FINAL DRAFT

INTERNATIONAL STANDARD

ISO/FDIS 22949-1

ISO/TC **34**/SC **16**

Secretariat: ANSI

Voting begins on:
2021-08-07

Voting terminates on:
2021-10-02

Molecular biomarker analysis — Methods of analysis for the detection and identification of animal species in foods and food products (nucleotide sequencing-based methods) —

Part 1:
General requirements

Analyse moléculaire de biomarqueurs — Méthodes d'analyse pour la détection et l'identification d'espèces animales dans les aliments et les produits alimentaires (méthodes basées sur le séquençage des nucléotides) —

Partie 1: Exigences générales

Reference number
ISO/FDIS 22949-1:2021(E)

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 documents 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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

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 ISO/TC 34, *Food products*, Subcommittee SC 16, *Horizontal methods for molecular biomarker analysis*.

A list of all parts in the ISO 22949 series can be found on the ISO website.

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

This document provides general guidance for deoxyribonucleic acid (DNA) sequencing for animal species detection or identification, or both, in food and feed products. DNA sequencing is the process of determining the order of the four nucleotide bases (adenine, guanine, cytosine and thymine) in a nucleic acid polymer. Nucleic acid polymers can range in length from a few nucleotides to hundreds of millions of bases.

Rapid DNA sequencing methods have been successfully validated and verified for detection and identification of animal species in food products[1]. Within the food industry, rapid, economical and high throughput access to whole genome sequences in foods has improved the control of both food quality and safety[2]. Two types of DNA sequencing methods are most widely used for food products: chain termination and high throughput sequencing[3][4][5][6][7][8].

Chain termination developed in 1977 by Frederick Sanger still bears his name. Sanger sequencing reactions can be prepared manually and electropherograms can be read directly by the user. Automated base calling capillary electrophoresis systems have mostly replaced manually read gels and rapid Sanger applications are being developed.

High throughput or next generation sequencing (NGS), including next generation short-read and third generation long-read methods, has reduced the cost of DNA sequencing, improved sequence readability and automated most of the steps from preparatory to bioinformatics. High throughput automated DNA sequencing applies base/wavelength specific fluorescence or ionic detection to determine the real-time enzymatic addition of nucleotides to a DNA template.

Sanger sequencing (dideoxy chain termination) generates high quality data for determining a single DNA sequence of an individual target. NGS by contrast, can be used to assess millions of individual DNA fragments of mixed markers and targets at the same time.

Sanger sequencing and NGS can both be used to verify animal species composition in a food sample or compare DNA sequence results to previously defined databases to identify its animal origin, or both[9].

Molecular biomarker analysis — Methods of analysis for the detection and identification of animal species in foods and food products (nucleotide sequencing-based methods) —

Part 1: General requirements

1 Scope

This document specifies general requirements for DNA sequencing method performance in the detection and identification of animal species in food and feed products. Performance requirements are limited to Sanger and next generation sequencing (NGS), including second and third generation sequencing, for analysis of single species products and multispecies products.

This document is applicable to DNA sequences for mammals, birds, fish, molluscs, crustaceans, amphibians, reptiles and insects, and to the validation of the applicable methods.

Methods for DNA species quantification are not considered under the scope of this document.

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 — Terms and definitions*

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*

ISO 21571, *Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — Nucleic acid extraction*

3 Terms and definitions

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

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

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

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

4 Performance characteristics of the methods

4.1 General

The identity of an organism at the species level is established by DNA sequencing through the comparison of DNA sequences obtained from samples with known reference sequences. Results can

also be obtained at multiple taxonomic levels (e.g. order, species, genus, family), depending on the type of database and DNA data analysis performed.

Methods used for animal species identification by DNA sequencing shall meet the performance guidance provided in this document. The method validation process described in this document, interlaboratory collaborative or single-laboratory studies or both shall be used to determine and elucidate sequencing method performance characteristics.

Single species products that have been produced from one piece of meat (e.g. fish fillet, beef tenderloin) are most appropriately analysed by Sanger sequencing, while NGS is the appropriate method for simultaneous multispecies identification.

4.2 Fitness for purpose of the method

The method shall be fit for purpose for the identification of organisms and their taxonomic relationship with other organisms. Identification at the sub-specific levels (e.g. breeds) can be included provided that the databases and bioinformatic analyses are available. Information regarding the applicability and limitations of the method shall be documented sufficiently to fulfil the criteria described in this document. The appropriate commercial DNA sequencing platform(s) suitable for use in all method procedures should be determined and identified. The target DNA can be located on the nuclear genome, but also on the mitochondrial genomes. A DNA barcoding or metabarcoding approach can be used based on one or more allelic regions to be sequenced[10][11].

4.3 Scientific basis

4.3.1 General

A general workflow of the laboratory analytical procedure is provided in Annex A.

