International Standard

ISO 17174

First edition
2024-09

Molecular biomarker analysis — DNA barcoding of fish and fish products using defined mitochondrial cytochrome b and cytochrome c oxidase I gene segments

Analyse de biomarqueurs moléculaires — Codes-barres d'ADN de poissons et de produits à base de poisson à l'aide de segments de gènes mitochondriaux de cytochrome b et cytochrome c oxydase I

Reference number
ISO 17174:2024(en)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee 34, *Food products*, Subcommittee SC 16, *Horizontal methods for molecular biomarker analysis*, in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 460, *Food authenticity*, in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Food safety is a key aspect in terms of consumer protection. In the last three decades, globalization has taken place in the trade of food. Fish trade channels are becoming steadily longer and more complicated so that sophisticated traceability tools are needed to ensure food safety. Correct food labelling is a prerequisite to ensure safe fish products and fair trade as well as to minimize illegal, unreported and unregulated (IUU) fishing. In particular, the fact that fish is increasingly being processed in export countries makes the identification of species by morphological characteristics impossible. The development of reliable, harmonized and standardized protocols for the authentication of fish products is necessary to ensure consumer protection and the detection of potential food fraud.

Molecular biomarker analysis — DNA barcoding of fish and fish products using defined mitochondrial cytochrome b and cytochrome c oxidase I gene segments

1 Scope

This document specifies a method for the identification of single fish and fish fillets to the level of genus or species. It allows the identification of a large number of commercially important fish species using DNA barcoding.

This method was validated on raw fish. Laboratory experience indicates additional applicability to processed fish products (e.g. cold smoked, hot smoked, salted, frozen, cooked, fried and deep-fried samples).

The described method is usually unsuitable for the analysis of highly processed foods (e.g. tins of fish with highly degraded DNA where the fragment lengths are not sufficient for amplification of the targets). Furthermore, it does not apply to complex fish products containing mixtures of two or more fish species.

The identification of fish species is carried out by PCR amplification of either a segment of the mitochondrial cytochrome b gene (*cytb*) or the cytochrome c oxidase I gene (*cox1, syn COI*), or both, followed by sequencing of the PCR products and subsequent sequence comparison with entries in databases.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 16577, *Molecular biomarker analysis — Vocabulary for molecular biomarker analytical methods in agriculture and food production*

ISO 20813, *Molecular biomarker analysis — Methods of analysis for the detection and identification of animal species in foods and food products (nucleic acid-based methods) — General requirements and definitions*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 16577 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
alignment
sequence alignment
arrangement of nucleic acid sequences or protein sequences according to regions of similarity

Note 1 to entry: Alignment is a process or result of matching up the nucleotide residues of two or more biological sequences to achieve maximal levels of *identity* (3.3).

[SOURCE: ISO 16577:2022, 3.7.18, modified — “alignment” was added as the preferred term; Note 1 to entry was added.]

3.2
FASTA format
text-based format for representing either nucleotide sequences or amino acid (protein) sequences, in which nucleotides or amino acids are represented using single-letter codes

Note 1 to entry: A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line (defline) is distinguished from the sequence data by a greater-than (“>”) symbol at the beginning. It is recommended that all lines of text be shorter than 80 characters in length.

Note 2 to entry: An example sequence in FASTA format is:

```
>Sample_04_cytb
ATGGCCAGCCTCCGAAAAACTCATCCCCTTCTAAAGATTGCTAATGATGCATTAGTAGACCTTCCTGCCCCCTCT
AACCTCTCAACATTATGAAACTTCGGGTCTCTCCTAGGCCTCTGCTTAGCCGCCCAAATCTTAACAGGACTATTT
CTAGCGATACATTATACCGCAAACGTCGAGATAGCTTTCTCATCCGTCGTACACATCTGCCGCGACGTAAATTAC
GGATGACTAATCCGCAACATACACGCCAACGGCGCTTCTTTCTTCTTCATCTGCCTCTACCTACACATTGCAGGA
GGCCTATATTACGGCTCCTACTTATTCATAGAGACCTGAAACATTGGAGTTGTACTATTCCTTTTAGTAATAATG
ACCGCCTTCGTAGGCTACGTCCTCCCT
```

Note 3 to entry: Blank lines are not allowed in the middle of FASTA input. Sequences are represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with these exceptions:

— lower-case letters are accepted and are mapped into upper-case;

— a single hyphen or dash can be used to represent a gap of indeterminate length.

