INTERNATIONAL STANDARD

ISO 21286

First edition
2019-03

Soil quality — Identification of ecotoxicological test species by DNA barcoding

Qualité du sol — Identification des espèces par code-bare ADN dans les essais d'écotoxicologie

Reference number
ISO 21286:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
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 on 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 the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 190, *Soil quality*, Subcommittee SC 4, *Biological characterization*.

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

Currently, test species identification is usually based on morphological characters. However, this does not always give clear results because

a) few taxonomic experts are available,

b) closely related species can differ by a few, easily overlooked characters, and

c) even more importantly, several test species are in fact complexes of cryptic species.

A good example is the compost worm *Eisenia fetida/andrei* (used in ISO 11268-1, ISO 11268-2 and ISO 17512-1), in which morphological traits alone may not be sufficient to discriminate between both species[5][36]. Another well-known case is the predatory mite, *Hypoaspis* (*Geolaelaps*) *aculeifer*[50], which might get confused with *H. miles*, widely used in biological pest control[31].

Species misidentifications, the use of a morphospecies which is actually a complex of cryptic species, or even species mixing in lab cultures, can be a serious problem for the reliability of the ecotoxicological tests. Sibling species in a morphospecies complex can exhibit ecological, behavioural, and physiological differences, and can differ also in their response to toxicants (e.g. References [2], [17], [35], [40]). This also seems to be the case of the springtail *Folsomia candida* (used in ISO 11267 and ISO 17512-2), in which considerable levels of genetic differentiation have been found among natural populations of *F. candida* and among laboratory strains[9][19][41]. Although different laboratory strains have been found to exhibit only minor differences in the sensitivity towards some chemicals[12][9], other studies have detected significant variation in phenmedipham avoidance behaviour and divergent fitness responses to cadmium exposure among genetically differentiated strains[14][30]. Moreover, even if two species have similar responses to toxicants, the presence of two species within the same laboratory culture can result in the production of sterile hybrids, which will bias the outcome of reproduction tests[36].

Implementing species identification via DNA barcoding can help to overcome these obstacles, ensuring that the species or strain used for testing is well characterized. As a result, quality assurance can be improved, making the results obtained by different ecotoxicological laboratories far more reliable and comparable. For *Eisenia fetida/E. andrei* this work, including an international ringtest, has already been performed[36], see Annex A. The conclusions of this ringtest can be summarized as follows.

— DNA barcoding is a reliable and practical method for identifying *Eisenia* species.

— Only 17 out of 28 ecotoxicological laboratories were correct in their taxonomic assignment. Most laboratories with wrong or unknown assignments actually have *E. andrei* in stock.

— The existence of a cryptic species pair within *E. fetida* is a plausible hypothesis.

— It is important that earthworms used for ecotoxicological tests are regularly (re-)identified by DNA barcoding.

Very probably, similar experiences and recommendations can be drawn for other invertebrates species used in terrestrial ecotoxicology, as well as plants. Indeed, DNA barcoding has proven to be useful for specimen identification and species delimitation in many organism groups, including other earthworms[13][37], enchytraeids[16], mites[15], collembolans[32], molluscs[42], nematodes[28] and terrestrial plants[8].

Soil quality — Identification of ecotoxicological test species by DNA barcoding

1 Scope

This document specifies a protocol to identify ecotoxicological test specimens (mainly invertebrates and plants) to the species level, based on the DNA barcoding technique. This protocol can be used by laboratories performing DNA barcoding in order to standardize both the wet-lab and data analysis workflows as much as possible, and make them compliant with community standards and guidelines.

This document does not intend to specify one particular strain for each test method, but to accurately document the species/strain which was used.

NOTE 1 This does not imply that DNA barcoding is performed in parallel to each test run, but rather regularly (e.g. once a year, such as reference substance testing) and each time a new culture is started or new individuals are added to an ongoing culture.

This document does not aim at duplicating or replacing morphological-based species identifications. On the contrary, DNA barcoding is proposed as a complementary identification tool where morphology is inconclusive, or to diagnose cryptic species, in order to ensure that the results obtained from different ecotoxicological laboratories are referring to the same species or strain.

This document is applicable to identifications of immature forms which lack morphological diagnostic characters (eggs, larvae, juveniles), as well as the streamline identification of specimens collected in field monitoring studies, where large numbers of organisms from diverse taxa are classified.

NOTE 2 In principle, all species regularly used in ecotoxicological testing can be analysed by DNA barcoding. Besides the earthwoms *Eisenia fetida* and *E. andrei*, further examples for terrestrial species are *Lumbricus terrestris*, *L. rubellus*, *Allolobophora chlorotica*, *Aporrectodea rosea,* and *A. caliginosa*, *Dendrodrilus rubidus, Enchytraeus albidus,* and *E. crypticus* (Haplotaxida); *Folsomia candida*, *F. fimetaria, Proisotoma minuta,* and *Sinella curviseta* (Collembola); *Hypoaspis aculeifer* and *Oppia nitens* (Acari); *Aleochara bilineata* and *Poecilus cupreus* (Coleoptera); *Scathophaga stercoraria, Musca autumnalis* (Diptera) or *Pardosa* sp. (Arachnida). Nematodes or snails and even plants can also be added to this list.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions 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/

3.1
amplicon
specific DNA product generated by *PCR* (3.5) using one pair of *PCR primers* (3.6)

3.2
DNA barcode
unique pattern of DNA sequence that identifies each species

3.3
electropherogram
trace file
combination of a graphical representation of a Sanger DNA sequence composed of colour-coded peaks with each colour corresponding to one nucleotide

Note 1 to entry: They are automatically supplied by DNA sequencing programs.

