Bulletin OEPP/EPPO Bulletin (2021) **51** (1), 100–143 ISSN 0250-8052. DOI: 10.1111/epp.12724

European and Mediterranean Plant Protection Organization
Organisation Européenne et Méditerranéenne pour la Protection des Plantes PM 7/129 (2)

Diagnostics

PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests

Specific scope

This Standard describes the use of DNA barcoding protocols in support of identification of a number of regulated pests and invasive plant species comparing DNA barcode regions with those deposited in publicly available sequence databases.[1]

It should be used in conjunction with PM 7/76 *Use of EPPO diagnostic protocols*.

Specific approval and amendment

First approved in 2016–09.
Revised in 2020–10.

1. Introduction

DNA barcoding is a generic diagnostic method that uses a short standardized genetic marker in an organism's DNA to aid identification at a certain taxonomic level. The chosen marker region should reflect the group taxonomy of the target species. Therefore, the marker region should provide a high interspecific variability and low intraspecific differences, and should enable the identification of as many species as possible belonging to a shared higher taxonomical level such as genus, family or order (e.g. Chen *et al*., 2013). An organism is identified by finding the closest matching reference record. The first genetic marker to be described as a 'barcode' was the mitochondrial cytochrome c oxidase I (*COI*) gene, which is used for species identification in the animal kingdom (Hebert *et al*., 2003). Later the chloroplast large subunit ribulose-1,5-bisphosphate carboxylase-oxygenase (*rbcL*) gene (Hollingsworth *et al*., 2009) and the nuclear ribosomal internal transcribed spacer (ITS) region (Schoch *et al*., 2012) were proposed as barcodes for the plant and fungi kingdoms, respectively.

The use of a single barcode region does not provide sufficient reliability for the identification of the majority of regulated pests. Therefore, several short standardized genetic markers have been identified as 'barcodes' for identification at the required taxonomic level in several pest groups.

DNA barcoding protocols for eukaryotes and prokaryotes (a novelty in the DNA barcoding field) were developed and validated within the Quarantine Organisms Barcoding of Life (QBOL) Project financed by the 7th Framework Programme of the European Union. Within the DNA barcoding EUPHRESCO II project, test protocols for several quarantine pests and invasive plant species were added, and the use of polymerases with proofreading abilities was introduced to minimize the risk of polymerase chain reaction (PCR) errors. In addition, amplification primers were M13-tailed when possible to improve the user-friendliness of the protocols, allowing the generation of sequence data with a minimum number of sequencing primers. Additional protocols have been added in the second version of the Standard based on national projects conducted at Austrian Agency for Health and Food Safety (AGES, AT), Flanders Research Institute for Agriculture, Fisheries and Food (ILVO, BE) and Netherlands Food and Consumer Product Safety Authority (NVWA, (NL).

Regulated organisms are identified by finding the closest matching reference record using a combination of Basic Local Alignment Search Tool (BLAST) hit identity, multilocus sequence analysis (MLSA) and clustering in species-specific clades using multiple databases containing sequence data of regulated organisms and related species. Pest species in this Standard were selected on the basis of their pest status, economic impact, availability of material

[1]Use of brand names of chemicals or equipment in these EPPO Standards implies no approval of them to the exclusion of others that may also be suitable.

and pre-existing knowledge of loci with sufficient resolution.

This EPPO Standard describes the DNA barcoding protocols developed for the identification of a number of regulated arthropods, bacteria, fungi and oomycetes, invasive plant species, nematodes and phytoplasmas. Each organism group is covered in a separate appendix. Protocols describe the extraction of nucleic acids and the amplification of short standardized marker(s). Since the identification of regulated pests is often based on several different markers, diagnostic schemes are provided to aid the selection of appropriate protocols. When more than one marker is necessary, the markers are either used in parallel for species identification (e.g. invasive plant species and phytoplasmas) or a single marker is first used for genus identification (e.g. 16S rDNA for bacteria) and, depending on the genus, a second marker (sometimes in parallel with a third marker) is used for identification to species level. For some *Xanthomonas* bacteria a third marker is needed for identification at the pathovar level. For each identification based on several markers all consensus sequences produced need to be analysed in an MLSA, which can be done in EPPO-Q-bank (https://qbank.eppo.int/). The generation of sequence data, assembly of raw sequence data and analysis of consensus sequences using BLAST and MLSA in online databases is discussed in Appendix 7. Appendix 8 provides an example of a sequencing analysis report that can be used to collate all relevant data and Appendix 9 provides information on synthetic positive amplification controls (PACs).

It should be noted that the outcome of DNA barcoding tests can be negatively affected by the incompleteness of databases, incorrectly identified species in databases, the amplification of pseudogenes or nuclear mitochondrial DNAs (NUMTs) and introgression or hybridization events. For that reason, the analysis of sequence data should be performed by proficient operators. DNA barcoding is consequently used in support of identification at a certain taxonomic level. Origin, host plant and other characteristics (e.g. morphological, biochemical, reactions on indicator plants) are typically needed to complete the diagnosis. Video tutorials of barcoding training sessions produced in the framework of the PRACTIBAR Euphresco project (Giovani *et al*., 2020) are available on YouTube (https://www.youtube.com/playlist?list=PLoVf4Pt04Db53pUVTI8qwcWkWgUgg46gm).

2. Reference material

A single synthetic PAC per organism group can be used to assess the efficiency of the PCR amplification. It can also be used as a standardized process control from amplification until sequence analysis and will give insight into the repeatability and reproducibility of each test (see also Appendix 7, section 5.2). The synthetic PACs presented in Appendix 9 were designed in such a way that all tests in one appendix [except test 2.4 Conventional PCR *egl* *Ralstonia solanacearum* species complex (Appendix 2) and test 2.8 *TEF1* (Appendix 3)] can be monitored using a single control. When amplified, the synthetic PACs yield amplicons ranging from 560 to 720 base pairs, depending on the primers used. When sequenced, the synthetic PACs can easily be identified since, after translation of the nucleic acid sequence (reading frame 1, standard code), the following amino acid sequence is obtained twice: *KEEP*-CALM*THIS*IS*MERELY*A*VERY*STRANGE*REFERENCE*PHRASE*WITH*EIGHTY*FIVE*CHARACTERS (stop codons are indicated as *). Synthetic PAC sequences are presented in Appendix 9 and are available from the NCBI: PAC arthropods v.1 (KT429638), PAC bacteria v.1 (KT429643), PAC fungi v.1 (KT429642), PAC invasive plant species v.1 (KT429639), PAC nematodes v.1 (KT429641) and PAC phytoplasmas v.1 (KT429640), and can be ordered from commercial companies producing synthetic genes or gBlocks (e.g. ThermoFisher, IDT, Biomatik).

3. Feedback on this Diagnostic Protocol

If you have any feedback concerning this Diagnostic Protocol, or any of the tests included, or if you can provide additional validation data for tests included in this protocol that you wish to share please contact diagnostics@eppo.int.

4. Protocol revision

An annual review process is in place to identify the need for revision of Diagnostic Protocols. Protocols identified as needing revision are marked as such on the EPPO website. When errata and corrigenda are in press, this will also be marked on the website.

Acknowledgements

This protocol was originally drafted by BTLH van de Vossenberg, M Westenberg and M Botermans, Dutch National Plant Protection Organization, PO Box 9102, 6700 HC Wageningen, the Netherlands, J Hodgetts, Fera Science Limited, Sand Hutton, York YO41 1LZ, UK, and B Cottyn, Institute for Agricultural and Fisheries Research, Plant Sciences Unit, Crop Protection, Burgemeester van Gansberghelaan 96, bus 2, 9820, Merelbeke, Belgium. The first revision was drafted by T. Warbroek, Dutch National Plant Protection Organization, PO Box 9102, 6700 HC Wageningen, the Netherlands, B Cottyn and J van Vaerenbergh, Institute for Agricultural and Fisheries Research, Plant Sciences Unit, Burgemeester van Gansberghelaan 96, 9820, Merelbeke, Belgium, and R Gottsberger, Austrian Agency for Health and Food Safety (AGES), Department for Molecular Diagnostics of Plant Diseases, Spargelfeldstr. 191, 1220 Vienna. It was reviewed by the EPPO Panel on Diagnostics and Quality Assurance. Most of the DNA barcoding protocols in this standard were developed, optimized

and validated in an international test performance study within the QBOL Project financed by the 7th Framework Program of the European Union and the DNA Barcoding EUPHRESCO II Project.

References

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J *et al*. (2013) GenBank. *Nucleic Acids Research* **41**, D36–D42.

Carbone I & Kohn LM (1999) A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* **91**, 553–556.

Chen W, Djama ZR, Coffey MD, Martin FN, Bilodeau GJ, Radmer L *et al*. (2013) Membrane-based oligonucleotide array developed from multiple markers for the detection of many *Phytophthora* species. *Phytopathology* **103**, 43–54.

Coenye T, Falsen E, Vancanneyt M, Hoste B, Govan JRW, Kersters K *et al*. (1999) Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *International Journal of Systematic Bacteriology* **49**, 405–13.

Edwards U, Rogall T, Blöcker H, Emde M & Böttger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research* **17**, 7843–53.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294–299.

Germain JF, Chatot C, Meusnier I, Artige E, Rasplus J-Y & Cruaud A (2013) Molecular identification of *Epitrix* potato flea beetles (Coleoptera: Chrysomelidae) in Europe and North America. *Bulletin of Entomological Research* **103**, 354–362.

Giovani B, van de Vossenberg B, Hanner R, Radulovici A & Petter F (2020) Use of barcoding, from theory to practice (PRACTIBAR). *Zenodo*. https://doi.org/10.5281/zenodo.3648171

Groenewald Jz, Nakashima C, Nishikawa J, Shin H-d, Park J-h, Jama An *et al*. (2013) Species concepts in Cercospora: spotting the weeds among the roses. *Studies in Mycology* **75**, 115–170.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W & Hebert PD (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 968–971.

Hajri A, Hunault G, Lardeux F, Lemaire C, Manceau C, Boureau T *et al*. (2009) A "Repertoire for Repertoire" Hypothesis: Repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS One* **4**, e6632.

Hebert PDN, Cywinska A, Ball SL & deWaard JR. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313–321.

Hebert PD, Penton EH, Burns JM, Janzen DH & Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812–14817.