It is common to end the sequence with an “*” (asterisk) character and to leave a blank line between the description and the sequence.

[SOURCE: ISO 16577:2022, 3.1.2, modified — Example in Note 2 to entry was replaced; the third list item in Note 3 to entry was deleted.]

3.3
identity
extent to which two (nucleotide or amino acid) sequences have the same residues at the same positions in an *alignment* (3.1)

Note 1 to entry: Identity is often expressed as a percentage.

Note 2 to entry: In the sequence database of the Barcode of Life (BOLD), the term "similarity" is used instead of identity.

3.4
introgressed DNA
allele from one species incorporated in the gene pool of another, divergent species

Note 1 to entry: Introgression has usually happened via hybridization and backcrossing of individuals belonging to different species.

3.5
query
sequence (or other type of search term) that is compared to entries in a database

3.6
query coverage
percentage of the *query* (3.5) covered by *alignment* (3.1) to the database sequence

4 Symbols and abbreviated terms

bp	base pairs
Glu	glutamic acid, glutamate
tRNA	transfer RNA
cox1, COI	cytochrome c oxidase I gene
cytb	cytochrome b gene
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
dNTP	deoxyribonucleotide triphosphate
A	adenine
C	cytosine
G	guanine
I	inosine
R	purine base (adenine or guanine)
T	thymine
Y	pyrimidine base (cytosine or thymine)
IUU	illegal, unreported and unregulated
PCR	polymerase chain reaction

5 Principle

DNA is extracted from fish and fish products applying a suitable method. Segments of approximately 460 base pairs of *cytb* and/or approximately 650 base pairs of *cox1* are amplified by PCR. For the amplification of the *cytb* segment a set of two primers is used. Amplification of *cox1* uses a set of four primers to ensure the amplification of DNA of as many fish species as possible. Some of the used primers include ambiguous bases to increase the number of species detected by the two methods. The nucleotide sequences of the PCR products are determined by a suitable DNA sequencing method (e. g. Sanger sequencing). The two PCR primers used to generate the *cytb* amplicon are also used for sequencing. The *cox1* primers have M13 tails that enable sequencing of the *cox1* amplicons using M13 sequencing primers. The determined sequence is evaluated by comparison to sequence entries in databases, thus allowing the assignment to a fish species or genus according to the degree of identity with available sequences.

The decision whether the *cytb* or *cox1* segment or both are used for fish identification depends on the declared fish species, especially the applicability of the PCR method for the fish species and the availability of comparative sequences in the public databases.

6 Reagents and materials

During the analysis, unless otherwise stated, use only reagents of recognized molecular biology grade and distilled, demineralized water or water of equivalent purity, in accordance with ISO 20813. Laboratory organization shall follow ISO 20813.

6.1 Thermostable DNA polymerase[1)].

6.2 PCR reaction buffer (including $MgCl_2$ or with separate $MgCl_2$ solution)[1)].

6.3 dNTP mix (dATP, dCTP, dGTP and dTTP).

NOTE dNTP can also be part of a commercial PCR master mix.

6.4 Oligonucleotides (see Tables 1 and 2).

NOTE 1 The abbreviations of the DNA bases in Tables 1 and 2 are based on the recommendations for unambiguous, uniform, and consistent nomenclature, published by the International Union of Pure and Applied Chemistry (IUPAC).[4]

Table 1 — Oligonucleotides for amplification of the *cytb* gene region[1]

Name	DNA sequence of oligonucleotide
L14735	5'-AAA AAC CAC CGT TGT TAT TCA ACT A-3'
H15149ad	5'-GCI CCT CAR AAT GAY ATT TGT CCT CA-3'

NOTE 2 Laboratory experience has shown that M13 tails can also be used with *cytb* primers.

Table 2 — Oligonucleotides for amplification of the *cox1* gene region[2][3]

Name	DNA sequence of oligonucleotide[a]
VF2_t1	5'-***TGT AAA ACG ACG GCC AGT*** CAA CCA ACC ACA AAG ACA TTG GCA C-3'
FishF2_t1	5'-***TGT AAA ACG ACG GCC AGT*** CGA CTA ATC ATA AAG ATA TCG GCA C-3'
FishR2_t1	5'-***CAG GAA ACA GCT ATG AC***A CTT CAG GGT GAC CGA AGA ATC AGA A-3'
FR1d_t1	5'-***CAG GAA ACA GCT ATG AC***A CCT CAG GGT GTC CGA ARA AYC ARA A-3'
[a] M13 tails of the primers are highlighted in bold and italic.	