4.3.2 Sanger sequencing

Sanger sequencing is based on a chain termination method for determining the nucleotide sequence of DNA. Elongation of the replicated strand of DNA by DNA polymerase is interrupted by the incorporation of non-elongatable nucleotides that terminate replication successively at each position of the sequenced fragment. Labelling the four non-elongatable nucleotides (dideoxy-A, G, C and T) permits determination of the molecular size of the replicated fragment by electrophoresis.

NOTE 1 Sanger sequencing can only be used when a sample contains tissue from a unique single species (e.g. one fish fillet, one steak, one shrimp), because multiple DNA target sequences in the same sample from more than one species can produce more than one end labelled chain terminated product of the same chain length. Multiple chain terminated fragments of the same chain length can present two or more terminated bases at each position and appear electrophoretically as convoluted or overlapping sequences. The electropherogram for the reaction is unlikely to be readable and the identification will be compromised.

Automated Sanger sequencing platforms have been available for many years and have appropriately been used for DNA barcoding to identify a single species, e.g. for fish species identification[12][13][14].

NOTE 2 For an example, see CEN/TS 17303:2019[15].

4.3.3 Next generation sequencing

NGS in combination with DNA metabarcoding enables the robust and reliable identification of species in complex food products (containing multiple species). NGS is a commercial term used to reference high throughput DNA sequencing methods (see Annex B)[16]. It is alternatively referred to as “massively parallel sequencing”. Results usually take the form of a text file with millions of individual sequence reads. Commercially available platforms support second and third generation NGS sequencing methods.

Second generation sequencing is based on the analysis of short DNA fragments generated by a polymerase chain reaction (PCR) or a fragmentation process to produce fragments of up to 600 bp in

size. Third generation sequencing is capable of analysing much longer DNA fragments (e.g. bacterial genomes) and consequently is sometimes referred to as “long-read DNA sequencing”.

4.4 Units of measurement

Sequence reads generated by Sanger sequencing or NGS methods are compared bioinformatically to previously determined and identified reference sequences in DNA databases. Percent similarity between the sample and reference sequences is calculated. Two sequences are homologous if they share more similarity than would be expected by chance ($p \leq 0{,}01$).

NOTE A percent similarity greater than 98 % is usually sufficient to identify sequences as the same species.

Identification data shall be provided as a qualitative result, in the form of list of appropriate taxa, their respective percent similarity and, as appropriate, the expected value.

4.5 Applicability

4.5.1 Meat or food product considerations

The method should contain a definition of the food or food product(s) that will be sequenced and, where appropriate, the raw materials from which it is derived and any necessary references to manufacturing processes. In addition, whether the sample(s) will be composed of a single species or mixed species or both.

4.5.2 Sampling plan

Representative samples taken through an applicable sampling plan should be provided to the sequencing laboratory. Appropriate methods for subsampling laboratory samples should continue to ensure the representativeness of the sequencing result. If applicable, statistical criteria for acceptance or rejection of the analytical result should include the sampling plan and subsampling. Requirements and limitations for the laboratory sample size and number of individual items forming the sample, and the handling of the sample and its storage, should be evaluated and described based upon:

— the degree of processing of the sample constituents;

— the different species and animal tissue types involved;

— the nature of the sample matrices;

— the preparation of the sample matrix for analysis.

4.5.3 Genomic considerations

When assessing whether a method is fit for purpose, the following aspects regarding the nature of the genomic region to be sequenced should be considered:

— the cellular and genomic disposition of the sequence, i.e. cellular: nuclear or mitochondrial, or genomic: allelic or gene specific;

— the length of the sequence that will be determined;

— the types and design of primer sequences used for sample and library preparation and sequencing.

4.5.4 Method testing

Applicability shall be tested by extracting DNA from matrix-matched test samples representative of the analytical scope of the method.

Highly processed food products can have fragmented DNA or low DNA concentration or both, making them difficult to sequence. To compensate, some strategies can be employed, e.g. process more samples, concentrate DNA extracts, work with smaller targeted genomic regions.

4.6 Primer selection and evaluation

Primer selection and evaluation is the first and most important step in the nucleotide sequencing analytical workflow when a PCR-based DNA sequencing technology is used. The design of new primers is a bioinformatics task performed using applicable software. There are many commercial and freely available primer selection tools, e.g. Primer-BLAST and Primer 3. The following primer design factors affect primer performance: secondary structures, partial sequence matches, hairpins, GC content, etc.

NOTE 1 Primer design and selection is not considered as part of the workflow to select genomic regions to be sequenced in third generation NGS technology and the DNA fragmentation strategy, e.g. second generation based NGS, shotgun sequencing.

Universal (consensus) primers can be defined for any taxonomic level (species, genus, family, order, class, domain, etc). The design of these primers is based on DNA sequence alignments and identification of consensus regions that are almost identical amongst the taxa to be identified. Degenerate primers can be designed to increase the universality of the primers. Determining the number of degenerations included in a single primer to avoid non-specific binding requires evaluation of unspecific annealing.