Hollingsworth P M, Forrest L l, Spouge J l, Hajibabaei M, Ratnasingham S, van der Bank M *et al*. (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12794–12797.

Holterman M, Holovachov O, van den Elsen S, van Megen H, Bongers T, Bakker J *et al*. (2008) Small subunit ribosomal DNA-based phylogeny of basal Chromadoria (Nematoda) suggests that transitions from marine to terrestrial habitats (and vice versa) require relatively simple adaptations. *Molecular Phylogenetics and Evolution* **48**, 758–763.

Holterman M, van der Wurff A, Van den Elsen S, Van Megen H, Bongers T, Holovachov O *et al*. (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular Biology and Evolution* **23**, 1792–1800.

Hu M, Hoglund J, Chilton NB, Zhu XQ & Gasser RB (2002) Mutation scanning analysis of mitochondria cytochrome c oxidase subunit 1 reveals limited gene flow among bovine lungworm subpopulations in Sweden. *Electrophoresis* **23**, 3357–3363.

Jones SJ, Hay FS, Harrington TC & Pethybridge SJ (2011) First report of Boeremia blight caused by *Boeremia exigua* var. *exigua* on Pyrethrum in Australia. *Plant Disease* **95**, 1478.

Kress WJ & Erickson DL (2007) A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS One* **2**, e508.

Kress W j, Erickson D l, Jones F a, Swenson N g, Perez R, Sanjur O *et al*. (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18621–18626.

Lemey P, Marco Salemi M & Irvine Vandamme A (2009) *The Phylogenetic Handbook, A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edn., Cambridge University Press, Cambridge (GB).

Makarova O, Contaldo N, Paltrinieri S, Kawube G, Bertaccini A & Nicolaisen M (2012) DNA barcoding for identification of 'Candidatus Phytoplasmas' using a fragment of the elongation factor Tu gene. *PLoS One* **7**, e52092.

O'Donnell K, Kistler HC, Cigelnik E & Ploetz RC (1998) Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 2044–2049.

Oliveira LSS, Harington TC, Freitas RG, McNew D & Alfenas AC (2015) *Ceratocystis tiliae* sp. nov., a wound pathogen on *Tilia americana*. *Mycologia* **107**, 986–995.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J & Stead D (2007) Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **57**, 2881–2887.

Ratnasingham S & Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* **7**, 355–364.

Richert K, Brambilla E & Stackebrandt E (2005) Development of PCR primers specific for the amplification and direct sequencing of gyrB genes from microbacteria, order Actinomycetales. *Journal of Microbiological Methods* **60**, 115–123.

Robideau GP, De Cock AW, Coffey MD, Voglmayr H, Brouwer H, Bala K *et al*. (2011) DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Molecular Ecology Resources* **11**, 1002–1011.

Sang T, Crawford DJ & Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* **84**, 1120–1136.

Schoch Cl, Seifert KA, Huhndorf S, Robert V, Spouge Jl, Levesque CA *et al*. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6241–6246.

Sjölund J. (2017) Non-destructive DNA extraction from Psyllids. https://www.ponteproject.eu/protocols-calsol/non-destructive-dna-extraction-psyllids/ [accessed on 18 April 2020].

Tate JA (2002) Systematics and evolution of *Tarasa* (Malvaceae): an enigmatic Andean polyploid genus. Ph.D. dissertation. The University of Texas at Austin.

White TJ, Bruns T, Lee S & Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*(eds. Innis MA, Gelfand DH, Sninsky JJ & White TJ), pp. 315–322. Academic Press, San Diego (US).

Wicker E, Grassart L, Coranson-Beaudu R, Mian D, Guilbaud C, Fegan M *et al*. (2007) *Ralstonia solanacearum* strains from Martinique (French west indies) exhibiting a new pathogenic potential. *Applied and Environmental Microbiology* **73**, 6790–6801.

Appendix 1 – DNA barcoding Arthropods

The tests below are described as they were carried out to generate the validation data provided in section 4 (unless stated otherwise) in combination with the guidance provided in Appendix 7. Other equipment, kits or reagents may be used provided that verification (see PM 7/98) is carried out.

1. General information

1.1 This appendix describes the protocols used to identify selected regulated arthropods by conventional PCR followed by Sanger sequencing analysis. Table 1 shows the regulated organisms that have successfully been tested with the protocols described in this section. It is very likely that other regulated arthropods can also successfully be identified using these protocols, but to date validation data has not been generated to support this.

1.2 Protocols in sections 2.2 and 2.4 in this appendix were developed by INRA (FR) as part of the QBOL Project financed by the 7th Framework Programme of the European Union (2009–12). The protocols were further optimized by the Fera Science Limited (GB) as part of the EUPHRESCO II DNA Barcoding project (2013–14). The protocol in section 2.3 in this appendix was developed by the Division of Tropical Ecology and Animal Biodiversity, Department of Botany and Biodiversity Research, University of Vienna and further optimized by the Austrian Agency for Health and Food Safety, Department for Molecular Diagnostics of Plant Diseases (AGES/MDPD).

1.3 The mitochondrial *COI* gene test described in section 2.2 in this appendix is used for species identification of selected regulated arthropods (see Fig. 1 and Table 1). Other tests to generate *COI* amplicons can be used and are described in sections 2.3 and 2.4 in this appendix.

1.4 Primer sequences, amplicon sizes and thermocycler settings are provided in the test-specific sections. HPLC-purified primers should be ordered to avoid nonspecific PCR amplification.

1.5 Reaction mixes are based on the MyFi™ Mix (Bioline) or on the 5× HOT FIREPol Blend Master Mix with 15 mM $MgCl_2$ (Solis Biodyne) reagents. The protocols described in sections 2.2 and 2.4 were initially validated using the BIO-X-ACT™ Short Mix (Bioline), which is no longer available. Verification performed in NVWA showed that the MyFi™ Mix and BIO-X-ACT™ Short Mix performed equally (Van de Vossenberg, pers. comm.).

1.6 Molecular-grade water is used to set up reaction mixes; this should be purified (deionized or distilled), sterile (autoclaved or 0.22 µm filtered) and nuclease-free.

1.7 Amplification is validated for Peltier-type thermocycler with heated lid, e.g. C1000 (Bio-Rad) except for the test in section 2.3 where a Biometra T 3000 thermocycler was used.

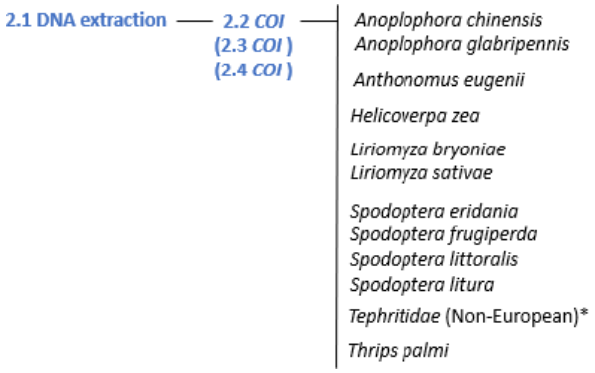


Fig. 1 Diagnostic testing scheme for identification of regulated arthropods using DNA barcodes. The steps shown refer to the sections in this appendix which should be followed to reach reliable identification of the corresponding taxa. *Several non-European *Tephritidae* sequences are available in the EPPO-Q-bank.

Table 1. Regulated arthropods successfully identified with barcoding protocols

Regulated organism	Test		
	2.2 *COI*†	2.3 *COI*†	2.4 *COI**
Anoplophora chinensis	x	x	
Anoplophora glabripennis	x	x	
Anthonomus eugenii	x	x	
Helicoverpa zea	x	x	
Liriomyza bryoniae	x	x	
Liriomyza sativae	x	x	
Spodoptera eridania	x	x	
Spodoptera frugiperda	x	x	
Spodoptera littoralis	x	x	
Spodoptera litura	x	x	
Tephritidae (Non-European)‡	x	x	
Thrips palmi	x	x	

†The test described in section 2.2 in this appendix needs to be performed to reach reliable identification of the corresponding taxa. The test described in section 2.3 in this appendix can be used as an alternative.

*In some cases, the *COI* tests using primers LCO1490 and HCO2198 (section 2.2) or LepF and LepR (section 2.3) can fail to produce an amplicon. In those cases, the *COI* test described in section 2.4 can be used alternatively. During validation, the 2.4 *COI* test was only used to identify nonregulated species, which explains why no cross is indicated in the table.

‡Several non-European *Tephritidae* sequences are available in EPPO-Q-bank.

2. Methods

2.1 Nucleic acid extraction and purification

2.1.1 Tissue material (typically 10–50 mg) of all life stages of a single specimen is used as input for DNA extraction.

2.1.2 DNA is extracted using the DNeasy Blood & Tissue kit (Qiagen) according to the "Purification of Total DNA from Animal Tissues" protocol.

2.1.3 When tissue material is stored in ethanol, all the ethanol should be removed prior to DNA extraction. To achieve this, the insects can be transferred for a few minutes to a dry filter paper and may be further dried in a SpeedVac centrifuge to facilitate evaporation of the solvent.

2.1.4 Grinding of the tissue material in a lysis buffer (provided), e.g. with a micro pestle prior to DNA extraction, can be performed but is not required in order to allow nondestructive DNA extraction (Sjölund, 2017).

2.1.5 When the specimen is crushed, the sample should be incubated at 56°C for at least 1 h. For nondestructive extraction, the specimen in lysis buffer should be incubated overnight at 56°C in a slow shaking heat block.

2.1.6 DNA is eluted in 200 µL of pre-heated (56°C) elution buffer (provided). When working with small amounts of tissue material (e.g. less than 10 mg), DNA is eluted in 50–100 µL of preheated elution buffer. The elution step can be repeated with the eluted buffer (from the previous step).

2.1.7 No DNA clean-up is required after DNA extraction.

2.1.8 The extracted DNA should either be used immediately or stored at −20°C until use.

2.2 PCR of the arthropod COI *gene*

2.2.1 PCR sequencing of 709 bp (amplicon size including primers) of the mitochondrial cytochrome c oxidase subunit I (*COI*) gene of arthropods is adapted from Folmer *et al*. (1994).

2.2.2 Primer sequences are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
LCO1490	GGTCAACAAATCAT AAAGATATTGG	X	X
HCO2198	TAAACTTCAGGGTG ACCAAAAAATCA	X	X

2.2.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
LCO1490	10 µM	0.5	0.2 µM
HCO2198	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.4 Thermocycler profile: 3 min 94°C, 5× (30 s 94°C, 30 s 45°C, 1 min 72°C), 35× (30 s 94°C, 1 min 51°C, 1 min 72°C), 10 min 72°C.