6.5 Trehalose.

6.6 Agarose.

6.7 DNA size standard.

6.8 Sequencing primers (see Table 3).

Table 3 — Sequencing primers for *cox1* PCR products[3]

Name	DNA sequence of oligonucleotide
M13F (–21)	5'-TGT AAA ACG ACG GCC AGT-3'
M13R (–27)	5'-CAG GAA ACA GCT ATG AC-3'

1) During the collaborative studies, the Maxima® Hot Start PCR Master Mix (2 ×) of Fermentas GmbH (ready-to -use PCR buffer solution including thermostable DNA polymerase) was used for the *cytb* amplification and the BIOTAQ™ DNA polymerase of Bioline with 10 × reaction buffer and separate $MgCl_2$ solution for the *cox1* amplification. In addition to the recommended BIOTAQ™ DNA polymerase, other mastermixes and polymerases were successfully used. Maxima® Hot Start PCR Master Mix (2 ×) of Fermentas GmbH (ready-to-use PCR buffer solution including thermostable DNA polymerase) and BIOTAQ™ DNA polymerase of Bioline are examples of suitable products available commercially. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of these products.

7 Apparatus

In addition to standard laboratory equipment, the following apparatus should be used.

7.1 UV-spectrophotometer or fluorometer, to determine the concentration of DNA.

7.2 Thermocycler.

7.3 Gel electrophoresis device.

7.4 Gel documentation system.

7.5 DNA sequencer.

8 Procedure

8.1 Sample preparation

The test portion used for DNA extraction shall be representative of the laboratory sample, for example, using the following guidance. In samples that consist of processed materials (e.g. convenience foods), single fish pieces shall be analysed separately. For the analysis of samples composed of several pieces (e.g. bags with different fillets), test portions for every putative fish species are taken and analysed separately. To minimize the risk of amplifying adhering contaminants, test sample material shall not be taken from the surface of the laboratory sample, see also ISO 20813 and ISO 22949-1[7].

8.2 DNA extraction

General instructions and measures specified in ISO 21571:2005[5] should be followed for the extraction of DNA from the test sample. For example, the DNA extraction methods specified in ISO 21571:2005, Annex A,[5] can be used. Commercial kits can be used for the extraction and purification of DNA, if their applicability for the extraction of DNA from fish has been experimentally confirmed.

8.3 PCR

8.3.1 General

Universal primers are used for amplification of segments of mitochondrial *cytb* and *cox1*. The primer H15149ad is a universal primer with fish-specific adaptations that binds to the *cytb* section. L14735 binds to a section in the neighbouring highly conserved tRNA-Glu gene.[1] The primers used for amplification of the mitochondrial *cox1* fragment were designed to the 5' region of the *cox1* of fish.[2][3]

Both primer sets have been tested against a very broad taxonomic range of fish species. It is currently known that, with some exceptions, the primer pair L14735/H15149ad did not react with samples labelled as barramundi (*Lates calcarifer*) or Nile perch (*Lates niloticus*).[6] The *cox1* primer set has only failed in a small minority of cases (< 5 % of species tested).[2]

8.3.2 PCR setup

The method was validated for a total volume of 25 µl (*cytb*) or 20 µl (*cox1*) per PCR. The reagents given in Tables 4 and 5 shall be used for the *cytb* and *cox1* PCR, respectively.

Reagents shall be completely thawed and should be centrifuged briefly before usage. A PCR reagent mixture is prepared containing all PCR components in the given concentrations except for the DNA extract or the controls. The amount of PCR mixture prepared depends on the total volume per PCR and the total number of reactions including a sufficient pipetting reserve.

Positive PCR results are expected when using a DNA concentration of approximately 1 ng DNA per µl of the final solution [25 ng (*cytb*) or 20 ng (*cox1*)].

To improve the PCR result, the DNA quantity can be increased (e.g. to increase the yield of PCR product) or decreased (e.g. to avoid PCR inhibition).

