S. sagax* had a pattern similar to *S. caeruleus*. However these last RFLP profiles provide initial information for discriminating

between sardine and sardine type products, which could be supplemented by sequencing for more precise identification.

No restriction polymorphism was detected when some *Sardina* haplotypes were checked for variability on the short fragments used (data not shown). However, it could be possible to observe such variations in the genus *Sardina*. In this case, other enzymes or sequencing could allow differentiation of this species from others. As many authors have mentioned the problem of intraspecific variation (*56*), further studies on the genetic variation of *S. pilchardus* may help resolve this problem.

In doubtful cases, direct analysis of the nucleotide sequence could be used. Nucleotide divergences estimated from the short DNA fragments were of the same order as those estimated from longer fragments (**Table 7**), which suggests that identification would be possible. In all events, PCR products have to be produced for the both methods. Today the decrease of the price for DNA sequencing of shorts fragments as those we expect to amplify from canned food allows one to estimate that the cost of sequencing and the cost of restriction digests with two restriction enzymes is becoming practically equivalent. To date, additional trials on canned material have provided good results (data not shown), but more samples of various origins need to be tested to validate the method.

LITERATURE CITED

(1) Durand, P.; Landrein, A.; Quéro, J. C. *Catalogue électrophorétique des poissons commerciaux*; IFREMER: Nantes, France, 1985.

(2) Lundstrom, R. Fish species identification by thin-layer polyacrylamide gel isoelectric focusing: Collaborative study. *J Assoc. Off. Anal. Chem.* **1980**, *63*, 69−73.

(3) Rehbein, H.; Mackie, I. M.; Gonzales-Sotelo, C.; Perez-Martin, R.; Quinteiro, J.; Rey-Mendez, M. Fish species identification in canned tuna by DNA analysis (PCR−SSCP). *Inf. Fischwirtsch.* **1995**, *42*, 209−212.

(4) Piñeiro, C.; Sotelo, C. G.; Medina, I.; Gallardo, J. M.; Pérez-Martín, R. Reverse-phase HPLC as a method for the identification of gadoid fish species. *Z. Lebensm.-Unters.-Forsch., A* **1997**, *204*, 411−416.

(5) Osman, M. A.; Ashoor, S. H.; Marsh, P. C. Liquid chromatographic identification of common fish species. *J. Assoc. Off. Anal. Chem.* **1987**, *70*, 618−625.

(6) Etienne, M.; Jérôme, M.; Fleurence, J.; Rehbein, H.; Kundiger, R.; Yman, I.; Ferm, M.; Craig, A.; Mackie, I.; Jessen, F.; Smelt, A.; Luten, J. A standardized method of identification of raw and heat-processed fish by urea isoelectric focusing: A collaborative study. *Electrophoresis* **1999**, *20*, 1923−1933.

Molecular Phylogeny of Sardines

J. Agric. Food Chem., Vol. 51, No. 1, 2003   **49**

(7) Etienne, M.; Jérôme, M.; Fleurence, J.; Rehbein, H.; Kündiger, R.; Mendes, R.; Costa, H.; Perez-Martin, R.; Pineiro-Gonzalez, C.; Craig-A.; Mackie, I.; Yman, I. M.; Ferm, M.; Martinez, I.; Jessen, F.; Smelt, A.; Luten, J. Identification of fish species after cooking by SDS−PAGE and urea IEF: A collaborative study. *J. Agric. Food Chem.* **2000**, *48*, 2653−2658.

(8) Etienne, M.; Jérôme, M.; Fleurence, J.; Rehbein, H.; Kündiger, R.; Mendes, R.; Costa, H.; Martinez, I. Species identification of formed fishery products and high pressure-treated fish by electrophoresis: a collaborative study. *Food Chem.* **2001**, *72*, 105−112.

(9) Mackie, I.; Craig, A.; Etienne, M.; Jérôme, M.; Fleurence, J.; Jessen, F.; Smelt, A.; Kruijt, A.; Yman, I. M.; Ferm, M.; Martinez, I.; Perez-Martin, R.; Pineiro, C.; Rehbein, H.; Kundiger, R. Species identification of smoked and gravad fish products by sodium dodecylsulphate polyacrylamide gel electrophoresis, urea isoelectric focusing and native isoelectric focusing: a collaborative study. *Food Chem.* **2000**, *71*, 1−7.