2.2.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.2.6 The mitochondrial *COI* is a protein coding region. Translation Table 5 (invertebrate mitochondrial) applies to the mitochondrial *COI* gene.

2.2.7 Primer pair LCO1490/HCO2198 results in a *COI* sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

2.3 Alternative PCR for the arthropod COI *gene – alternative 1.*

2.3.1 PCR sequencing of 709 bp (amplicon size incl. primers) of the mitochondrial cytochrome c oxidase subunit I (*COI*) gene of arthropods adapted from Hebert *et al*. (2004) and Hajibabaei *et al*. (2006).

2.3.2 Primer sequences are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
LepF	ATTCAACCAATCA TAAAGATATTGG	X	X
LepR	TAAACTTCTGGATG TCCAAAAAATCA	X	X

2.3.3 Master mixes are prepared according to the table below (the volumes per reaction can be up scaled to a larger total volume).

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	6	NA
HOT FIREPol Blend Master Mix with 15 mM $MgCl_2$ (Solis Biodyne)*	5×	2	1×
LepF	10 µM	0.5	0.5 µM
LepR	10 µM	0.5	0.5 µM
Subtotal		9.0	
Genomic DNA extract		1.0	
Total		10.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.3.4 Thermocycler profile: 15 min 95°C, 5× (45 s 95°C, 45 s 44°C, 45 s 72°C), 35× (45 s 95°C, 45 s 49°C, 45 s 72°C), 7 min 72°C.

2.3.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.3.6 The mitochondrial *COI* is a protein coding region. Translation Table 5 (invertebrate mitochondrial) applies to the mitochondrial *COI* gene.

2.3.7 Primer pair LepF/LepR result in a *COI* sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

2.4 *Alternative PCR for the arthropod* COI *gene – alternative 2.*

2.4.1 PCR sequencing of 745 bp (amplicon size including primers) of the mitochondrial cytochrome c oxidase subunit I (*COI*) gene of arthropods is adapted from Germain *et al*. (2013).

2.4.2 Primer sequences are described in the table below. The M13-tailed *COI* Hemiptera primer cocktail is prepared by pooling an equal volume of 10 µM of the five primers LCO1490puc-t1, LCO1490Hem1-t1, HCO2198puc-t1, HCO2198Hem1-t1 and HCO2198Hem2-t1.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
LCO1490puc-t1	caggaaacagctatgacc TTTCAACWAATC ATAAAGATATTGG*	X	
LCO1490Hem1-t1	caggaaacagctatgaccTTT CAACTAAYCATAA RGATATYGG	X	
HCO2198puc-t1	tgtaaaacgacggccagtTAA ACTTCWGGRTGWC CAAARAATCA	X	
HCO2198Hem1-t1	tgtaaaacgacggccagtTAA ACYTCDGGATGBCC AAARAATCA	X	
HCO2198Hem2-t1	tgtaaaacgacggccagtTAA ACYTCAGGATGAC CAAAAAAYCA	X	
M13rev-29	caggaaacagctatgacc		X
M13uni-21	tgtaaaacgacggccagt		X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.4.3 Master mixes are prepared according to the table below

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	10	NA
MyFi™ mix (Bioline)*	2 x	12.5	1×
Hemiptera primer cocktail	10 µM total	0.5	0.2 µM total
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.4.4 Thermocycler profile: 3 min 94°C, 5× (30 s 94°C, 30 s 45°C, 1 min 72°C), 35× (30 s 94°C, 1 min 51°C, 1 min 72°C), 10 min 72°C.

2.4.5 Cycle sequencing reactions are performed using the primers targeting the respective M13-tags in separate reactions.

2.4.6 The mitochondrial *COI* is a protein coding region. Translation Table 5 (invertebrate mitochondrial) applies to the mitochondrial *COI* gene.

2.4.7 The M13-tailed primer cocktail result in a *COI* sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

3. Essential procedural information

3.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each series of nucleic acid extraction and amplification of the target organism and target nucleic acid, respectively:

- Negative isolation control (NIC) to monitor contamination during DNA extraction: include an empty tube in the DNA extraction procedure as if it were a real sample.
- Negative amplification control (NAC) to rule out false positives due to contamination during the preparation of the reaction mix: include a tube with no added template, instead add 2 µL of molecular-grade water that was used to prepare the reaction mix.
- Positive amplification control (PAC) to monitor the efficiency of the amplification: amplification of gBlock EPPO_PAC_Arthropods_1 (0.1 ng μL^{-1}) (see Appendix 9) or genomic DNA of a relevant target organism (see Table 1).

3.2 Interpretation of results

Verification of the controls

- NIC and NAC: no band is visualized.
- PAC: a band of the expected size is visualized.

When these conditions are met

- Tests yielding amplicons of the expected size are used for cycle sequencing.
- Tests should be repeated if any contradictory or unclear results are obtained.

4. Performance characteristics available

Performance characteristics for the tests 2.2 *COI* and 2.4 *COI* alternative 2 in this appendix were determined under the EUPHRESCO DNA Barcoding project in an international consortium of 11 participants (see the EPPO validation sheet, https://dc.eppo.int/validation_data/dwvalidation?id=114). Additional data was generated by the Dutch NPPO laboratory. Performance characteristics for the test 2.3 *COI* alternative 1 in this Appendix were generated by AGES/MBDP.

4.1 Analytical sensitivity

Tissue material (typically 10–50 mg) of all life stages of a single specimen is used as input for DNA extraction. Concentration of the DNA extract was determined with Qubit dsDNA Assay (Invitrogen). For all protocols a DNA extract dilution of approximately 4 ng μL^{-1} was sufficient to generate an amplicon that can be sequenced, leading to a high-quality (HQ) consensus sequence (Phred score >40) of at least 99%. For more information, see the relevant EPPO validation sheet.

4.2 Analytical specificity

The locus indicated in Table 1 possesses sufficient interspecies variation to allow for identification to species level. In addition to the species listed in Table 1, species from several genera have successfully been amplified and sequenced by the Dutch NPPO (test 2.2 *COI* and test 2.4 *COI* alternative 2) or by AGES/MBDP (test 2.3 *COI* alternative 1) using the protocols in this appendix. The number of species tested is indicated in parentheses below; for the list of species, see the relevant EPPO validation sheet.

Test 2.2 *COI*:

Acanthocinus (1), *Acleris* (1), *Adoxophyes* (1), *Anastrepha* (1), *Anoplophora* (8), *Apriona* (1), *Argyrogramma* (1), *Atherigona* (1), *Autographa* (1), *Bactrocera* (5), *Bombus* (1), *Cameraria* (1), *Carpomya* (1), *Ceratitis* (3), *Chloridea* (2), *Chromatomyia* (1), *Chrysodeixis* (1), *Chymomyza* (1), *Clepsis* (1), *Clytus* (1), *Conogethes* (1), *Contarinia* (1), *Copitarsia* (2), *Coremagnatha* (1), *Cydalima* (1), *Cydia* (1), *Dasineura* (3), *Deroceras* (1), *Desmiphora* (1), *Deudorix* (1), *Diabrotica* (1), *Diaphania* (2), *Dorata* (1), *Drosophila* (2), *Dryocosmus* (1), *Earias* (1), *Elaphria* (2), *Enarmonia* (1), *Ephestia* (1), *Epiphyas* (1), *Euclea* (1), *Euleia* (1), *Frankliniella* (1), *Grapholita* (1), *Helicoverpa* (2), *Heliothis* (1), *Hesperophanes* (1), *Himacerus* (1), *Hylotrupes* (1), *Hymenia* (1), *Hypena* (1), *Janetiella* (2), *Janus* (1), *Lasioptera* (2), *Liriomyza* (5), *Mamestra* (1), *Maruca* (1), *Mesopolobus* (1), *Monochamus* (7), *Muscina* (1), *Napomyza*

(2), *Neoleucinodes* (1), *Orgyia* (1), *Ornidia* (1), *Ovachlamys* (1), *Ozodes* (1), *Palpita* (1), *Pemphredon* (1), *Placochela* (1), *Planococcus* (1), *Platynota* (2), *Pomacea* (1), *Prays* (1), *Psapharochrus* (1), *Pyroderces* (1), *Rhagoletis* (1), *Rhectocraspeda* (1), *Rhinoncus* (1), *Sesia* (1), *Sinibotys* (1), *Spodoptera* (15), *Sternochetus* (1), *Strymon* (1), *Tetranychus* (1), *Thaumatotibia* (1), *Thecabius* (1), *Thrips* (3), *Torymus* (1), *Trichoferus* (2), *Tuta* (1), *Vittaplusia* (1), *Xylodiplosis* (1), *Xylotrechus* (1), and *Xystrocera* (1).

Test 2.3 *COI* alternative 1:

Aleyrodes (1), *Aromia* (1), *Bactrocera* (3), *Bemisia* (1), *Bradysia* (1), *Ceratitis* (1), *Contarinia* (1), *Cydia* (1), *Delia* (1), *Grapholita* (1), *Helicoverpa* (1), *Hendecasis* (1), *Janus* (1), *Lasioptera* (1), *Maruca* (1), *Megalurothrips* (1), *Megoura* (1), *Noctua* (1), *Opogona* (1), *Phthorimaea* (1), *Rhynchites* (1), *Spodoptera* (3), *Therioaphis* (1), *Thrips* (2), *Trialeurodes* (1).

Test 2.4 *COI* alternative 2:

Anoplophora (4), *Apriona* (1) and *Argyresthia* (1).

It should be recognized that the potential for amplification and sequencing with the generic primers in this appendix is much larger.

4.3 Selectivity

Selectivity does not apply as individual specimens are used.

4.4 Diagnostic sensitivity

Test performance study (TPS) partners in the EUPHRESCO II DNA Barcoding project analysed five DNA samples of the following species: *Vespa crabro* (not regulated), *Bemisia tabaci*, *Liriomyza huidobrensis*, *Spodoptera eridania* and *Anoplophora glabripennis.* The overall diagnostic sensitivity obtained was 98%. All except one sample was correctly identified. One partner used conservative identification for the *Spodoptera eridania* sample (i.e. Lepidoptera sp.: order level identification) which resulted in a diagnostic sensitivity of 91% for this sample. Re-analysis of data produced by this partner showed that species level identification is possible and an overall diagnostic sensitivity of 100% could be obtained. The blind samples of the TPS could accordingly be resolved using the test 2.3 *COI* alternative 1 (this comparative verification was only performed by the AGES/MBDP and resulted in 100% concordance in diagnostic sensitivity).