Table 4 — Components for the *cytb* PCR

Reagent (stock solution)	Final composition in the reaction solution
PCR buffer	1 ×
$MgCl_2$[a]	1,5 mmol/l
dNTP mix[a]	0,2 mmol/l for each dNTP
Primer L14735	500 nmol/l
Primer H15149ad	500 nmol/l
Thermostable DNA polymerase	0,5 units to 1 unit
Water	Add to obtain final volume
Sample DNA	Approximately 1 ng/µl
[a] Use a reagent only if one is not already included in the PCR buffer.	

Table 5 — Components for the *cox1* PCR

Reagent (stock solution)	Final composition in the reaction solution
Trehalose[a]	5 %
PCR buffer	1 ×
$MgCl_2$[b]	2,5 mmol/l
dNTP mix[b]	0,2 mmol/l for each dNTP
Primer VF2_t1	100 nmol/l
Primer FishF2_t1	100 nmol/l
Primer FishR2_t1	100 nmol/l
Primer FR1d_t1	100 nmol/l
Thermostable DNA polymerase	0,5 units
Water	Add to obtain final volume
Sample DNA	Approximately 1 ng/µl
[a] Optional.	
[b] Use a reagent only if one is not already included in the PCR buffer.	

Mix the PCR reagent mixture, centrifuge briefly and split into the individual reactions. Pipette the DNA extracts to be examined or the PCR controls (see 8.3.3) into the different reaction solutions.

For further information on PCR controls, see also ISO 20813.

8.3.3 PCR controls

In addition to the reaction setups for the samples to be analysed, an amplification reagent control and an extraction blank control in accordance with ISO 20813 shall be included.

A positive DNA target control (see ISO 20813) can be used to demonstrate the ability of the PCR to amplify the target sequence. As positive control material, genomic DNA extracted from a known fish species or an available plasmid containing the target sequence can be used.

If a sample shows no amplification in both targets, an inhibition control reaction should be performed to exclude an inhibition of the PCR as the cause (see ISO 20813). This can be done either by dilution of sample DNA or by using an internal inhibition control assay.

Additional PCR controls can be used, see ISO 20813.

8.3.4 Thermal cycling

Transfer the reaction setups into the thermal cycler and start the temperature-time programme. The temperature-time programmes as outlined in Tables 6 and 7 have been successfully used in the collaborative studies.

NOTE The use of different reagent conditions and thermocyclers can require specific optimization. The time for initial denaturation depends on the thermostable DNA polymerase used.

Table 6 — Temperature-time programme for the *cytb* PCR

Step	Parameter		Temperature	Time	Cycles
1	Initial DNA denaturation and activation of the hot-start DNA polymerase (if used)		95 °C	15 min	1
2	Amplification	Denaturation	95 °C	40 s	35
		Annealing	50 °C	80 s	
		Elongation	72 °C	80 s	
3	Final elongation		72 °C	10 min	1

Table 7 — Temperature-time programme for the *cox1* PCR

Step	Parameter		Temperature	Time	Cycles
1	Initial DNA denaturation and activation of the hot-start DNA polymerase (if used)		94 °C	2 min	1
2	Amplification	Denaturation	94 °C	30 s	35
		Annealing	52 °C	40 s	
		Elongation	72 °C	60 s	
3	Final elongation		72 °C	10 min	1

After the PCR thermal cycling is finished, follow with the evaluation of the PCR products or store samples in the refrigerator until further analysis.

8.4 Evaluation of PCR products

The PCR product should be assessed for quality, and its quantity estimated (e.g. by agarose gel electrophoresis).

Gel electrophoresis of DNA in an agarose gel is a standard technique in molecular biology. Therefore, only general conditions that need to be adapted to each laboratory are suggested.

A volume of 1 µl to 10 µl of each PCR product is separated in, for example, an agarose gel of suitable concentration (e.g. a mass fraction of 1 % to 2 % per volume) and evaluated with a gel documentation system. In one lane, an appropriate DNA size standard is included for comparison. For the *cytb* PCR a product of approximately 460 bp and for the *cox1* PCR a product of approximately 700 bp (which includes M13 tails) should be clearly visible after gel electrophoresis.

No amplicons should be visible for the amplification reagent control and the extraction blank control. For the positive DNA target control, PCR products of the expected size should be visible.