3.4
Phred quality score
Q score
quality measure used to assess the accuracy of a sequencing reaction

Note 1 to entry: This quality measure indicates the probability that a given base is called incorrectly by the sequencer. Phred scores are on a logarithmic scale. Therefore, if Phred assigns a Q score of 30 (Q30) to a base, this is equivalent to the probability of an incorrect base call 1 in 1 000 times. A lower base call accuracy of 99 % (Q20) will have an incorrect base call probability of 1 in 100, meaning that every 100 base pairs sequencing read will likely contain an error.

3.5
polymerase chain reaction
PCR
molecular biology technique for rapidly synthesising multiple copies of a given DNA segment by using a DNA polymerase and an oligonucleotide primer pair

3.6
PCR primer
short oligonucleotides (usually 15 to 30 nucleotides in length) that allow PCR amplification of DNA between specific sites

Note 1 to entry: The two primers (a forward and a reverse) are base-paired to the top and bottom strand of the template DNA, and their 3'-OH ends are in convergent direction.

4 Principle

DNA barcoding is a molecular method that uses a short and standardized DNA region (the DNA barcode) as a genetic tag for species-level identification[22].

Since its inception in 2003 and the launch of the Barcode of Life project, DNA barcoding has systematically been applied not only to biological research, but also to several industrial fields where a correct identification of biological materials is essential, such as the food industry. For example, it is helping to detect fraud in herbal medicinal products[29], and it has been adopted by the Food and Drug Administration (FDA) for seafood and fish identification[21][44]. In fact, DNA barcoding is likely to become a routine test in many fields, in particular in food quality control and traceability[20].

Briefly, the goal of DNA barcoding is:

a) to obtain the nucleotide sequence of a standardised DNA region from an unidentified sample (a test specimen),

b) to compare that sequence with known sequences in a reference database by using bioinformatic methods, and

c) based on such comparison, to identify the sample to the species level.

Therefore, DNA barcoding cannot be a useful identification tool without a reliable and comprehensive reference database, which includes enough samples of each species from across its geographic range to account for intraspecific variability. Also, DNA barcoding relies on the premise that sequences in this barcode region are more similar between members of a species than to sequences of any other species (the so called barcode gap). Therefore, before applying DNA barcoding, a species delimitation study

of the target organismal group should have been carried out to assess its efficacy for discriminating species.

It is essential that the DNA barcoding method is carried out by trained staff. On the one hand, trained laboratory technicians are needed to optimize the wet-lab protocols for each organismal group. On the other hand, the wet-lab pipeline needs to be supervised by scientists trained in genomics and systematics. These scientists should also be in charge of the electropherogram and/or raw DNA sequence file analysis and species assignment.

5 Reagents and material

5.1 Biological material

Adequate specimen preservation is a critical factor to obtain good-quality DNA from samples. Whenever possible, specimen samples for DNA barcoding should be taken from freshly harvested or fresh-frozen tissue. Exposure to preservation agents such as ethyl acetate or formaldehyde should be avoided, as they destroy DNA.

Freezing at –80 °C or in liquid nitrogen (–196 °C) is the preferred method for long-term storage of tissue samples. DNA in dried specimens generally remains stable for at least one year, but degradation becomes increasingly problematic over time[23].

Ethanol-preserved material is easily analysed when fresh, but DNA will slowly become acidified and degraded unless ethanol is regularly refreshed or buffered. For proper tissue preservation, use an ethanol concentration of 95 % to 99 %, and ensure that the volume of ethanol is at least three times greater than the volume of tissue. In order to maintain the ethanol concentration to at least 95 %, it is necessary to replace the ethanol solution within the first days (at least three days) after sampling, and tightly seal the vial to avoid evaporation[23]. A combination of low temperatures (–20 °C) and ethanol will help preserve the samples for long-term storage and helps prevent degradation during thawing and re-freezing cycles.

As a general rule, DNA barcoding analysis should follow tissue collection as soon as possible, but specimens adequately preserved and stored for several months will perform well in DNA extraction[21][23].

5.2 Enzyme

Taq Polymerase from *Thermus aquaticus* is standard for PCR. Hot start Taq polymerases and/or high fidelity DNA polymerases have been shown to offer a high performance in DNA barcoding, allowing for greater amplification sensitivity and increased ease of reaction setup than standard polymerases (http://ccdb.ca/resources/).

Alternatively, pre-optimised commercial master mixes may be used. These consist of a premixed, ready-to-use solution containing Taq DNA polymerase, dNTPs, $MgCl_2$ and reaction buffers at optimal concentrations for efficient amplification of DNA templates in routine PCR.

5.3 Oligonucleotide PCR primers

For Oligonucleotide PCR primers, see 8.3.2 and 8.3.3.

5.4 Reagents

5.4.1 **Nuclease-free water molecular grade water** (dd H_2O).

5.4.2 **TE buffer (Tris-EDTA buffer)**, 1-fold, pH 8,0.

Dissolve 1 ml of 1 mol/l Tris base (pH 8,0), 0,2 ml EDTA (0,5 mol/l) in 98,8 ml of molecular grade water. Adjust the pH to 8,0 with concentrated HCl.

5.4.3 Deoxynucleoside triphosphates (dNTPs).

5.4.4 PCR buffer, without Mg (500 mmol/l KCl, 100 mmol/l Tris-HCl, pH 8,3 at 25 °C).

Buffer is usually supplied with each enzyme as a 10-fold or fivefold concentrate. Use only the buffer supplied with each particular enzyme.

5.4.5 Magnesium chloride, $MgCl_2$.

5.4.6 PCR additives (optional): trehalose dihydrate, bovine serum albumin (BSA), formamide, dimethyl sulfoxide (DMSO).

5.4.7 Agarose (analytical grade, standard melting temperature).

5.4.8 TAE (gel-running buffer), 50-fold stock solution, pH 8,3.

Dissolve 242 g of Tris base [tris(hydroxymethyl)aminomethane], 57,1 ml of glacial acetic acid (17,4 mol/l), 100 ml of 500 mmol/l EDTA solution (pH 8,0) in 842,9 ml of molecular grade water.