4.5 Reproducibility

The same DNA samples were analysed by different partners. Therefore, in this situation the reproducibility is identical to diagnostic sensitivity. For the test 2.3 the same DNA samples were tested by two operators.

The outcome of data analysis is dependent on the databases used and relies on a combination of nucleotide similarity, specific clustering in tree views and the ability of end users to recognize sequence data deposited in databases which are likely to be misidentified. The analysis of sequence data using online resources and the interpretation of BLAST and MLSA results heavily depends on the proficiency of operators handling the data. All relevant (online) resources should be used to draw a final conclusion for the data analysis. See Appendix 7 for guidance on data analysis.

Appendix 2 – DNA barcoding Bacteria

The tests below are described as they were carried out to generate the validation data provided in section 4 (unless stated otherwise) in combination with the guidance provided in Appendix 7. Other equipment, kits or reagents may be used provided that a verification (see PM 7/98) is carried out.

1. General information

1.1 This appendix outlines protocols for the identification of selected regulated bacteria using conventional PCR followed by Sanger sequencing analysis. Table 2 shows the regulated organisms that have successfully been tested with the protocols described in this section. It is very likely that other regulated bacteria can successfully be identified using these protocols, but validation data has not been generated to support this.

1.2 The protocol was developed by the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), the Laboratory for Microbiology, Ghent University, Belgium, and Agroscope, Switzerland, as part of the QBOL Project financed by the 7th Framework Programme of the European Union (2009–12). As part of the EUPHRESCO II DNA Barcoding Project (2013–14), the protocols were further optimized by ILVO, Belgium.

1.3 A combination of two to three genes out of four genes are used to identify selected regulated bacteria, i.e. the *16S* ribosomal DNA (*16S* rDNA), *gyrB* (2×), *avrBs2* and *egl*. After *16S* rDNA-based identification of the bacterial genus, the protocol follows the barcoding strategy as presented in the diagnostic testing scheme (see Fig. 2) to determine the (sub)species and pathovars. Table 2 gives an overview of the loci needed for the selected regulated bacteria.

1.4 Primer sequences, amplicon sizes and thermocycler settings are provided in the test-specific sections. HPLC-purified primers should be ordered to avoid nonspecific PCR amplification.

1.5 Reaction mixes are based on the MyFi™ Mix (Bioline) reagents or on the Roche FastStart system. The tests 2.2 *16S* rDNA, 2.3 *gyrB*, 2.5 *gyrB* and 2.6 *avrBs2* were initially validated using BIO-X-ACT™ Short Mix (Bioline), which is no longer available. Verification performed in NVWA showed that MyFi™ Mix and BIO-X-ACT™ Short Mix performed equally (Van de Vossenberg, pers. comm.).

1.6 Molecular-grade water is used to set up reaction mixes; this should be purified (deionized or distilled), sterile (autoclaved or 0.22 -µm filtered) and nuclease-free.
1.7 Amplification is validated for Peltier-type thermocycler with heated lid, e.g. C1000 (Bio-Rad) or Applied Biosystems PCR cycler 9700 (test 2.4 *egl*).

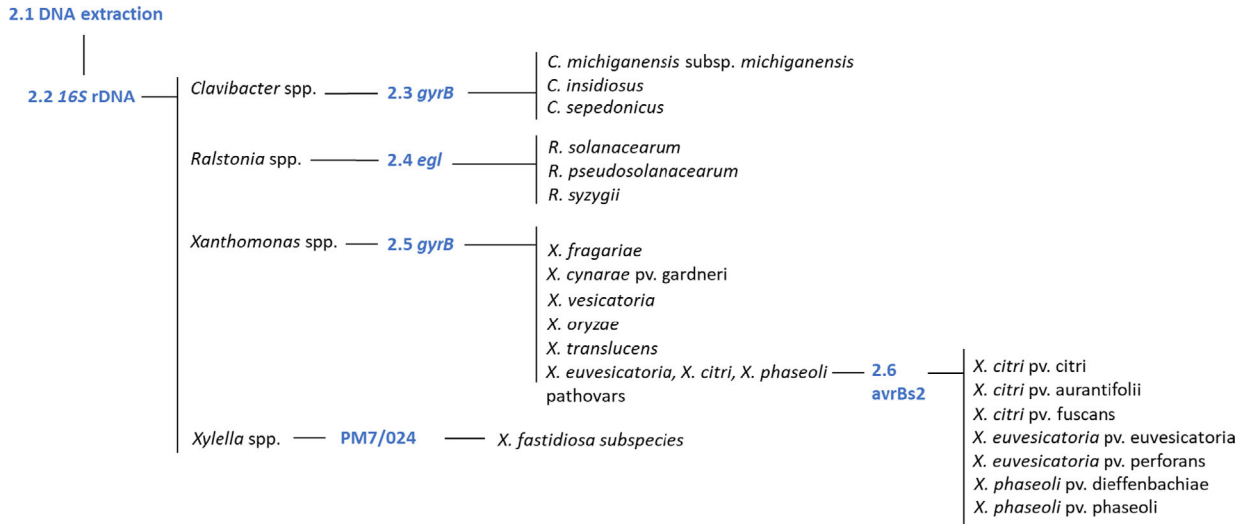


Fig. 2 Diagnostic testing scheme for identification of regulated bacteria using DNA barcodes. The steps shown refer to the sections in this appendix which should be followed to reach reliable identification of the corresponding taxa. When sequence data of multiple loci are generated, the MLSA tools in EPPO-Q-bank need to be used.

Table 2. Regulated bacteria successfully identified with barcoding protocols

	Test*					
Regulated organism	2.2 *16S* rDNA	2.3 *gyrB Clavibacter*	2.4 *egl Ralstonia*	2.5 *gyrB Xanthomonas*	2.6 *avrBs2 Xanthomonas*	Remarks
Clavibacter michiganensis subsp. *michiganensis*	x	x				Gram +
Clavibacter insidiosus	x	x				Gram +
Clavibacter sepedonicus	x	x				Gram +
Ralstonia solanacearum species complex	x		x			Gram −
Xanthomonas phaseoli pv. *dieffenbachiae*	x			x	x	Gram −
Xanthomonas phaseoli pv. *phaseoli*	x			x	x	Gram −
Xanthomonas citri pv. *citri*	x			x	x	Gram −
Xanthomonas euvesicatoria pv. *euvesicatoria*	x			x	x	Gram −
Xanthomonas fragariae	x			x		Gram −
Xanthomonas citri pv. *aurantifolii*	x			x	x	Gram −
Xanthomonas citri pv. *fuscans*	x			x	x	Gram −
Xanthomonas cynarae pv. *gardneri*	x			x		Gram −
Xanthomonas oryzae	x			x		Gram −
Xanthomonas euvesicatoria pv. *perforans*	x			x	x	Gram −
Xanthomonas translucens	x			x		Gram −
Xanthomonas vesicatoria	x			x		Gram −
Xylella spp.	x					Gram − (see PM 7/024 for identification at subspecies level of *X. fastidiosa*)

*Tests marked with 'x' need to be performed to reach reliable identification of the corresponding taxa. When multiple loci are indicated in the table, the MLSA tools in EPPO-Q-bank should be used.

2. Methods

2.1 Nucleic acid extraction and purification

2.1.1 Cell pellets of pure cultures (maximum 2×10^9 cells) are used as starting material for the DNA extraction.

2.1.2 DNA is extracted using the Blood & Tissue kit (Qiagen) using the pre-treatment for Gram-negative or Gram-positive bacteria followed by the Animal Tissues protocol according to the manufacturer's instructions. The pre-treatment for Gram-positive bacteria can also be used for the DNA extraction of Gram-negative bacteria.

2.1.3 DNA is eluted in 100 µL elution buffer (provided in the kit). As the first elution fraction may still contain impurities, elution is performed two times using 50 µL elution buffer and the two fractions are collected in a single microcentrifuge tube.

2.1.4 No DNA clean-up is required after DNA extraction. Note: High DNA concentrations may hamper the PCR test and diluting the DNA extract decreases potential PCR inhibitors in the extract. Diluting the DNA extract to approximately 10 ng μL^{-1} for the purpose of using approximately 20 ng DNA as template in the PCR tests is recommended.

2.1.5 The extracted DNA should either be used immediately or stored until use at −20°C or below.

2.2 Conventional PCR 16S rDNA bacteria

2.2.1 PCR of approximately 1500 bp of the *16S* ribosomal DNA (*16S* rDNA) amplification is adapted from Edwards *et al*. (1989), followed by sequencing of a partial 309–350 bp fragment using the two reverse primers as adapted from Coenye *et al*. (1999).

2.2.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
pA (forward primer)	AGAGTTTGATCCTGGCTCAG	X	
pH (reverse primer)	AAGGAGGTGATCCAGCCGCA	X	
16R339	ACTGCTGCCTCCCGTAGGAG		X
16R519	GTATTACCGCGGCTGCTG		X

Note: For sequencing, the pA primer can also be used in combination with any of the reverse primers listed in the table.

2.2.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.0	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
pA (forward primer)	10 µM	0.75	0.3 µM
pH (reverse primer)	10 µM	0.75	0.3 µM
Subtotal		23.0	
Genomic DNA extract	~10 ng μL^{-1}	2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.4 Thermocycler profile: 1 min 30 s 98°C, 30× (20 s 98°C, 20 s 60°C, 1 min 72°C), 5 min 72°C.

2.2.5 Cycle sequencing reactions of a small fragment from the amplified 1500 bp are performed using primers 16R339 and 16R519 in separate reactions. The obtained dual coverage sequence (309–350 bp) fragment is used for genus identification.

2.2.6 *16S* rDNA is a noncoding but conserved locus that is transcribed in *16S* ribosomal RNA. Translation tables do not apply to *16S* rDNA.

2.3 Conventional PCR gyrB Clavibacter species.

2.3.1 PCR sequencing of 598 bp (amplicon size including primers) of the gyrase subunit B gene (*gyrB*) for *Clavibacter* species is adapted from Richert *et al*. (2005).