8.5 Evaluation of the PCR results

The *cytb* PCR or the *cox1* PCR or both can show a positive or a negative result for amplification of the target sequence(s).

Depending on the outcome of the PCR, the next step is to evaluate the following:

— If the sample is positive for one or both targets (*cytb* and/or *cox1*), sequencing of one or both PCR products is the next step (see 9.1).

— If the sample is negative for the targets, the control results are acceptable, and the sample showed no inhibition (8.3.3) it is possible that one of the following cases applies:

  — The primers do not sufficiently anneal to the target sequence of the DNA extracted from the fish species under analysis. In this case, species identification of the sample is not possible with this method and analyses with further universal primer pairs (i.e. 16S rRNA gene-specific primers) can follow the tests (see ISO 22949-1:2021, 7.5.2.2)[7].

  — The DNA extracted was degraded or not of sufficient quantity or quality for PCR.

NOTE 1 For some fish species, the primer pairs of the PCR systems do not amplify the target sequences. Examples are listed in Table A.1.

NOTE 2 Additional universal primer pairs are described in the literature [e.g. Fish *cytb*-F/*Cytb*1–5R for full-length barcode (approximately 720 bp) cytochrome b amplification,[8] mlCOIintF/jgHCO2198 for mini-barcode (313 bp) COI amplification[9] and 16SH/16SR for amplification of 16S rDNA gene segments[10]]. Verified or validated barcoding methods are listed in the Barcoding Table of Animal Species (BaTAnS).[11] The current version of the BaTAnS can be obtained via the webpage www.bvl.bund.de.

9 Sequencing

9.1 Sequencing of PCR products

Sequencing of PCR products is carried out in-house or outsourced according to the method available for the testing laboratory.

If PCR products obtained from a sample show a single band in the gel, the (remaining) PCR reaction mixture can be purified directly using a suitable commercial kit. If more bands are present, cut out the appropriate band from the gel prior to purification.

A commonly applied standard procedure is Sanger sequencing (a cycle sequencing method) using fluorescence-labelled dideoxynucleotides. For sequencing the *cytb* PCR products, the primers used for the generation of the amplicons serve as sequencing primers. The sequencing primers for *cox1* PCR products bind only to the M13 tail and therefore differ from those used in the PCR amplification (see Table 5).

DNA fragments from the sequencing reaction are subsequently separated by means of a DNA sequencer (e.g. using capillary electrophoresis). Fluorescence signals are recorded and analysed with the device software (see ISO 22949-1:2021, 7.5.2.5)[7].

NOTE Laboratory experience has shown that M13 tails can also be added to the *cytb* primers facilitating the use of M13 primers for sequencing *cytb* PCR products.

9.2 Evaluation of sequence data

The sequence trace data (or electropherogram) shall be checked visually to ensure the sequencing reaction has worked appropriately, and base calling is accurate. In case of misassigned nucleotides to electropherogram peaks, sequences should be edited using appropriate software and evaluating the fluorescence peak data. Based on experience, the length of the determined sequence should be in general approximately 80 % of the expected read length.

A sequence analysis should be preferably performed of both DNA strands. These complementary/overlapping sequences should be assembled into a consensus sequence. This serves as an important way of checking the accuracy of the sequence and can help remove any ambiguous bases.

The sequences of the primers are excluded from the determined sequences before the comparison to database sequences. The consensus sequence resulting from the forward and the reverse reaction should be used for sequence comparison with public databases.

9.3 Comparison of the sequence with public databases

9.3.1 General

The *cytb* and *cox1* amplicon sequences are evaluated according to the taxon by comparison to sequence entries in the nucleotide collection (nr/nt) by BLAST®[12].[2)] For *cox1* DNA sequences, the *cox1* sequence database of the Barcode of Life (BOLD)[13] project is used in parallel.

NOTE GenBank®'s (nr/nt) nucleotide collection is a non-redundant database in which identical sequences have been merged into one single entry. GenBank® is hosted by the National Center for Biotechnology Information (NCBI).

Prior to queries in public databases, it is important to gather information about the taxon under investigation (e.g. from FishBase,[15] NCBI taxonomy browser and/or BOLD taxonomy section[16],[3)]), including:

— additional species in the same genus;

— presence of declared and related species in GenBank® and BOLD[3)];

— amount of *cytb* and/or *cox1* sequences of respective species in GenBank® and/or BOLD.