5.4.9 Size standard 100 base pair (bp) DNA ladder, a commercially available molecular-weight marker suitable for sizing double-stranded DNA from 100 to 1 000 base pairs during gel electrophoresis.

5.4.10 6-fold Loading buffer, 3 ml of 100 % glycerol, 0,025 g of bromophenol blue, 0,025 g of xylene cyanol FF in 7 ml of molecular grade water.

5.4.11 Ethidium bromide solution (0,5 µg/ml) or any safer alternative nucleic acid stain.

5.4.12 PCR purification kit, either using enzymatic reactions, magnetic beads or silica-membrane-based cleanup.

5.4.13 5-fold Sequencing buffer (400 nm Tris-HCl, pH 9,0, 10 mmol/l $MgCl_2$).

5.4.14 BigDye® Terminator v3.1 cycle sequencing kit[1)].

5.4.15 Pop-7 Polymer for 3730 DNA analyzers[1)].

5.4.16 3730 DNA analyser capillary array, 50 cm[1)].

5.4.17 GeneScanTM 500 LIZTM DYE Size Standard[1)].

5.4.18 Highly deionized formamide.

6 Apparatus

The usual laboratory equipment, including micropipettes, centrifuge, and the following specific equipment.

6.1 Spectrophotometer, to measure the concentration and purity of double-stranded DNA at 260 nm.

1) This protocol has been validated using the 3730 DNA Analyzer capillary electrophoresis system and the BigDye terminator chemistry. They are registered trademarks of Applied Biosystems. 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.

6.2 Laminar flow hood.

6.3 PCR thermal cycler.

6.4 Horizontal electrophoresis system.

6.5 Electrophoresis power supply.

6.6 Gel documentation system.

6.7 Automated DNA sequencing system, for DNA sequencing (e.g. 3730 DNA analyser, Applied Biosystems)[2].

7 General requirements

7.1 Experimental precaution and contamination avoidance

Good laboratory practice and specific anti-contamination strategies are necessary to minimize the chance of contamination during the DNA isolation and PCR steps, either from DNA previously handled in the laboratory, amplicon carry-over from previous PCR assays, or sample-to-sample cross-contamination.

There are some basic steps to be followed to prevent exogenous contamination and sample carry-over: work on a clean surface, wear gloves, and use disposable or sterilised instruments. If possible, laboratories should use separate rooms – or, at least, different benchtops and work spaces of the laboratory – for template extraction, PCR reagent preparation, and amplification. Work shall always flow from the cleanest to the dirtiest area. Each work area should have dedicated supplies and reagents, as well as lab coats and gloves. It is recommended that the setting up of PCR reactions is performed in a laminar flow hood. It is also necessary to use sterile plastic-ware and aerosol-resistant filtered pipette tips.

When handling multiple specimens, care shall be taken to avoid cross-contamination between samples. Anything (gloves, surfaces) that comes in contact with one sample shall be discarded or cleaned before proceeding with the next one by wiping it with 1 % bleach solution and then thoroughly rinsing it with water. Tools used to collect a fragment of tissue (tweezers, scissors, scalpels, etc.) shall be sterilized before and after handling each sample. To do so, soak the instruments into ethanol (70 % to 96 %) and then hold them briefly over the flame of a Bunsen burner or a lighter to burn off the alcohol.

7.2 Safety precautions

7.2.1 Chemical hazards

WARNING 1 — Ethidium bromide (EtBr) staining is commonly used to visualize DNA in agarose gels. EtBr is a potential mutagen and is a skin, eye, and respiratory irritant. Avoid direct skin contact, wear nitrile gloves, and use adequate eye protection. All staining of gels should be done in a designated area in the laboratory, and all EtBr waste should be disposed of in labelled containers.

2) This protocol has been validated using the 3730 DNA Analyzer capillary electrophoresis system and the BigDye terminator chemistry. They are registered trademarks of Applied Biosystems. 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.

WARNING 2 — In the past few years, diverse alternatives to EtBr (e.g. GreenSafe®, Sybr® Safe, GelRed™)[3] have become commercially available. These products are generally considered to be less hazardous than EtBr. They are, however, still mutagenic and should also be handled and disposed of with care.

WARNING 3 — Formamide used in sequencing reactions causes eye, skin, and respiratory tract irritation. It is a possible developmental and birth defect hazard. When handling formamide, wear appropriate protective eyewear, clothing, and gloves.

7.2.2 Physical hazards

Electrophoresis experiments present a potential electrical hazard if used incorrectly. Electrophoresis units and their power supplies shall be routinely inspected to ensure that they are working properly, that wires, leads, and connectors are undamaged and properly insulated, and that buffer tanks have no cracks or leaks. During electrophoresis, since any wet surface can become conductive, it is recommended not to touch any part of the equipment (tank, wires) while the power supply is on.

Ultraviolet (UV) transilluminators are often used to visualize fluorescent dyes used in gel electrophoresis and pose potential exposures to UV radiation. Laboratory coats, gloves and appropriate glasses and face visor shall be worn if there is any risk of exposure, especially when using unshielded transilluminators.

8 Procedure

8.1 DNA isolation

This process can either be carried out by using a commercial DNA isolation kit or by following a standard DNA isolation protocol. The DNA isolation kit or protocol can be selected depending on the specific features of the sample. For instance, for fresh or recently collected tissue a Chelex-based DNA release method usually provides enough DNA for DNA barcoding. However, for archival samples or very small specimens, more sensitive approaches should be used (e.g. silica membrane-based DNA extraction methods)[23]. Using commercial kits can be helpful, as the pre-prepared reagents facilitate the standardization of the technique.