(10) Piñeiro, C.; Barros-Velazquez, J.; Perez-Martin, R. I.; Martinez, I.; Jacobsen, T.; Rehbein, H.; Kundiger, R.; Mendes, R.; Etienne, M.; Jérôme, M.; Craig, A.; Mackie, I. M.; Jessen, F. Development of a sodium dodecyl sulfate-polyacrylamide gel electrophoresis reference method for the analysis and identification of fish species in raw and heat-processed samples: A collaborative study. *Electrophoresis* **1999**, *20*, 1425−1432.

(11) Rehbein, H.; Kundiger, R.; Yman, I. M.; Ferm, M.; Etienne, M.; Jérôme, M.; Craig, A.; Mackie, I.; Jessen, F.; Martinez, I.; Mendes, R.; Smelt, A.; Luten, J.; Pineiro, C.; Perez-Martin, R. Species identification of cooked fish by urea isoelectric focusing and sodium dodecyl sulfate polyacrylamide gel electrophoresis: a collaborative study. *Food Chem.* **1999**, *67*, 333−339.

(12) Bossier, P. Authentication of seafood products by DNA patterns. *J. Food Sci.* **1999**, *64*, 189−193.

(13) Lockley, A. K.; Bardsley, R. G. DNA-based methods for food authentication. *Trends Food Sci. Technol.* **2000**, *11*, 67−77.

(14) 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.

(15) 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.

(16) Bartlett, S. E.; Davidson, W. S. FINS (Forensically Informative Nucleotide Sequencing): A procedure for identifying the animal origin of biological specimens. *Biotechnology* **1991**, *12*, 408−411.

(17) Albert, B.; Lewis, D.; Raff, J.; Roberts, K.; Watson, J. D. In *Molecular Biology of the Cell*; Garland Publishing: New York, 1994; p 94.

(18) Carvalho, G. R.; Hauser, L. Advances in the molecular analysis of fish population structure. *It. J. Zool.* **1998**, *65*, 21−33.

(19) Sezaki, K.; Kuboshima, Y.; Mitani, I.; Fukui, A.; Watabe, S. Identification of chub and spotted mackerels with mitochondrial cytochrome b genes and its application to respective pelagic eggs fixed with formalin. *Nippon Suisan Gakkaishi* **2001**, *67*, 17−22.

(20) Bembo, D. G.; Carvalho, G. R.; Snow, M.; Cingolani, N.; Pitcher, T. J. Stock discrimination among European anchovies, *Engraulis encrasicolus*, by means of PCR-amplified mitochondrial DNA analysis. *Fish. Bull.* **1995**, *94*, 31−40.

(21) Grant, W. S.; Clark, A. M.; Bowen, B. W. Why restriction fragment length polymorphism analysis of mitochondrial DNA failed to resolve sardine (Sardinops) biogeography: insights from mitochondrial DNA cytochrome b sequences. *Can. J. Fish. Aquat. Sci.* **1998**, *55*, 2539−2547.

(22) Hauser, L.; Ward, R. D. Population identification in pelagic fish: the limits of molecular markers. *Adv. Mol. Ecol.* **1998**, 191−224.

(23) Meyer, A. DNA technology and phylogeny of fish. In *Genetics and Evolution of Aquatic Organisms*; Chapman and Hall: Beaumont, AR, 1994.

(24) Grant, W. S.; Bowen, B. W. Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *J. Hered.* **1998**, *89*, 415−426.

(25) Froese, R.; Pauly, D. World Wide Web electronic publication, FishBase: www.fishbase.org, 2002.

(26) Parrish, R. H.; Serra, R.; Grant, W. S. The monotypic sardines *Sardina* and *Sardinops*: their taxonomy, distribution, stock structure, and zoogeography. *Can. J. Fish. Aquat. Sci.* **1989**, *46*, 2019−2036.

(27) Chickhi, L.; Agnèse, J.-F.; Bonhomme, F. Fortes différences des ADN mitochondriaux de populations de *Sardinella aurita* de la mer Méditerranée et de l'Atlantique Est. *C. R. Acad. Sci., Ser. 3* **1997**, 289−297.

(28) Chickhi, L.; Bonhomme, F.; Agnèse, J.-F. Low genetic variability in a widely distributed and abundant clupeid species, *Sardinella aurita*. New empirical results and interpretations. *J. Fish Biol.* **1998**, *52*, 861−878.

(29) Magoulas, A.; Tsimenides, N.; Zouros, E. Mitochondrial DNA phylogeny and the reconstruction of population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Mol. Biol. Evol.* **1996**, *13*, 178−190.