2.3.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)*	Primer used for	
		PCR	Sequencing
GyrB 2F (M13- tagged)	caggaaacagctatgacc ACCGTCGAGTTC GACTACGA	X	
GyrB 4R (M13- tagged)	tgtaaaacgacggccagtC CTCGGTGTTGC CSARCTT	X	
M13rev-29	caggaaacagctatgacc		X
M13uni-21	tgtaaaacgacggccagt		X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.3.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.0	NA
MyFi™ Short mix (Bioline)*	2×	12.5	1×
GyrB 2F (M13- tagged)	10 µM	0.75	0.3 µM
GyrB 4R (M13- tagged)	10 µM	0.75	0.3 µM
Subtotal		23.0	
Genomic DNA extract	~10 ng μL^{-1}	2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.3.4 Thermocycler profile: 1 min 30 s 98°C, 30× (10 s 98°C, 10 s 60°C, 30 s 72°C), 5 min 72°C.

2.3.5 Cycle sequencing reactions are performed using the primers targeting the respective M13 tags in separate reactions.

2.3.6 The *gyrB* gene is a protein coding region. Translation Table 11 (The Bacterial, Archaeal and Plant Plastid Code) applies to the bacterial *gyrB* gene.

2.3.7 The M13 tailed primer pair GyrB 2F/GyrB 4R results in a *gyrB* sequence with codon start in reading frame 3 of the primer-trimmed consensus sequence.

2.4. Conventional PCR egl for the *Ralstonia solanacearum* species complex

2.4.1 PCR amplification of 837–846 bp (amplicon size including primers) of the endoglucanase (*egl*) gene for the identification of the *Ralstonia solanacearum* species complex, i.e. *Ralstonia solanacearum*, *Ralstonia pseudosolanacearum* and *Ralstonia syzygii*, is adapted from Wicker *et al.* (2007). The gene is located on the megaplasmid.

2.4.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)*	Primer used for	
		PCR	Sequencing
EndoF (M13- tagged)	caggaaacagctatgaccA TGCATGCCGCTGG TCGCCGC	X	
EndoR (M13- tagged)	tgtaaaacgacggccagtG CGTTGCCCGGCAC GAACACC	X	
M13rev-29	caggaaacagctatgacc		X
M13uni-21	tgtaaaacgacggccagt		X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.4.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	26.8	NA
FastStart PCR buffer (incl. 20 mM $MgCl_2$)*	10×	5.0	1× (2.0 mM $MgCl_2$)
FastStart GC rich solution	5×	10	1×
PCR grade dNTPs	10 mM	1.0	200 µM
FastStart Taq DNA polymerase	5U μL^{-1}	0.2	1 U
Endo-F (M13-tagged)	10 µM	1.0	0.2 µM
Endo-R (M13-tagged)	10 µM	1.0	0.2 µM
Subtotal		45.0	
Genomic DNA extract	~10 ng μL^{-1}	5.0	
Total		50.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.4.4 Thermocycler profile: 5 min 94°C, 30× (30 s 94°C, 30 s 65°C, 1 min 72°C), 5 min 72°C.

2.4.5 Cycle sequencing reactions are performed using the primers targeting the respective M13-tags in separate reactions.

2.4.6 The *egl* gene is a protein coding sequence (CDS). Translation Table 11 (The Bacterial, Archaeal and Plant Plastid Code) applies to the bacterial *egl* gene.

2.4.7 For phylotype I: egl barcode on the positive strand starts at ACCGACACC and ends at CAGTGG, with codon (ACC) starting at position 74 of the primer-trimmed sequence = 666 bp.

For phylotype IIA: egl barcode on the positive strand starts at ACCGACACC and ends at CAGTGG, with codon (ACC) starting at position 80 of the primer-trimmed sequence = 666 bp.

For phylotype IIB: egl barcode on the positive strand starts at ACGGACACC and ends at CAGTGG, with codon (ACG) starting at position 83 of the primer-trimmed sequence = 666 bp.

For phylotype III: egl barcode on the positive strand starts at GCCGACACC or ACCGACACC and ends at CAGTGG with codon (GCC or ACC)

starting at position 80 of the primer-trimmed sequence = 666 bp.

For phylotype IV: egl barcode on the positive strand starts at ACCGACACC and ends at CAGTGG or CAATGG with codon (ACC) starting at position 80 of the primer-trimmed sequence = 666 bp.

2.5 Conventional PCR gyrB Xanthomonas *species*

2.5.1 PCR amplification 765 bp (amplicon size including primers) of the gyrase subunit B gene (g*yrB*) for *Xanthomonas* species. Identification is adapted from Parkinson *et al*. (2007).

2.5.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)*	Primer used for	
		PCR	Sequencing
XgyrPCR2F (M13-tagged)	caggaaacagctatgacc AAGCAGGGCAA GAGCGAGCTGTA	X	
X.gyrrsp1 (M13-tagged)	tgtaaaacgacggccagt CAAGGTGCTGAA GATCTGGTC	X	
M13rev-29	caggaaacagctatgacc		X
M13uni-21	tgtaaaacgacggccagt		X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.5.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.0	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
XgyrPCR2F (M13- tagged)	10 µM	0.75	0.3 µM
X.gyrrsp1 (M13- tagged)	10 µM	0.75	0.3 µM
Subtotal		23.0	
Genomic DNA extract	~10 ng μL^{-1}	2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.5.4 Thermocycler profile: 1 min 30 s 98°C, 30× (10 s 98°C, 10 s 60°C, 30 s 72°C), 5 min 72°C.

2.5.5 Cycle sequencing reactions are performed using the primers targeting the respective M13 tags in separate reactions.

2.5.6 The *gyrB* gene is a protein coding region. Translation Table 11 (The Bacterial, Archaeal and Plant Plastid Code) applies to the bacterial *gyrB* gene.

2.5.7 The M13-tailed primer pair XgyrPCR2F/X.-gyrrsp1 results in a *gyrB* sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

2.6 Conventional PCR avrBs2 Xanthomonas *pathovars*

2.6.1 PCR amplification of approx. 905 bp (amplicon size including primers) of the avirulence protein gene (*avrBs2*) for *Xanthomonas* pathovar identification is adapted from Hajri *et al*. (2009).

2.6.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)*	Primer used for	
		PCR	Sequencing
AvrBs2F (M13-tagged)	caggaaacagctatgacc GGACTAGTCCTG CCGGTGTTGATG CACGA	X	
AvrBs2R (M13-tagged)	tgtaaaacgacggccagt CGCTCGAGCGGTG ATCGGTCAACAG GCTTTC	X	
M13rev-29	caggaaacagctatgacc		X
M13uni-21	tgtaaaacgacggccagt		X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.6.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.0	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
AvrBs2F (M13-tagged)	10 µM	0.75	0.3 µM
AvrBs2R (M13-tagged)	10 µM	0.75	0.3 µM
Subtotal		23.0	
Genomic DNA extract	~10 ng μL^{-1}	2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.6.4 Thermocycler profile: 1 min 30 s 98°C, 30× (10 s 98°C, 10 s 60°C, 30 s 72°C), 5 min 72°C.

2.6.5 Cycle sequencing reactions are performed using the primers targeting the respective M13 tags in separate reactions.

2.6.6 The *avrBs2* gene is a protein coding region. Translation Table 11 (The Bacterial, Archaeal and Plant Plastid Code) applies to the bacterial *avrBs2* gene.

2.6.7 The M13-tailed primer pair AvrBs2F/AvrBs2R results in an *avrBs2* sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

3 Essential procedural information

3.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each series of nucleic acid extraction and amplification of the target organism and target nucleic acid, respectively.

- Negative isolation control (NIC) to monitor contamination during DNA extraction: include an empty tube in the DNA extraction procedure as if it were a real sample.
- Negative amplification control (NAC) to rule out false positives due to contamination during the preparation of the reaction mix: include a tube with no added template, instead add adequate volume of molecular-grade water that was used to prepare the reaction mix.
- Positive amplification control (PAC) to monitor the efficiency of the amplification: amplification of gBlock EPPO_PAC_Bacteria_1 (0.1 ng μL^{-1}) (see Appendix 9) or genomic DNA of a relevant target organism (see Table 2). It should be noted that gBlock EPPO_PAC_Bacteria_1 cannot be used for *2.4 egl* Protocol.

3.2 Interpretation of results

Verification of the controls

- NIC and NAC: no band is visualized.
- PAC: a band of the expected size is visualized.

When these conditions are met

- Tests yielding amplicons of the expected size are used for cycle sequencing.
- Tests should be repeated if any contradictory or unclear results are obtained.

4. Performance characteristics available

Performance characteristics for the tests 2.2 *16S* rDNA, 2.3 gyrB *Clavibacter*, 2.5 gyrB *Xanthomonas* and 2.6 avrBs2 *Xanthomonas* pathovars in this appendix were determined under the EUPHRESCO DNA Barcoding project in an international consortium of 11 participants (see the EPPO validation sheet, https://dc.eppo.int/validation_data/dwvalidation?id=115). Additional data was generated by the Dutch NPPO laboratory. Performance characteristics for the test 2.4 *egl Ralstonia solanacearum* species complex in this appendix were determined by ILVO.

4.1 Analytical sensitivity

Pellets of pure cultures are used for the DNA extraction. Concentration of the DNA extract was determined with Qubit dsDNA Assay (Invitrogen). For protocols 2.2 *16S* rDNA, 2.3 gyrB *Clavibacter*, 2.5 gyrB *Xanthomonas* and 2.6 avrBs2 *Xanthomonas* pathovars, a DNA extract dilution of approximately 1.1 ng μL^{-1} is sufficient to generate an amplicon that can be sequenced, leading to a consensus sequence with a HQ (Phred > 40) of at least 84%. For more information, see the relevant EPPO validation sheet.

4.2 Analytical specificity

The combination of loci indicated in Table 2 possess sufficient interspecies variation to allow for identification to species level and, when relevant, also to the subspecies or pathovar level. Apart from the species listed in Table 2, species from several genera have successfully been amplified and sequenced using the protocols in this appendix by the Dutch NPPO. The number of species tested is indicated in brackets below; for the list of species, see the relevant EPPO validation sheet.

Test 2.2 16S rDNA: *Acidovorax* (4), *Clavibacter* (1), *Curtobacterium* (1), *Dickeya* (7), *Pantoea* (1), *Pseudomonas* (2), *Ralstonia* (1), *Rhodococcus* (1) and *Xanthomonas* (4).