In the case of minute organisms (i.e. length < 0,5 cm) — such as collembolans, mites, or small insects — the entire specimen may be processed. For larger organisms, various body parts can be selected, including muscle biopsies, abdomen, legs, antennae, or eggs. When possible, external body surfaces, the digestive tract and body parts containing hairs, bristles, or hardened exoskeletons, should be avoided.

For earthworms, potworms, and other annelids, it is recommended to keep the specimens for about 24 h on moist filter-paper before tissue sampling, so their gut contents are evacuated.

DNA isolation from plant tissue can be challenging due to their high content in complex polysaccharides, polyphenols, and other secondary metabolites that can affect DNA quality and inhibit downstream reactions. Compounds such as CTAB (cetyl trimethylammonium bromide) and PVP (polyvinylpyrrolidone) can aid in removing polysaccharides and polyphenols, respectively, and are therefore widely used in plant DNA isolation buffers.

Detailed protocols for DNA isolation from either animal or plant tissues can be found in Reference [43] and at the DNA barcoding website of the DNA Learning Center (http://www.dnabarcoding101.org/).

A negative control (with no tissue) shall be run in parallel with all batches of sample extraction, in order to detect DNA contamination of the analytical reagents or sample-to-sample contamination. This

3) GreenSafe®, Sybr® Safe, GelRed™ are registered trademarks of NZYTech, Thermo Fisher Scientific, and Biotium, respectively. 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.

negative control shall be processed through the DNA isolation and PCR steps, in parallel with the test samples.

8.2 Quantification

After DNA isolation, the yield and purity of the DNA sample should be determined by either

a) using a spectrophotometer equipped with a UV lamp, to measure sample absorbance at 260 nm, or

b) using a fluorometer that employs fluorescent DNA-binding dyes to specifically quantify double-stranded DNA.

Typically, the DNA sample is diluted to 10 ng/µl using molecular grade water and should be stored at −20 °C for extended periods, or at 4 °C until use.

8.3 PCR

8.3.1 Target genomic region

The target genomic region (or DNA barcode) varies across taxa. The standard DNA barcode for almost all animal groups is a 658-base pairs region of the mitochondrial cytochrome c oxidase 1 gene (COI). For land plants, a combination of the chloroplastic genes *mat*K and *rbc*L has been recommended as the plant DNA barcode[8], although the success of species discrimination using this combination can be limited in certain plant groups.

Please note that the usual DNA barcode may not be sufficient in particular cases. In these cases, additional genomic regions may be analysed.

8.3.2 Primer design

A critical factor in the PCR stage is the selection of appropriate primer pairs. Therefore, care shall be taken when choosing or designing optimal primers.

The first step is to check the literature and databases (e.g. the BOLD Primer Database, available at http://www.boldsystems.org/index.php/Public_Primer_PrimerSearch) for existing primers that have been successfully used to amplify the DNA barcode in the target organism group.

If there are no public primers available, or the selected primer pairs do not work on the target species, then new primer pairs may be designed. Sequences of the same genomic region belonging to the same or related taxa, publicly available in the BOLD database (http://www.boldsystems.org/) and the National Center for Biotechnology Information (NCBI) GenBank database (https://www.ncbi.nlm.nih.gov/genbank), can be used for primer design. Free online resources such as Primer-BLAST[45] are a useful tool to find primers specific to the target genomic region. This program also checks primer specificity against a user-selected database, to confirm that the designed primers do not amplify genomic regions other than the specific target region.

8.3.3 Primer synthesis

Primers may be ordered from a primer supplier who specializes in the synthesis of oligonucleotides. A standard desalting purification is sufficient for the primers used in DNA barcoding.

The primers are usually shipped and delivered in a lyophilized state in nanomoles quantity. Before PCR, they should be centrifuged and resuspended in TE buffer or molecular grade water to make a 100 µmol/l stock. Alternatively, primers can be ordered already resuspended at a given concentration (typically between 10 µmol/l to 100 µmol/l) from several manufacturers. In either case, resuspended primers shall be dispensed into single-use aliquots and stored at −20 °C.

8.3.4 PCR

For standard PCR, reactions are set up in sterile, thin-walled 0,2 ml tubes or an equivalent PCR plate format. The final concentrations of reagents for a typical PCR reaction, with a final volume of 25 µl, are as follows (see Table 1).

Table 1 — Final concentrations of reagents for a typical PCR reaction, with a final volume of 25 µl

Reagent	Initial concentration	Volume µl	Final concentration
PCR buffer	10-fold	2,5	1-fold
$MgCl_2$	50 mmol/l	0,75	1,5 mmol/l
dNTPs	2,5 mmol/l	2,0	0,2 mmol/l
Forward primer	0,1 mmol/l	0,125	0,5 µmol/l
Reverse primer	0,1 mmol/l	0,125	0,5 µmol/l
Taq polymerase	5 units/µl	0,125	0,025 units/µl
Template DNA	10 ng/µl	2,5	1 ng/µl
Nuclease-free sterile water		16,875	

Always add water and buffer first and Taq polymerase last.

When multiple PCR reactions are set up simultaneously, it is useful to prepare a master mix that contains all the reagents except for DNA template. This reduces preparation time and the potential for pipetting errors. Make up enough master mix for the number of reactions to be amplified, and an extra volume to allow for pipetting errors (as a rule of thumb, count one extra reaction every 10 samples).

Addition of PCR-enhancing agents in the PCR mix can increase the yield of the specific PCR product and overcome the effect on PCR inhibitors present in the reaction. Some of these enhancers are trehalose, bovine serum albumin (BSA), betaine, formamide or DMSO (e.g. References [18] and [39]). Alternatively, pre-optimised commercial master mixes may be used.