A scientific literature search should be performed to obtain information about the genetic relationship between and among species within a taxon, the possibility of hybrid fish species or the occurrence of introgressed DNA in a particular species (e.g. some tuna species from the genus *Thunnus*)[17][18].

Information from the AquaGene database[19],[3)] a curated genetic information database for fish species based on various genetic markers, can also be used to assess results. Other databases such as the French Polynesia Fish Barcoding Database[20] [3)] can provide useful information for fish species identification.

FASTA format should be used when pasting sequences into the query boxes of BLAST® and BOLD, so that the query results are displayed together with the names of the sequences.

9.3.2 Sequence comparison of *cytb* and/or *cox1* DNA sequences with GenBank®

The edited *cytb* and/or *cox1* DNA sequences are subjected to a comparison with sequences from the nucleotide collection (nr/nt) of GenBank® by BLAST®, optimized for highly similar sequences (Megablast) in order to identify the species from which the sequences originated. The identified target sequences (hits) are displayed as a list. Before assigning a species, re-sort the hits by maximum identity. The hits are additionally presented as alignments with the query sequence. When more than 100 hits with ≥ 98 % identity, the number of maximal target sequences shall be increased (under algorithm parameters) to identify all relevant taxa. The BLAST® tool can also produce a neighbour-joining tree to graphically display the results of the homology search. Neighbour-joining is a bottom-up (agglomerative) clustering method for the creation of phylogenetic trees. The query sequence is highlighted in the tree diagram.

The query result should be saved as a PDF-file (or similar) to document the output of the database at the time of comparison and shall include:

— species of sequence entries with identities ≥ 98 % (including gaps);

— degree of identity, in %;

2) BLAST® and GenBank® are the trademarks of products supplied by National Center for Biotechnology Information (NCBI). This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named. Equivalent products may be used if they can be shown to lead to the same results.

3) BOLD, FishBase[15], NCBI taxonomy browser, AquaGene[19] and the French Polynesia Fish Barcoding Database[20] are examples of suitable tools and databases available. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of these products.

— degree of query coverage, in %;

— search database used and the date of search.

If a declared species does not appear in the first 100 hits of the BLAST® search, then the BLAST® search can be repeated with restriction to database entries belonging to the declared species. Repeating the search can be done by entering the species name under "Organism" in the BLAST® form. The discontiguous megablast (optimized for more dissimilar sequences) should be used to find all relevant entries. The result is recorded, as above.

NOTE The search database is non-redundant which means that sequences with 100 % identity and the same length are merged into one hit. This can be determined by moving the mouse over the hit and reading the appearing text, or by investigating the alignments section. Hidden sequence entries are displayed under "See x more title(s)". This is frequently found for entries belonging to one species and can disguise sequence abundance but is also possible for sequences derived from different species.

9.3.3 Sequence comparison of *cox1* DNA sequences with BOLD

The *cox1* sequence database of BOLD can also be used to assign species with *cox1* sequences.[3][16] The web page "Identification Request/Animal identification" acts as a portal allowing the edited *cox1* query sequence to be compared with the BOLD reference database. Various search options are possible that relate to different collections of reference data, but the default settings, selecting only records at the species level, provide an excellent initial step for identifying the species. The search result provides the following information:

— At the top of the query result sheet, a species is assigned, based on the identity values from the top hits.

— The percent similarity for the top matching records in the database against the query sequence obtained from the sample is displayed in a graph, full records for these corresponding top matches can be shown (except for entries with the status "private").

— BOLD can also produce a simple neighbour-joining tree to graphically display the search result. The query sequence is highlighted in the tree diagram.

Sequences should be subjected to BOLD identification engine in forward orientation using FASTA format. If an identification result is not obtained, the reverse complement sequence of the input sequence should be subjected for BOLD analysis.

NOTE In some cases, species with high identity can be missing in the result sheet, if there are more than the displayed sequences with equally high identities.

In cases where BOLD returns more than one species or none at all, and a species match cannot be made, an additional search can be performed, selecting a different set of reference data, e.g. the "Public Record Barcode Database" (this restricts the search to sequences that have been published) or "All Records on BOLD" (this is the broadest database). In cases where BOLD cannot identify the query sequence, other publicly available reference databases should be searched, e.g. GenBank®[14].