Test 2.3 gyrB *Clavibacter: Clavibacter* (1).

Test 2.5 gyrB *Xanthomonas*: *Xanthomonas* (10).

Test 2.6 avrBs2 *Xanthomonas: Xanthomonas* (7).

It should be recognized that potential for amplification and sequencing with the generic primers in this appendix is much larger.

For the test 2.4 *egl Ralstonia solanacearum* species complex, barcoding was performed on over 200 exotic isolates of *Ralstonia solanacearum* species complex. A set of 25 European and nonEuropean strains from official collections was used as a benchmark.

4.3 Selectivity

Selectivity does not apply as pure cultures are used.

4.4 Diagnostic sensitivity

TPS partners in the EUPHRESCO II DNA Barcoding project analysed five DNA samples of the following species: *Clavibacter michiganensis* subsp. *michiganensis*, *Ralstonia solanacearum sensu lato*, *Xanthomonas axonopodis* pv. *begoniae* (not regulated), *Xanthomonas phaseoli* pv. *dieffenbachiae* and *Xylella fastidiosa.* The overall diagnostic sensitivity obtained was 67% (*C. michiganensis* subsp. *michiganensis* 55 %, *X. a.* pv. *begoniae* 45%, *X. phaseoli* pv. *dieffenbachiae* 45% and *X. fastidiosa* 100%). Conservative identification at higher taxonomical level due to a lack of confidence of the operators to make the identification at

subspecies or pathovar level [i.e. *C. michiganensis* instead of *C. michiganensis* subsp. *michiganensis* (n = 5) and *X. axonopodis* instead of *X. a.* pv. b*egoniae* (n = 2) or *X. ph.* pv. *dieffenbachiae* (n = 3)] and incorrect identifications led to relative low diagnostic sensitivity values for some samples. Reanalysis of the data provided by partners show that identification at the required taxonomical level as listed in Table 2 is possible and an overall diagnostic sensitivity of 96% could be obtained.

The Euphresco validation data were generated before the reclassification of *Ralstonia* so the data are not shown.

4.5 Reproducibility data

The same DNA samples are analysed by different partners. Therefore, in this situation the reproducibility is identical to diagnostic sensitivity.

The outcome of data analysis is dependent on the databases used and relies on a combination of nucleotide similarity, specific clustering in tree views and the ability of end users to recognize sequence data deposited in databases which are likely to be misidentified. The analysis of sequences data using online resources and the interpretation of BLAST and MLSA results heavily depends on the proficiency of operators handling the data. All relevant (online) resources should be used to draw a final conclusion for the data analysis. See Appendix 7 for guidance on data analysis.

Appendix 3 – DNA barcoding Fungi and Oomycetes

The tests below are described as they were carried out to generate the validation data provided in section 4 (unless stated otherwise) in combination with the guidance provided in Appendix 7. Other equipment, kits or reagents may be used provided that verification (see PM 7/98) is carried out.

1. General information

1.1. This appendix describes the protocols for the identification of selected regulated fungi and oomycetes using conventional PCR followed by Sanger sequencing analysis. Table 3 shows the regulated organisms that have successfully been tested with the protocols described in this section. It is very likely that other regulated fungi and oomycetes can successfully be identified using these protocols, but validation data has not been generated to support this.

1.2. Protocols were developed by the CBS-KNAW Fungal Biodiversity Centre, Utrecht, the Netherlands (KNAW-CBS), Plant Research International, Business Unit Biointeractions and Plant Health, Wageningen, the Netherlands (PRI) and the Food and Environment Research Agency, York, United Kingdom (Fera Science Limited), as part of the QBOL Project financed by the 7th Framework Programme of the European Union (2009–12). As part of the EUPHRESCO II DNA Barcoding Project (2013–14), the protocols were further optimized by the Dutch NPPO.

1.3 A combination of two out of six tests is used to identify selected regulated fungi and oomycetes: *ITS*, *TEF1*, *TUB2*, *CALM*, *ACT* and the mitochondrial *COI* gene (see Fig. 3). Table 3 gives an overview of the loci needed for the selected regulated fungi and oomycetes.

1.4 Primer sequences, amplicon sizes and thermocycler settings are provided in the test-specific sections. HPLC-purified primers should be ordered to avoid nonspecific PCR amplification.

1.5 Reaction mixes are based on the MyFi™ Mix (Bioline) reagents (cat. no. BIO-25049). The tests were initially validated using the BIO-X-ACT™ Short Mix (Bioline) which is no longer available. Verification performed in NVWA shows that the MyFi™ Mix and BIO-X-ACT™ Short Mix performed equally (Van de Vossenberg, pers. comm.).

1.6 Molecular-grade water is used to set up reaction mixes; this should be purified (deionized or distilled), sterile (autoclaved or 0.22-µm filtered) and nuclease-free.

1.7 Amplification is performed in a Peltier-type thermocycler with a heated lid, e.g. C1000 (Bio-Rad).

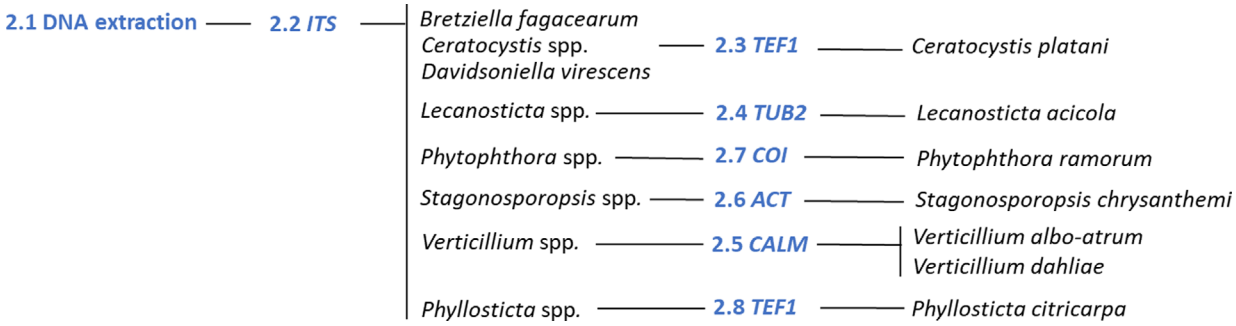


Fig. 3 Diagnostic testing scheme for identification of regulated fungi and oomycetes using DNA barcodes. The steps shown refer to the sections in this appendix which should be followed to reach reliable identification of the corresponding taxa. When sequence data of multiple loci are generated, the MLSA tools in EPPO-Q-bank need to be used.

Table 3. Regulated fungi and oomycetes successfully identified with barcoding protocols

	Tests*						
Regulated organism	2.2 ITS	2.3 TEF1	2.4 TUB2	2.5 CALM	2.6 ACT	2.7 COI	2.8 TEF1
Bretziella fagacearum	x						
Ceratocystis platani	x	x					
Davidsoniella virescens	x						
Lecanosticta acicola	x		x				
Phytophthora ramorum	x					x	
Stagonosporopsis chrysanthemi	x				x		
Verticillium albo-atrum	x			x			
Verticillium dahliae	x			x			
Phyllosticta citricarpa	x						x

*Tests marked with 'x' need to be performed to reach reliable identification of the corresponding taxa. When multiple loci are indicated in the table, the MLSA tools in EPPO-Q-bank should be used.

2. Methods

2.1 Nucleic acid extraction

2.1.1 Mycelium of pure cultures is removed from the agar surface (approximately 2 cm^2) using a sterile scalpel or micro pestle and used as starting material for the DNA extraction.

2.1.2 DNA is extracted using the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions.

2.1.3 Particular care should be given to ensure the sample is adequately homogenized in the step. Grind fungal tissue manually using micro pestles or automatically (e.g. using Retsch Mixer Mill MM301).

2.1.4 DNA is eluted twice in 50 µL elution buffer (provided in the extraction kit).

2.1.5 DNA extracts should be used immediately or stored at −20°C until use.

2.2 Conventional PCR ITS fungi and oomycetes

2.2.1 PCR sequencing of approx. 550–1700 bp (amplicon size including primers) of the nuclear ribosomal internal transcribed spacer (ITS) locus is adapted from White *et al*. (1990).

2.2.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
ITS5	GGAAGTAAAAGTCGTAACAAGG	X	X
ITS4	TCCTCCGCTTATTGATATGC	X	X

2.2.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
ITS5	10 µM	0.5	0.2 µM
ITS4	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.4 Thermocycler profile: 5 min 95°C, 40× (30 s 94°C, 30 s 52°C, 1 min 40 s 72°C), 10 min 72°C.

2.2.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.2.6 Sequencing of the ITS PCR product (covering the partial 18S rDNA, ITS1, 5.8S rDNA, ITS2, and the partial 28S rDNA) results in a nonprotein coding sequence, and translation tables do not apply.

2.3 Conventional PCR TEF1 ceratocystis

2.3.1 PCR Sequencing of approx. 680 bp (amplicon size including primers) of the translation elongation factor 1 alpha (*TEF1*) gene is adapted from Jones *et al*. (2011) and Oliveira *et al*. (2015).

2.3.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
EFCF1	AGTGCGGTGGTATCGACAAG	X	X
EFCF2	TGCTCACGGGTCTGGCCAT	X	X

2.3.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
EFCF1	10 µM	0.5	0.2 µM
EFCF2	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.3.4 Thermocycler profile: 5 min 95°C, 40× (30 s 94°C, 30 s 52°C, 30 s 72°C), 10 min 72°C.

2.3.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.3.6 The nuclear *TEF1* is a protein coding region. Translation Table 1 (the standard code) applies to the nuclear *TEF1* gene.

2.3.7 Primer pair EFCF1/EFCF2 results in a *TEF1* sequence containing two introns and starting at one of them in the primer-trimmed consensus sequence.

2.4 Conventional PCR TUB2 fungi

2.4.1 PCR sequencing of approx. 450 bp (amplicon size including primers) of the nuclear beta-tubulin (*TUB2*) gene is adapted from Groenewald *et al*. (2013).

2.4.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
TUB2Fd	GTBCACCTYCARA CCGGYCARTG	X	X
TUB4Rd	CCRGAYTGRCCR AARACRAAGTTGTC	X	X

2.4.3 Mastermixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
TUB2Fd	10 µM	0.5	0.2 µM
TUB4Rd	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.4.4 Thermocycler profile: 5 min 95°C, 40× (30 s 94°C, 30 s 52°C, 30 s 72°C), 10 min 72°C.