The high sensitivity of PCR makes it especially vulnerable to trace contamination. Therefore, it is necessary to use sterile tips (and preferably filter tips) for pipetting the PCR reagents. DNA templates (and other PCR products) shall also be kept away from the PCR reagents while setting up the PCR reaction mix. In addition, and besides the DNA isolation negative control, every set of reactions shall include a PCR negative control, which contains all reagents except for the DNA template. This PCR negative control is used to detect DNA contamination in the PCR reagents.

PCR is performed in a thermocycler (PCR block), according to the following program:

— one cycle of 4 min at 94 °C;

— 35 cycles of: 30 s at 94 °C, 40 s at specific annealing temperature (*Ta*) for the PCR primers, 1 min at 72 °C;

— a final extension step at 72 °C for 5 min.

NOTE Primers T_a is usually set at 2 °C to 6 °C lower than their melting temperature, usually specified by the primer supplier. Alternatively, T_a can be estimated using an online calculator (http://tmcalculator.neb.com). Not only the T_a, but also other PCR cycling parameters can be modified depending on the nature of the template DNA or the DNA polymerases (see manufacturers' recommendations).

Once completed the PCR reactions can be stored in the fridge at 4 °C. For long-term storage (i.e. more than a week), freezing at –20 °C is recommended.

8.4 Checking the amplicon size

PCR success can be checked by agarose gel electrophoresis.

— Prepare a 0,5 % to 2 % agarose gel in TAE buffer and add 1 µl of nucleic acid stain per each 10 ml of gel solution. A 0,7 % gel will show good separation (resolution) of large DNA fragments (5 kb to 10 kb), while a 2 % gel will show good resolution for small fragments (0,2 kb to 1 kb).

— Pipette 4 µl of each PCR product to a new tube, and add 1 µl of DNA loading buffer to each 4 µl aliquot.

— Load 5 µl of each sample into the pre-cast gel wells.

— Include the appropriate size standard in one well.

— Run the electrophoresis and visualize the results under UV light in a gel documentation system.

A successful PCR reaction should yield a single sharp amplicon (of approximately 700 base pairs, in the case of COI). Both the DNA isolation and the PCR negative controls should contain no bands.

Photograph and keep a picture of the gel (electronic and/or hard copy) for records.

8.5 Purification

The PCR products can require a purification step, in order to prepare good quality DNA templates for sequencing. This step is intended to eliminate unspecific bands which may have been observed in the gel. Unincorporated nucleotides and residual primers are also eliminated in this step.

Besides traditional ethanol precipitation, there are varied commercial kits for PCR product purification, which can be based on an enzymatic reaction, magnetic beads or silica-membrane columns.

8.6 Sequencing

The PCR products or the purified PCR products may be sequenced with the same primers used for PCR.

Depending on the sequencing platform and reagents used, this step may need some optimization. Assuming that most laboratories do not have access to Sanger sequencing equipment in-house, it is recommended to send the purified PCR products to an external company, where the sequencing service will be performed by professionally trained staff. The requirements of the sequencing services in terms of volume and concentration of PCR products and sequencing primers should be checked with the service provider. Typically, an automated capillary electrophoresis system will produce a sequence read of circa 750 base pairs.

Whenever possible, PCR products should be sequenced bidirectionally. Bidirectional sequencing enables the generation of full length barcode sequences by avoiding low-quality base calls which generally occur towards the end of the reads. Based on a bi-directional sequencing, two electropherograms or trace files (generally .ab1 files) are obtained per DNA barcode.

The electropherogram files or other applicable raw file format shall be saved and stored, since they serve as quality control for the original DNA barcode sequences.

8.7 Bioinformatics

8.7.1 General

The electropherogram files or other applicable raw file can be opened, quality-checked, and edited using DNA analysis software packages. For an example, see http://www.geneious.com[25].

8.7.2 Electropherogram or raw sequence quality checking

The bidirectional sequencing strategy enables automatic sequence assembly using the DNA analysis software package.

Nonetheless, visual inspection and manual editing of the electropherograms is still necessary to verify sequence quality. Discrimination of closely related species using DNA barcodes often relies on differences in one or a very few base-pair sites. Therefore, electropherograms quality checking is a critical factor to ensure the accuracy and reliability of the resulting DNA barcode sequences[7].

If sequencing quality is high, the majority of the electropherogram should consist of a series of clear peaks, corresponding to the signal from each of the nucleotides in the DNA sequence. If sequencing has failed, or contamination is present, the peaks will look weak and/or it will be impossible to clearly resolve a single peak at each nucleotide position. If this is the case, the results are not high quality enough for use.

The Geneious software implements a "Base Calling" tool which detects peaks in the electropherograms and assigns the most probable base at each position. It may also assign a quality measure for such a call, in terms of the expected probability of making an erroneous call (Phred quality scores or Q scores).

According to the Database Working Group (DBWG) of the Consortium for the Barcode of Life (CBOL), Q scores > 20 are generally considered to be high quality base calls, and scores > 30 are very high quality[7]. In editing single reads, base calls with quality score of less than 20 should be recorded as "N". There shall be less than 10 ambiguous bases of 500 to validate the sequence analysis [see Clause 10, c)].

Should a new generation of sequencing technology be used, the Phred-scaled quality score should be considered in a similar manner, and alignments to reference sequences can be visualized in a genome browser to ensure accuracy.

8.7.3 Trimming of low-quality regions and primers sequences

Basic information on the process of sequence editing and assembly, common sequence errors, and standards for producing high quality barcodes can be found in the BOLD Handbook[3].

Errors in base calling usually occur at the beginning and the end of the electropherogram, where the signal intensity is weakest. Therefore, low-quality regions (i.e. $Q < 20$) at both ends of each file can simply be deleted.