The query result shall be saved as a PDF-file (or similar) to document the output of the database at the time of comparison and shall be recorded as follows:

— species, to which BOLD has matched the sequence;

— species name and similarity of hits with ≥ 98 %;

— search database used and the date of search.

10 Interpretation of database query results

The specificity of the results from the database queries depends on the genetic diversity of the organism group to which the sample belongs and the database breadth and quality. As a consequence, the specificity of this method can vary between organism groups. Identification will not be possible at the species level for

all fish samples if genetic diversity of the organism group, database breadth and database quality are not sufficient.

For an accurate assignment of fish species apply the following criteria:

— Sequence matches should have ≥ 98 % identity (similarity) with ≥ 90 % query coverage.

— In a neighbour-joining tree, the queried "unknown specimen" should cluster only with sequences originating from a single species (i.e. monophyletic group).

In assigning identities, users should recognize the following:

— Related species can show sequence identities < 98 %. However, an exact threshold cannot be specified as this depends on the taxon under investigation. Identification of a sample to species or genus level, or not at all, should be decided on a case-by-case basis. Analysis of both *cytb* and *cox1* gene regions should be performed for inconclusive results.

— Species assignments in the database can be incorrect:

  — If only one particular sequence of a certain species or sequences deposited by only one group of researchers matches the query sequence, the result should be regarded as unconfirmed.

  — Single entries showing equally high identities to the query sequence that belong to a divergent species compared to a high amount of top hit entries could have been misidentified in the respective queried database. Many database sequences have been assigned to species by means of identity instead of morphological characteristics. Without morphological data supporting the hit, an abundance of sequences from a particular species can be misleading.

— Hybrids cannot be differentiated from maternal species, because the described identification method is based on maternally inherited mitochondrial gene sequences.

— Some genera show only small nucleotide differences within the *cytb* and *cox1* gene segments. This is especially but not exclusively true for the genera *Thunnus*[11] or *Coregonus*[21].

NOTE For samples that cannot be identified using the described method, alternative barcoding methods can be used, as appropriate (e.g. see References [8], [9], [10] and [11]).

11 Validation status and performance criteria

11.1 Collaborative study for the identification of fish species based on *cytb* sequence analysis

The *cytb* method was validated in a collaborative study with a total of 11 participants in 2010. The participants received six encoded fish samples in ethanol for the analysis. The extraction of DNA from the samples was carried out by the laboratories using the CTAB method.[22]

Sequence analysis of one or both strands was performed by the participants themselves or by contract laboratories. The sequences of the primers were excluded from the determined sequences. The edited sequences were subjected to comparison with GenBank sequences by BLAST®,[12] optimized for highly similar sequences using megablast and analysis sorted by maximum identity.

An overview of the results of the *cytb* collaborative study and their evaluation is shown in Tables 8 and 9.

Table 8 — Validation data from the *cytb* collaborative study

Parameter	Result	
Number of laboratories	11	
Number of samples per laboratory	6	
Number of expected results	66	
Samples with correctly identified fish species	63	(95 %)
Samples without identification of fish species	3	(5 %)
Samples with correctly identified fish genus	66	(100 %)
Samples falsely identified	0	(0 %)

Table 9 — Evaluation of the results of the *cytb* collaborative study depending on fish species

Fish species	Samples with correctly identified fish species	Samples without identified fish species	Samples with falsely identified fish species
Pleuronectes platessa (European plaice)	10	1	0
Solea solea (sole)	11	0	0
Glyptocephalus cynoglossus (witch flounder)	10	1	0
Psetta maxima (turbot)	11	0	0
Limanda limanda (common dab)	11	0	0
Limanda aspera (yellowfin sole)	10	1	0

Of the participating laboratories, nine laboratories correctly identified all samples. The three samples not identified at the species level were correctly identified at the genus level. One laboratory could not provide any information about the European plaice and witch flounder samples, another laboratory could not identify the yellowfin sole sample. The inability to correctly identify the sample was attributable to poor quality of the DNA sequences generated.

11.2 Collaborative study for the identification of fish species based on *cox1* sequence analysis

The *cox1* method was validated in a European collaborative study led by the consortium of the EU-project LABELFISH[23] with a total of 13 participants in 2014. The participants received 12 encoded fish samples in ethanol for the analysis.