2.4.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.4.6 The nuclear *TUB2* is a protein coding region. Translation Table 1 (the standard code) applies to the nuclear *TUB2* gene.

2.4.7 Primer pair TUB2Fd/TUB4Rd results in a *TUB2* sequence containing three introns and starting at one of them in the primer-trimmed consensus sequence.

2.5 Conventional PCR CALM fungi

2.5.1 PCR sequencing of approx. 520 bp (amplicon size including primers) of the nuclear calmodulin (*CALM*) gene is adapted from Carbone & Kohn (1999).

2.5.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
CAL-228F	GAGTTCAAGGAGGCCTTCTCCC	X	X
CAL-737R	CATCTTTCTGGCCATCATGG	X	X

2.5.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
CAL-228F	10 µM	0.5	0.2 µM
CAL-737R	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.5.4 Thermocycler profile: 5 min 95°C, 40× (30 s 94°C, 30 s 50°C, 30 s 72°C), 10 min 72°C.

2.5.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.5.6 The nuclear *CALM* is a protein coding region. Translation Table 1 (the standard code) applies to the nuclear *CALM* gene.

2.5.7 Primer pair CAL-228F/CAL-737R results in a *CALM* sequence starting with an intron of the primer-trimmed consensus sequence.

2.6. Conventional PCR ACT fungi

2.6.1 PCR sequencing of approx. 290 bp (amplicon size including primers) of the nuclear actin (*ACT*) gene is adapted from Carbone & Kohn (1999).

2.6.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
ACT-512F	ATGTGCAAGGCCGGTTTCGC	X	X
ACT-783R	TACGAGTCCTTCTGGCCCAT	X	X

2.6.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
ACT-512F	10 µM	0.5	0.2 µM
ACT-783R	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.6.4 Thermocycler profile: 5 min 95°C, 40× (30 s 94°C, 30 s 52°C, 30 s 72°C), 10 min 72°C

2.6.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.6.6 The nuclear *ACT* is a protein coding region. Translation Table 1 (the standard code) applies to the nuclear *ACT* gene.

2.6.7 Primer pair ACT-512F/ACT-783R results in an *ACT* sequence with codon start in reading frame 3 of the primer-trimmed consensus sequence and containing two introns.

2.7 Conventional PCR COI *oomycetes*

2.7.1 PCR sequencing of 727 bp (amplicon size including primers) of the mitochondrial cytochrome c oxidase I (*COI*) gene is adapted from Robideau *et al*. (2011).

2.7.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
OomCoxI-Levup	TCAWCWMGAT GGCTTTTTTCAAC	X	X
OomCoxI-Levlo	CYTCHGGRTGW CCRAAAAACCAAA	X	X

2.7.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
OomCoxI-Levup	10 µM	0.5	0.2 µM
OomCoxI-Levlo	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.7.4 Thermocycler profile: 5 min 95°C, 40× (30 s 94°C, 30 s 52°C, 45 s 72°C), 10 min 72°C.

2.7.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.7.6 The mitochondrial *COI* is a protein coding region. Translation Table 5 (Invertebrate Mitochondrial Code) applies to the mitochondrial *COI* gene.

2.7.7 Primer pair OomCoxI-Levup/OomCoxI-Levlo results in a *COI* sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

2.8 Conventional PCR TEF1 Phyllosticta

2.8.1 PCR sequencing of 434 bp (amplicon size including primers) of the translation elongation factor 1-α (*TEF1*) gene, adapted from Carbone & Kohn (1999) and O'Donnell *et al*. (1998).

2.8.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
EF1-728F	CATCGAGAAGTT CGAGAAGG	X	X
EF2	GGARGTACCAGT SATCATGTT	X	X

2.8.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
Primer EF1-728F	10 µM	0.5	0.2 µM
Primer EF2	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.8.4 Thermocycler profile: 5 min 94°C, 40× (30 s 94°C, 30 s 52°C, 30 s 72°C), 5 min 72°C.

2.8.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.8.6 The nuclear *TEF1* is a protein coding region. Translation Table 1 (Standard Code) applies to the nuclear *TEF* gene.

2.8.7 Primer pair EF1-728F/ EF2 results in an *TEF1* sequence containing one intron starting in the primer-trimmed consensus sequence.

3. Essential procedural information

3.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each series of nucleic acid extraction and amplification of the target organism and target nucleic acid, respectively.

- Negative isolation control (NIC) to monitor contamination during nucleic acid extraction: DNA extraction of an Eppendorf tube containing 25 µL of molecular-grade water.

- Negative amplification control (NAC) to rule out false positives due to contamination during the preparation of the reaction mix: amplification of molecular-grade water that was used to prepare the reaction mix.
- Positive amplification control (PAC) to monitor the efficiency of the amplification: amplification of gBlock EPPO_PAC_Fungi_1 (0.1 ng μL^{-1}) (see Appendix 9) or genomic DNA of a relevant target organism (see Table 3). It should be noted that gBlock EPPO_PAC_Fungi_1 cannot be used for 2.8 *TEF1 Phyllosticta* Protocol.

3.2. Interpretation of results

Verification of the controls
- NIC: no band is visualized.
- PAC: a band of the expected size is visualized.

When these conditions are met
- Tests yielding amplicons of the expected size are used for cycle sequencing.
- Tests should be repeated if any contradictory or unclear results are obtained.

4. Performance characteristics available

Performance characteristics for the tests in this appendix (except 2.8 *TEF1 Phyllosticta*) were determined under the Euphresco DNA Barcoding project in an international consortium of nine participants (see EPPO validation sheet, https://dc.eppo.int/validation_data/dwvalidation?id=116). Additional data was generated by the Dutch NPPO laboratory (including on 2.8 *TEF1 Phyllosticta*).

4.1 Analytical sensitivity

Pure cultures are used for the DNA extraction. Concentration of the DNA extract was determined with Qubit dsDNA Assay (Invitrogen). For all protocols a DNA extract dilution of approximately 0.05 ng μL^{-1} was sufficient to generate an amplicon that can be sequenced, leading to a consensus sequence with a HQ (Phred > 40) of at least 83%. For more information, see the relevant EPPO validation sheet.

4.2 Analytical specificity

The locus or combination of loci indicated in Table 3 possess sufficient interspecies variation to allow for identification to species level. Apart from the species listed in Table 3, species from several genera have successfully been amplified and sequenced by the Dutch NPPO laboratory using the protocols in this appendix (the number of species tested is indicated in brackets below; for the list of species, see the relevant EPPO validation sheet).

Test 2.2 *ITS*: *Atropellis* (1), *Boeremia* (1), *Ceratocystis* (1), *Chalara* (1), *Ciborinia* (1), *Colletotrichum* (1), *Diaporthe* (4), *Diplocarpon* (1), *Elsinoe* (3), *Epicoccum* (1), *Fusarium* (1), *Geosmithia* (1), *Gremmeniella* (1), *Heterobasidion* (1), *Melampsora* (2), *Ophiognomonia* (1), *Penicillium* (1), *Peyronellaea* (1), *Phialophora* (1), *Phoma* (2), *Phomopsis* (1), *Phytophthora* (8), *Phytopythium* (1), *Pseudocercospora* (1), *Pyrenochaeta* (1), *Stagonosporopsis* (1) and *Venturia* (1).

Test 2.4 *TUB2*: *Ciborinia* (1), *Colletotrichum* (1), *Fusarium* (1) and *Penicillium* (1).

Test 2.5 *CALM*: *Penicillium* (1).

Test 2.6 *ACT*: *Colletotrichum* (1), *Entoleuca* (1), *Epicoccum* (1), *Phoma* (2) and *Stagonosporopsis* (1).

Test 2.8 *TEF1: Phyllosticta* (2).

It should be recognized that the potential for amplification and sequencing with the generic primers in this appendix is much larger.

4.3 Selectivity

Selectivity does not apply as pure cultures are used.

4.4 Diagnostic sensitivity

TPS partners in the EUPHRESCO II DNA Barcoding project analysed five DNA samples of the following species: *Ceratocystis fimbriata* f. sp. *platani*, *Lecanosticta acicola*, *Phytophthora ramorum*, *Stagonosporopsis chrysanthemi* and *Verticillium dahliae*. The overall diagnostic sensitivity obtained was 96% (*C. fimbriata* f. sp. *platani* 89%, *L. acicola* 100%, *P. ramorum* 100%, *S. chrysanthemi* 89% and *V. dahliae* 100%). One of the partners was not able to correctly identify the sample *Stagonosporopsis chrysanthemi* as no amplicon was obtained for the ACT locus, which is necessary for reliable species identification. Reanalysis of the data provided by partners show that identification at the required taxonomical level as listed in Table 3 is possible and an overall diagnostic sensitivity of 98% could be obtained.

4.5 Reproducibility

The same DNA samples are analysed by different partners. Therefore, in this situation the reproducibility is identical to diagnostic sensitivity.

The outcome of data analysis is dependent on the databases used and relies on a combination of nucleotide similarity, specific clustering in tree views and the ability of end users to recognize sequence data deposited in databases, which are likely to be misidentified. The analysis of sequences data using online resources and the interpretation of BLAST and MLSA results heavily depends on the proficiency of operators handling the data. All relevant (online) resources should be used to draw a final conclusion for the data analysis. See Appendix 7 for guidance on data analysis.

Appendix 4 – DNA barcoding Invasive Plant species

The tests below are described as they were carried out to generate the validation data provided in section 4 (unless stated otherwise) in combination with the guidance provided in Appendix 7. Other equipment, kits or reagents may be used provided that verification (see PM 7/98) is carried out.

1. General information

1.1 This appendix describes protocols for the identification of selected invasive plant species using conventional PCR followed by Sanger sequencing analysis. Table 4 shows the selected invasive plant species that have successfully been tested with the protocols described in this section. It is very likely that other invasive plant species can successfully be identified using these protocols, but validation data has not been generated to support this.

1.2 Protocols were developed by the Dutch NPPO.

1.3 Two tests in parallel are used to identify selected invasive plant species: targeting the chloroplast *trnH-psbA* intergenic spacer and the *rbcL* gene. *rbcL*, one of the standardized DNA barcodes for plants, does not give sufficient resolution for species demarcation for the selected invasive plant species, therefore *trnH-psbA* is added as an additional barcode region (see Fig. 4). Table 4 gives an overview of the selected invasive plant species.