Primer-annealing regions should also be trimmed from each sequence. This can be done by referring to the primer sequences and removing them from both ends of the sequence. Sometimes it is not possible to recognize the PCR primers in an electropherogram. To ensure that it is trimmed at the correct nucleotide position, a sequence with the correct length from a closely related species can be downloaded from the GenBank or BOLD database and aligned to the original trace. Using the GenBank/BOLD sequence as a reference, the trace can be trimmed to the same starting and ending point.

Ambiguous nucleotides can also arise within the middle of the sequences. These can simply be overwritten with the letter "N" (indicating uncertainty about the call), but should not be deleted.

8.7.4 Sequence overlapping

Since each of the samples has been sequenced in both the forward and reverse direction, these complementary sequences can be assembled into a longer consensus sequence.

Bidirectional sequencing also allows checking the accuracy of the base calls in the overlapping region. The forward and reverse sequences should match perfectly with no mismatching nucleotides. Moreover, any ambiguous nucleotides ("N") in one sequence can now be resolved using the complementary sequence. Base calls of high or very high quality in one direction should be maintained over those with lower quality in the other read. If, however, this is not possible an "N" can be used at the position where the sequences mismatch[7].

8.7.5 Sequence verification

The sequence obtained may be compared against the GenBank database in order to verify that the appropriate genomic region has been sequenced and that the sample belongs to the appropriate organismal group.

8.7.6 Reviewing the edited sequence

Once the consensus sequence has been generated, it should be translated into amino acids to detect stop codons. Most DNA analysis software recognizes and flags the presence of stop codons in a sequence, which should be not present in COI barcode sequences[3]. They can be the result of a shift in the reading frame introduced during the editing process (i.e. by accidentally adding or removing a base). This type of error can usually be corrected by inspecting the original electropherograms to determine the location of the mistake.

However, stop codons can also signal the amplification of a nuclear pseudogene derived from mitochondrial DNA or numt (nuclear copies of mtDNA). If no accidental shifts in the reading frame are detected, sequences with stop codons should be excluded from the analysis, as they confuse DNA barcoding.

When a sequence is translated into amino acids, it is important to ensure that the correct genetic code is being used. Most invertebrates have a generic "invertebrate mitochondrial" code. However, vertebrates and plants have their own specific code. If the wrong code is used, false stop codons can appear in the sequence.

Finally, the edited sequences should also be inspected for the presence of indels (insertions/deletions). Indels will be represented as gaps in a multiple sequence alignment. Therefore, indels can be detected by comparing the DNA barcode sequences from closely related species. Indels can occur naturally in the genome, or can be the result of incorrect sequence editing. In a protein-coding region, naturally occurring indels can be identified if

a) the number of inserted/deleted bases is a multiple of 3, and

b) the translated alignment results in the correct amino acid sequence (i.e. without stop codons or frameshifts)[3].

8.7.7 Species assignment

In this last step, the consensus sequence obtained is compared against the DNA barcoding reference database using appropriate software.

The reference database should contain sequences from voucher specimens from all the species that the sample can belong to, a priori. International databases such as the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/GenBank/) or the Barcode of Life Database (BOLD, (http://www.boldsystems.org/) allow access to all public DNA barcode sequence data. The BOLD database contains mainly metazoan COI sequences, whereas the GenBank database contains nucleotide sequences of taxa across all domains of life.

Species assignments can be performed by submitting the query sequence to the search engines BOLD Identification System (BOLD-IDS)[34] and NCBI's BLAST[1]. Both search engines employ the BLAST algorithm to perform pairwise global alignments between each query sequence and each sequence in the reference database. From each pairwise alignment, an "overlap identity" percentage is calculated for the region in common between two sequences. The likely species identity of the query sequence is then determined by picking the database sequence with highest overlap identity to the query (i.e. best match identification *sensu* Meier et al. 2006[27]). The results are reported in a rank list, in which the efficiency of species identification is measured by different statistics: the E-value, Max score, and Total score (in BLAST) or the maximum identity (in BOLD). The closer a hit approaches 100 % in sequence identity, or the E-value to 0, the better the identification efficiency.

However, for many taxonomic groups, reference libraries are still incomplete. Some of the species regularly used in ecotoxicological testing (e.g. *Enchytraeus crypticus* or *Sinella curviseta*) are poorly represented or even absent from those reference databases. In these cases, DNA barcoding can be heavily biased by Type II errors (misidentifications of queries that lack conspecific database entries) and therefore, species-level identifications need to be confirmed by qualified taxonomists.

A query that is not represented by a conspecific DNA barcode in the reference library will be erroneously identified according to the most similar non-conspecific reference barcode. The number of these false positive identifications can be greatly reduced by using the modified assignment criterion "best close match"[27]. With this criterion, species names are only assigned when the distance between the query and its best DNA barcode match is below a given distance threshold value. For instance, the BOLD Identification Systems uses an arbitrary distance threshold of 1 %. Yet, since no universal distance threshold is applicable to all taxonomic groups, it is advisable to use an ad hoc distance threshold that is calculated for each particular reference library of DNA barcodes, based on the expected separation between intra-specific and interspecific distances (i.e. the local barcode gap). Software such as TaxonDNA/SpeciesIdentifier 1.7.8[27], the R packages "adhoc"[38] and SPIDER[4], Automatic Barcode Gap Discovery (ABGD)[33] or the Barcode Gap Analysis tool in the BOLD workbench[34] can be used to calculate the ad hoc distance threshold.

In brief, firstly, uncorrected p-distances are calculated between all sequence pairs and used to estimate intra- and interspecific genetic distances. Secondly, the maximum intraspecific and minimum interspecific distances are calculated for each sequence. The distance threshold for the data set is set at a value below which 95 % of all intraspecific pairwise distances are found.

Briefly, the "best close match" approach, as implemented in TaxonDNA/SpeciesIdentifier 1.7.8, comprises the following steps.

a) Uncorrected p-distances are calculated between all sequence pairs and used to estimate intra- and inter-specific genetic distances.