During the collaborative study, it was recommended that the participants performed a sequence analysis of both DNA strands. These complementary sequences were assembled. The sequences of the primers were excluded from the determined sequences. The edited sequences were subjected to sequence comparison using BOLD.[13][24]

An overview of the results of the *cox1* collaborative study and their evaluation is shown in Tables 10 and 11.

Table 10 — Results from the *cox1* collaborative study

Parameter	Result	
Number of laboratories	13	
Number of samples per laboratory	12	
Number of expected results	156	
Samples with correctly identified fish species	136	(87 %)
Samples without identification of fish species	20	(13 %)
Samples with correctly identified fish genus	156	(100 %)
Samples falsely identified	0	(0 %)

Table 11 — Evaluation of the results of the *cox1* collaborative trial depending on fish species

Fish species	Samples with correctly identified fish species	Samples without identified fish species	Samples with falsely identified fish species
Solea solea (sole)	13	0	0
Glyptocephalus cynoglossus (witch flounder)	12	1	0
Pollachius pollachius (pollack)	13	0	0
Pollachius virens (saithe)	13	0	0
Gadus chalcogrammus (Alaska pollock)	12	1	0
Boops boops (bogue)	10	3	0
Merluccius merluccius (European hake)	10	3	0
Merluccius hubbsi (Argentine hake)	12	1	0
Thunnus albacares (yellowfin tuna)	5	8	0
Thunnus alalunga (albacore)	10	3	0
Lophius piscatorius (angler)	13	0	0
Lophius budegassa (blackbellied angler)	13	0	0

In all cases where the fish species was not identified the genus of the fish was correctly assigned. The high number of samples where the fish was not identified at the species level (eight laboratories) in the case of yellowfin tuna was expected beforehand. Several phylogenetic studies reported that *Thunnus albacares* and *Thunnus obesus* show only small sequence differences in the *cox1* fragment.[11][25][26] The collaborative study confirmed the finding that *cox1* is unable to fully resolve the species status of these two species.

12 Test report

The test report shall contain at least the following:

a) all information necessary for the identification of the sample;

b) a reference to this document (including its year of publication), i.e. ISO 17174:2024;

c) date and type of the sampling procedure (if known);

d) date of receipt of sample in the laboratory;

e) date of the test;

f) database used (version) and date of search;

g) indication of the identified fish species or genus, respectively, in Latin and the language of the test report;

h) determined identity with deposited sequences in % (number of identical nucleotides / number of compared nucleotides);

i) in the case of a deviation from the declared fish species: determined identity to the declared species in % (number identical nucleotides / number of compared nucleotides);

j) any deviation from the procedures described in this document;

k) any particular points observed in the course of the test;

l) any operation not specified in the method or regarded as optional which might have affected the results.

When agreed with the customer, the results may be reported in a simplified way. Any information that is not reported to the customer shall be readily available.

Annex A
(informative)

Practical laboratory experiences with the amplifiability of *cytb* and *cox1* segments from tested fish species

Examples for fish species where laboratory experiences for amplification of target sequences exist are listed in Table A.1. Results for further fish species can be found in the Barcoding Table of Animal Species (BaTAnS) at www.bvl.bund.de.[11]

Table A.1 — Practical laboratory experiences

Latin name	*cytb*	*cox1*
Acipenser baerii	+	+
Acipenser gueldenstaedtii	+	/
Anarhichas spp.	+	+
Anguilla anguilla	+	+
Anguilla australis	+	+
Anguilla japonica	+	+
Anguilla rostrata	+	+
Aphanopus spp.	+	+
Argyrosomus regius	+	+
Atherestes stomias	+	+
Boops boops	+	+
Brosme brosme	+	+
Centroberyx affinis	–	+
Chelidonichthys cuculus (syn. *Aspitrigla cuculus*)	+	+
Chelidonichthys spp.	+	+
Chiloscyllium punctatum	/	+
Clarias macrocephalus	+	+
Clupea harengus	+	–
Conger conger	+	+
Coregonus spp.	+	/
Ctenopharyngodon idella	+	+
Cynoglossus browni	+	–
Cynoglossus senegalensis	+	±
Cynoglossus spp.	+	/
Cypselurus hiraii	+	+
Dagetichthys lusitanicus (syn. *Synaptura lusitanica)*	±	+

Key

+ positive result
/ no data available
– negative result
± inconsistent result