Primers should be evaluated in silico irrespective of their use in the laboratory workflow. They shall be assessed using appropriate bioinformatics tools for their match with the species (taxon) to be detected/identified.

NOTE 2 An example of an in silico workflow for primer set selection is shown in Annex C.

The final evaluation of the performance of each set of primers should be done by applying the laboratory workflow using reference DNA or tissue material (whenever it is available) representative of the identifiable group of species (taxa).

4.7 Database construction and evaluation

The DNA database used for species detection or identification or both will determine the number of different species identifiable by the laboratory workflow. Database construction and evaluation is a bioinformatics task and many different tools can be used, including in-house developed software. Publicly available databases can also be used, e.g. GenBank. The databases should have the following information available:

— the species list;

— the number of different entries for each species;

— the DNA region(s), gene(s) and length of DNA sequences;

— a measure of the validation quality of the sequence entry(ies).

The final performance evaluation of a bioinformatics search of a database shall be done by applying the laboratory workflow using reference material (whenever possible) that is representative of the species included in the database (vouchered). The use of higher taxonomic levels is appropriate when the identification at the species is inconclusive, i.e. when the sample is not differentiable by a more targeted search. In GenBank, the designation "*Bos* sp." can be used as the search term for all of the species of the genus *Bos*.

4.8 Selectivity and nucleotide sequence specificity

4.8.1 General

Excessive amplification of non-targeted species DNA can interfere with the detection of targeted species DNA, e.g. increased limit of detection (LOD) or critical value or both. Targeted (PCR-based) NGS and Sanger sequencing methods should provide experimental evidence for nucleotide specificity with non-targeted species (see ISO 20813). The degree of interference (selectivity) from non-targeted DNA should be determined. The nucleotide sequence specificity should be assessed in a two-step procedure: theoretical and experimental evaluation of the inclusivity and exclusivity. Experimental evaluation is done by PCR with the selected primers. For NGS, the performance of each set of primers should also be evaluated with adapters or barcodes or both. The minimum concentration of DNA required for specific sequence identity analysis can be determined as the probability of detection (POD) (see ISO/TS 16393) [17]. Because sequence data are used for the verification of animal speciation results, these data should be based on evaluated databases with due consideration of the timing of submission of individual entries and any subsequent changes in taxonomic classification or naming (i.e. provenance). In cases of unexpected results, further investigation should be carried out to confirm reference material identity. DNA amplification is not required when using a non-PCR based sequencing analysis, e.g. shotgun sequencing.

4.8.2 Requirements for inclusivity testing

Experimental results from testing a method with a target animal taxon should be provided. This testing should include relevant species for the scope of the method and the taxonomic level covered by the primers to be used. An appropriate number of different species should be used (see ISO 20813). For primers with less than 100 % sequence homology with target DNA species (based on in silico evaluation), the amplifiability of those species should be tested in the laboratory to define the limits of acceptable mismatches. The user shall define a theoretical list of species that can be identified and the list of species that were tested in the laboratory. It is known that depending on the DNA region that is targeted by the primers there can be very different conservation levels. Different genes have different rates of intraspecific DNA sequence variability. Appropriate numbers of individuals from each species should be tested accordingly.

Material for experimental inclusivity testing should contain sufficient DNA concentration for PCR and DNA sequencing, as described in ISO 20813. Replicates of each sample material shall be tested with controls appropriate to the different workflow steps, e.g. PCR. If present, sequence variants of the target animal species should be identified with comparable amplification efficiency.

4.8.3 Evaluation of non-targeted DNA interference

Experimental results from testing the sequencing method with non-target animal species shall be provided. This testing should include both taxonomically close and not-closely-related animal species. Animal species or taxonomic groups relevant to the scope of the method should be tested, e.g. species commonly used in food in general and particularly in matrices considered in the scope of the method. The method should clearly distinguish between target and non-target animal species. In silico analysis should be performed with the appropriate non-target taxa to theoretically evaluate the sequence homology of the universal primers. Sequence mismatches are good indicators of specificity. They indicate the degree to which a set of primers and probe will bind to unintended sequences to produce a false-positive result. A minimum number of mismatches should be defined as acceptable for in silico analysis. Exclusivity is determined more effectively when a higher number of mismatches is permitted in the analysis. Select an appropriate number of species that can cause interference with the target animal species present in the food test material (as described in ISO 20813). Examples of suitable organisms are listed in ISO 20813:2019, Annex A. Other species should be included if relevant, e.g. if there are sequence homologies of oligonucleotides to nucleic acid sequences. The suitability of the DNA used for amplification should be confirmed by using appropriate controls. Sufficient DNA should be used for experimental PCR exclusivity testing.