   NOTE Uncorrected p-distances (number of base differences per site) are preferred over the widely used Kimura-two-parameter (K2P) model because the latter seems to be a poorly fitting model for distance measures between closely related COI sequences[10].

b) The maximum intraspecific and minimum interspecific distances are determined for each sequence.

c) The distance threshold for the data set is set at a value below which 95 % of all intraspecific pairwise distances are found.

d) The query sequence is assigned to the species with the reference DNA barcode with the lowest distance within the estimated threshold value. If multiple species have equally small distance matches, the result is considered ambiguous. If the distance to the most similar sequence(s) is outside the threshold level, the query is classified as "no match".

Besides the described distance-based methods, DNA barcoding studies have widely used tree-based approaches for species assignment[11]. These approaches focus on whether the query sequence clusters together with sequences originating from a single species or not. The 'Taxon ID Tree' functionality in BOLD allows for the generation of dendrograms from the multiple alignment of the query and database sequences using the Neighbour-joining (NJ) algorithm. Although it has been well documented that NJ trees perform poorly for specimen identification purposes (see References [6] and [11] and references therein), they still provide a useful graphical summary of the data. The tree can be saved as a PDF file, as a permanent record of the results.

8.7.8 Quality of the reference databases

The success of DNA barcoding is highly dependent on the accuracy and quality of the sequences available in the reference databases. Both GenBank and BOLD are built by direct submissions from researchers and organisations across the globe, and they are constantly being updated with new data. However, since these databases rely on the submitters' information, they may contain low-quality

sequences or sequences from misidentified specimens, which can constrain the identification success of DNA barcoding (see Reference [26]).

Sequence quality standards, however, are expected to be higher in BOLD. Most records included in this database are linked to a voucher specimen in a biological collection, a tissue sample in a biorepository, collection data, a specimen photograph, and sequence trace files. This allows straightforward traceability of the data back to the original source.

Although DNA barcodes submitted to BOLD are not subjected to any centralized review, BOLD employs several tools to identify data anomalies or low quality records. If any potential errors are detected (e.g. a contaminated sequence or a misidentified species) the submitter is informed and the sequence is flagged. In addition, any specimen record will not gain formal barcode status until it meets the Barcode Compliance Standards (species name, voucher data, collection records, identifier of the specimen, minimum sequence length of 500 bp, less than 1 % ambiguous bases, high-quality trace files)[33].

Therefore, when possible, the BOLD Identification System and a "barcode-compliant" reference sequences should be used for species assignment. This will ensure that the identification results are based on high quality and reliable data.

9 Calculation and expression of results

The results of a DNA barcoding analysis shall include the following.

— A table with the 20 top matching records in the reference database against the query sequence, with the corresponding overlap identity values and the distance threshold used in the 'best close match' analysis. The source and the number of sequences included in the reference database should be specified here.

— The multiple alignment of the query and database sequences and the resulting phylogenetic tree showing sequence clustering.

— A clear statement of the result of the species assignment (i.e. Unambiguous identification/ Ambiguous/No match) and identity of the query sequence at the lowest taxonomic level possible. If sample identification is ambiguous (e.g. two different species have 100 % similarity to the query sequence because they are closely related), further testing can be required to successfully identify them. In such circumstances, laboratories should indicate on their report that the sample was identified to a higher taxonomic level (genus, family), and/or indicate the two species creating ambiguity.

10 Validity of the test

The results are considered to be valid if the following conditions are met in the different experimental steps.

a) DNA isolation: all samples should yield ≥ 5 ng/µl (with an elution volume of at least 30 µl) of good-quality DNA (260/280 nm ratio of approximately 1,8).

   The negative control with no tissue should give a negligible reading during quantification (<1 ng/µl).

b) PCR: an optimized PCR reaction should yield a single sharp amplicon of the expected size, with no spurious amplification products when examined on an agarose gel. If this is not achieved, PCR may require further optimization (i.e. change the annealing temperature and/or alter the concentration of PCR reagents such as magnesium, dNTPs and primers).

   Neither the DNA isolation nor the PCR negative controls should yield any band in the PCR reactions. If any detectable PCR products are observed in the negative controls, the sources of contamination should be sought out and eliminated before proceeding further.

c) Sequence analysis: if the complete fragment of the COI barcode is amplified, bidirectional sequences should be at least 500 base pairs in length, with <2 % ambiguous bases in the consensus, and no stop codons.

11 Test report

The test report shall include the following information:

a) a reference to this document, i.e. ISO 21286:2019;

b) a full description of the experimental design and procedures, including a description of the equipment, reagents (with batch numbers) and bioinformatics software used;

c) a full description of the experimental design and procedures, including DNA isolation, PCR amplification, PCR product purification, sequencing and post-sequencing analysis;

d) quantification measures of the DNA samples (after DNA isolation) and the DNA isolation negative control;

e) record of the agarose gel (electronic and/or hard copy) showing the result of PCR amplification;

f) raw electropherogram files and edited nucleotide sequences;

g) final results of the analysis (as detailed in Clause 9);

h) discussion of the results;

i) any details not specified in this document or which are optional, as well as any incident which may have affected the results.

Annex A
(informative)

Eisenia Barcoding Initiative: A ring test to evaluate the applicability of DNA barcoding for the identification of *Eisenia* species

The *Eisenia* Barcoding Initiative (EBI) was a project launched in 2013 by 3 partners. As part of the EBI, a ring test was carried out by five DNA barcoding laboratories to analyse samples of *E. fetida*, *E. andrei*, and *Eisenia sp*. from 28 ecotoxicological laboratories throughout the world. This ringtest was supported by the Global Soil Advisory Group (GSAG) of the Society of Environmental Toxicology and Chemistry (SETAC).