(30) Tringali, M. D.; Wilson, R. R. Differences in haplotype frequencies of mtDNA of the Spanish sardine *Sardinella aurita* between specimens from the eastern Gulf of Mexico and southern Brazil. *Fish. Bull. U.S.* **1993**, *91*, 362−370.

(31) Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111−120.

(32) Grant, W. S.; Leslie, R. W. Late Pleitocene dispersal of Indian-Pacific sardine populations in an ancient lineage of the genus *Sardinops*. *Mar. Biol.* **1996**, *126*, 133−142.

(33) Quinteiro, J.; Sotelo, C. G.; Rehbein, H.; Pryde, S. E.; Medina, I.; 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.

(34) 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*, 95−98.

(35) Kumar, S.; Tamura, K.; Jakobsen, I.; Nei, M. MEGA2: Molecular Evolutionary Genetics Analysis Software. *Bioinformatics* **2001**, *17*, 1244−1245.

(36) Brown, W. M.; Prager, E. M.; Wang, A.; Wilson, A. C. Mitochondrial DNA sequence of primates: tempo and mode evolution. *J. Mol. Evol.* **1982**, *18*, 239−255.

(37) Holmquist, R. Transitions and transversions in evolutionary descent: an approach to understanding. *J. Mol. Evol.* **1983**, *19*, 134−144.

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

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

(40) Fitch, W. M. Toward defining the course of evolution: Minimum change for a specific tree topology. *Syst. Zool.* **1971**, *20*, 406−416.

(41) Felsenstein, J. *Phylip 3.5 (Phylogeny Inference Package)*; University of Washington: Seattle, WA, 1993.

(42) Inoue, J. G.; Miya, M.; Tsukamoto, K.; Nishida, M. Complete mitochondrial DNA sequence of the Japanese sardine *Sardinops melanosticus*. *Fish. Sci.* **2000**, *66*, 924−932.

(43) Miya, M.; Nishida, M. Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): First example of transfer RNA gene rearrangements in bony fishes. *Mar. Biotechnol.* **1999**, *1*, 416−426.

(44) Quinteiro, J.; Vidal, R.; Rey-Mendez, M. Phylogeny and biogeographic history of hake (genus Merluccius), inferred from mitochondrial DNA control-region sequences. *Mar. Biol.* **2000**, *136*, 163−174.

(45) Tinti, F.; Piccinetti, C.; Tommasini, S.; Vallisneri, M. Mitochondrial DNA variation, phylogenetic relationships, and evolution of four Mediterranean genera of soles (Soleidae, Pleuronectiformes). *Mar. Biotechnol.* **2000**, *2*, 274−284.

(46) Graur, D.; Li, W.-H. *Fundamentals of Molecular Evolution*, 2nd ed.; Sinauer Associates: Sunderland, MA, 1999.

(47) Johns, G. C.; Avise, J. C. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Mol. Biol. Evol.* **1998**, *15*, 1481−1490.

(48) Okazaki, T.; Kobayashi, T.; Uozomi, Y. Genetic relationships of pilchards (genus *Sardinops*) with anti-tropical distributions. *Mar. Biol.* **1996**, *126*, 585−590.

(49) Bowen, B. W.; Grant, W. S. Phylogeography of the sardines (*Sardinops spp*): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution* **1997**, *51*, 1601−1610.

(50) Svetovidov, A. N. *Fishes: Clupeidae. Fauna U.S.S.R. (New Ser.)* **1952**, *2*, 1.

(51) Pamilo, P.; Nei, M. Relationship between gene trees and species trees. *Mol. Biol. Evol.* **1988**, *5*, 568−583.

(52) Avise, J. C. *Molecular markers, natural history and evolution*; Chapman and Hall: London, 1994.

(53) Lydeard, C.; Roe, K. J. *The phylogenetic utility of the mitochondrial cytochrome b gene for inferring relationships among actinopterygian fishes*; Kocher, T. D., Stepien, C. A., Eds.; Academic Press: San Diego, CA; pp 285−303.

(54) Rehbein, H.; Kress, G.; Schmidt, T. Application of PCR−SSCP to species identification of fishery products. *J. Sci. Food Agric.* **1997**, *74*, 35−41.

(55) 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.

(56) Wolf, C.; Burgener, M.; Hubner, P.; Luthy, J. PCR−RFLP analysis of mitochondrial DNA: Differentiation of fish species. *Lebensm.-Wiss. Technol.* **2000**, 144−150.