The main goals of this work were:

a) to evaluate whether COI sequencing is a robust method for molecular characterization of *Eisenia*;

b) to assess the consistency of results among DNA barcoding laboratories;

c) to check the *Eisenia* species / strains currently used by ecotoxicology laboratories.

A total of 144 coded specimens of *E. fetida*, *E. andrei*, and *Eisenia* sp. were provided by 28 ecotoxicological laboratories from 15 countries on four continents. The specific identity of each specimen was indicated by the laboratories, according to the identifications performed by external experts, suppliers, or scientists of the respective institutions. Each individual worm was photographed and the posterior part of each worm was divided into five pieces (one for each DNA barcoding laboratory). The anterior part of the body was kept as voucher specimen.

DNA barcoding was performed in parallel in five laboratories from Germany, Spain, Canada and Belgium. All steps of the sample preparation were described by Standard Operating Procedures (SOP), which covered the following steps in detail: selection and preservation of the earthworms to be analysed, transport of the worm samples to the laboratories for DNA extraction, amplification, and sequencing of the standard mitochondrial COI gene fragment. General guidelines (SOP) for the DNA barcoding procedure were distributed among the laboratories prior to the experiment. Each DNA barcoding laboratory provided a report describing their methodologies and results (including results files).

A total of 668 out of the 720 samples (144 in five laboratories) were successfully DNA barcoded (overall success rate 92,8 %). Success rates differed among laboratories and varied between 75,7 % and 98,6 %. The inter-laboratory variation of the results was extremely low. From the possible 647,321 pairwise nucleotide site comparison, 91 showed discrepancies (0,000 14). Base calling differences among laboratories can have been due to different interpretations by the researchers, the different sequencer models used, different sequencing chemistries, different array lengths, differences in the dye set, and different software versions at the sequencing facilities involved. However, the extremely low amount of sequence discrepancies and the consistency of the conclusions among laboratories suggest that such DNA barcoding studies can be performed by any standard sequencing laboratory.

The analysis of the COI sequences (581 bp) obtained revealed three distinct haplotype clusters: one including only *E. andrei* sequences and two with only *E. fetida* sequences, referred to as *E. fetida* 1 and *E. fetida* 2. The attribution of the individual worms to these three clusters was completely consistent among the five DNA barcoding laboratories, which demonstrates that DNA barcoding is a reproducible and robust method for the identification of *Eisenia* specimens.

Comparisons with the prior taxonomic assignments of the ecotoxicological laboratories showed that, whereas specimens of the molecular *E. fetida* clusters were always identified morphologically as *E. fetida*, some specimens of the molecular *E. andrei* cluster were identified morphologically as *E. fetida*. Only 17 of the 28 laboratories (61 %) provided correct identification of their laboratory stocks. Most laboratories with wrong or unknown assignments actually had *E. andrei* in culture, or a mixture of

both species. This finding highlights the need to verify the identity of the taxa used in ecotoxicological testing.

Remarkably, the mean p-distance between *E. fetida* 1 and *E. fetida* 2 was 0,112, a COI divergence level which is usually indicative of species level differentiation in earthworms. Therefore, the existence of a cryptic species pair within *E. fetida* is therefore a pausible hypothesis in need of further investigation.

The results of this ring test are available in Reference [36]. Since all COI sequences obtained in this work are linked to morphologically identified voucher specimens, they constitute a valuable curated reference library for future studies aimed at identifying *Eisenia* specimens through DNA barcoding. The DDBJ/EMBL/GenBank accession numbers for these COI sequences are listed in Table A.1.

Table A.1 — List of the DDBJ/EMBL/GenBank accession numbers of the *Eisenia* COI sequences used in the study with their respective taxonomic designation

Accession number	Identical haplotypes	Accession number	Identical haplotypes
AY874511.1|Eisenia andrei	AY874509.1|Eisenia andrei	JN869997.1|Eisenia andrei	JN870070.1|Eisenia andrei
	AY874507.1|Eisenia andrei		JN870074.1|Eisenia andrei
	AY874506.1|Eisenia andrei		JN870078.1|Eisenia andrei
	AY874505.1|Eisenia andrei		JN870081.1|Eisenia andrei
	AY874504.1|Eisenia andrei		AY874493.1|Eisenia andrei
	AY874497.1|Eisenia andrei		AY874494.1|Eisenia andrei
	JN870088.1|Eisenia andrei		AY874495.1|Eisenia andrei
	JN870087.1|Eisenia andrei		AY874496.1|Eisenia andrei
	JN870086.1|Eisenia andrei		AY874498.1|Eisenia andrei
	JN870085.1|Eisenia andrei		AY874500.1|Eisenia andrei
	JN870084.1|Eisenia andrei		AY874502.1|Eisenia andrei
	JN870083.1|Eisenia andrei		AY874503.1|Eisenia andrei
	JN870082.1|Eisenia andrei		AY874512.1|Eisenia andrei
	JN870080.1|Eisenia andrei		
	JN870079.1|Eisenia andrei	AY874513.1|Eisenia fetida	AY874514.1|Eisenia fetida
	JN870077.1|Eisenia andrei		
	JN870076.1|Eisenia andrei	AY874515.1|Eisenia fetida	AY874516.1|Eisenia fetida
	JN870075.1|Eisenia andrei		AY874517.1|Eisenia fetida
	JN870073.1|Eisenia andrei		AY874518.1|Eisenia fetida
	JN870072.1|Eisenia andrei		AY874519.1|Eisenia fetida
s	JN870071.1|Eisenia andrei		AY874520.1|Eisenia fetida
	JN870069.1|Eisenia andrei		AY874521.1|Eisenia fetida
	JN870068.1|Eisenia andrei		AY874522.1|Eisenia fetida
	JN870067.1|Eisenia andrei		AY874523.1|Eisenia fetida