1.4 Primer sequences, amplicon sizes and thermocycler settings are provided in the test-specific sections. HPLC-purified primers should be ordered to avoid nonspecific PCR amplification.

1.5 Reaction mixes are based on the MyFi™ Mix (Bioline) reagents (cat.no. BIO-25049). The tests were initially validated using BIO-X-ACT™ Short Mix (Bioline), which is no longer available. Verification performed in NVWA shows that MyFi™ Mix and BIO-X-ACT™ Short Mix performed equally (Van de Vossenberg, pers. comm.).

1.6 Molecular-grade water is used to set up reaction mixes; this should be purified (deionized or distilled), sterile (autoclaved or 0.22 µm filtered) and nuclease-free.

1.7 Amplification is performed in a Peltier-type thermocycler with heated lid, e.g. C1000 (Bio-Rad).

Table 4. Regulated invasive plant species successfully identified with barcoding protocols.

	Tests*	
Regulated organism	2.2 rbcL	2.3 trnH-psbA
Ludwigia peploides	x	x
Ludwigia grandiflora	x	x
Hydrocotyle ranunculoides	x	x
Myriophyllum aquaticum	x	x
Myriophyllum heterophyllum	x	x

*Tests marked with 'x' need to be performed to reach reliable identification of the corresponding taxa. When multiple loci are indicated in the table, the MLSA tools in EPPO-Q-bank should be used.

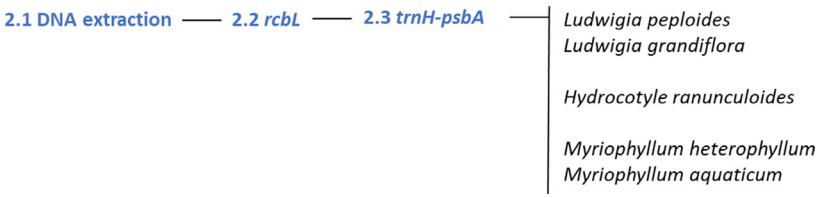


Fig. 4 Diagnostic testing scheme for identification of regulated invasive plant species using DNA barcodes. The steps shown refer to the sections in this appendix which should be followed to reach reliable identification of the corresponding taxa. When sequence data of multiple loci are generated, the MLSA tools in EPPO-Q-bank need to be used.

2. Methods

2.1 Nucleic acid extraction

2.1.1 About 1 g fresh or frozen (green) plant tissue from a single plant is ground in 5 ml GH + grinding buffer (6 M guanidine hydrochloride, 0.2 M sodium acetate pH 5.2, 25 mM EDTA and 2.5% PVP-10), in a plastic grinding bag using Homex 6 (Bioreba AG) and 75 µL of this extract is used as starting material for the DNA extraction.

2.1.2 DNA is extracted using the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions.

2.1.3 DNA is eluted twice in 50 µL of elution buffer (provided in the isolation kit).

2.1.4 DNA extracts should be used immediately or stored at −20°C until use.

2.2 Conventional PCR rbcL invasive plants

2.2.1 PCR sequencing of 599 bp (amplicon size including primers) of the chloroplast large subunit ribulose-1,5-bisphosphate caboxylase-oxygenase (rbcL) gene is adapted from Kress & Erickson, (2007) and Kress *et al*. (2009).

2.2.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
rbcL-a_f	ATGTCACCACA AACAGAGACTAAAGC	X	X
rbcLa SI_Rev	GTAAAATCAAGTCCACCRCG	X	X

2.2.3 Mastermixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
rbcL-a_f	10 µM	0.5	0.2 µM
rbcLa SI_Rev	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.4 Thermocycler profile: 5 min 95°C, 5× (30 s 94°C, 30 s 45°C, 30 s 72°C), 35× (30 s 94°C, 30 s 50°C, 30 s 72°C), 10 min 72°C.

2.2.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.2.6 The chloroplast *rbcL* is a protein coding region and approximately 1430 bp in length. Translation Table 11 (The Bacterial, Archaeal and Plant Plastid Code) applies to the chloroplast *rbcL* gene.

2.2.7 Primer pair rbcL-a_f/rbcLa SI_Rev results in a sequence with codon start in reading frame 2 of the primer-trimmed consensus sequence.

2.3 Conventional PCR trnH-psbA invasive plants

2.3.1 PCR sequencing of 300–900 bp (amplicon size including primers) of the chloroplast intergenic spacer between the histidine transfer tRNA (*trnH*) and the D1 protein of photosystem II (*psbA*) is adapted from Sang *et al*. (1997) and Tate (2002).

2.3.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
trnH2	CGCGCATGGTGGATTCACAATCC	X	X
psbAF	GTTATGCATGAACGTAATGCTC	X	X

2.3.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ Short mix (Bioline)*	2×	12.5	1×
trnH2	10 µM	0.5	0.2 µM
psbAF	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proof reading activity.

2.3.4 Thermocycler profile: 5 min 95°C, 5× (30 s 94°C, 30 s 45°C, 50 s 72°C), 35× (30 s 94°C, 30 s 50°C, 50 s 72°C), 10 min 72°C.

2.3.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.3.6 The chloroplast *trnH-psbA* intergenic spacer is a noncoding region. Translation tables do not apply to *trnH-psbA*.

3. Essential procedural information

3.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each series of nucleic acid extraction and amplification of the target organism and target nucleic acid, respectively.

- Negative isolation control (NIC) to monitor contamination during nucleic acid extraction: DNA extraction of an Eppendorf tube containing 25 µL of molecular-grade water.
- Negative amplification control (NAC) to rule out false positives due to contamination during the preparation of the reaction mix: amplification of molecular-grade water that was used to prepare the reaction mix.
- Positive amplification control (PAC) to monitor the efficiency of the amplification: amplification of gBlock EPPO_PAC_Invasive_Plants_1 (0.1 ng μL^{-1}) (see Appendix 9) or genomic DNA of a relevant target organism (see Table 4).

3.2 Interpretation of results

Verification of the controls

- NIC and NAC: no band is visualized.
- PAC: a band of the expected size is visualized.

When these conditions are met

- Tests yielding amplicons of the expected size are used for cycle sequencing.
- Tests should be repeated if any contradictory or unclear results are obtained.

4. Performance characteristics available

Performance characteristics for the tests in this appendix were determined under the EUPHRESCO DNA Barcoding project in an international consortium of eight participants (see the EPPO validation sheet, https://dc.eppo.int/validation_data/dwvalidation?id=117). Additional data were generated by the Dutch NPPO laboratory.

4.1 Analytical sensitivity

Plant tissue is used for the DNA extraction. Concentration of the DNA extract was determined with Qubit dsDNA Assay (Invitrogen). For all protocols a DNA extract dilution of approximately 5 ng μL^{-1} was sufficient to generate an amplicon that can be sequenced, leading to a consensus sequence with a HQ (Phred > 40) of at least 98%. For more information, see the relevant EPPO validation sheet.

4.2 Analytical specificity

The combination of loci indicated in Table 4 possesses sufficient interspecies variation to allow for identification to species level. Apart from the species listed in Table 4, species from several genera have successfully been amplified and sequenced by the Dutch NPPO using the protocols in this appendix (the number of species tested is indicated in brackets below; for the list of species, see the relevant EPPO validation sheet).

Test 2.2 *rbcL*: *Carex* (1), *Centella* (1), *Cyperus* (3), *Hydrocotyle* (6), *Impatiens* (3), *Kyllinga* (1), *Lagarosiphon* (1), *Ludwigia* (2), *Myriophyllum* (16), *Oxalis* (1), *Rotala* (1) and *Wolffia* (4).

Test 2.*3 trnH-psbA*: *Carex* (1), *Centella* (2), *Cyperus* (3), *Hydrocotyle* (6), *Impatiens* (3), *Kyllinga* (1), *Lagarosiphon* (1), *Ludwigia* (2), *Myriophyllum* (17), *Oxalis* (1), *Rotala* (1) and *Wolffia* (4).

It should be recognized that the potential for amplification and sequencing with the generic primers in this appendix is much larger.

4.3 Selectivity

Selectivity does not apply as single specimens are used.

4.4 Diagnostic sensitivity

TPS partners in the EUPHRESCO II DNA Barcoding project analysed five DNA samples of the following species: *Ludwigia peploides*, *Ludwigia grandiflora*, *Hydrocotyle ranunculoides*, *Hydrocotyle vulgaris* and *Myriophyllum heterophyllum*. The overall diagnostic sensitivity obtained was 68% (*L. peploides* 50%, *L. grandiflora* 63%, *H. ranunculoides* 75%, *H. vulgaris* 63% and *M. heterophyllum* 88%). Conservative identification at a higher taxonomical level (genus instead of species level) led to relative low diagnostic sensitivity values for some samples. Reanalysis of the data provided by partners show that identification at the required taxonomical level as listed in Table 4 is possible and an overall diagnostic sensitivity of 100% could be obtained.

4.5 Reproducibility

The same DNA samples are analysed by different partners. Therefore, in this situation the reproducibility is identical to diagnostic sensitivity.

The outcome of data analysis is dependent on the databases used and relies on a combination of nucleotide similarity, specific clustering in tree views and the ability of end users to recognize sequence data deposited in databases which are likely to be misidentified. The analysis of sequences data using online resources and the interpretation of BLAST and MLSA results heavily depends on the proficiency of operators handling the data. All relevant (online) resources should be used to draw a final conclusion for the data analysis. See Appendix 7 for guidance on data analysis.

Appendix 5 – DNA barcoding Nematodes

The tests below are described as they were carried out to generate the validation data provided in section 4 in combination with the guidance provided in Appendix 7. Other equipment, kits or reagents may be used provided that verification (see PM 7/98) is carried out.

1. General information

1.1 This appendix describes protocols for the identification of selected regulated nematodes using conventional PCR followed by Sanger Sequencing analysis. Table 5 shows the selected regulated organisms that have successfully been tested with the protocols described in this section. Other (regulated) nematode species can successfully be identified using these protocols, but validation data has not been generated to support this.

1.2 The protocols were developed by Agroscope, Switzerland, and the Laboratory of Nematology, Wageningen University, the Netherlands, as part of the QBOL Project financed by the 7th Framework Programme of the European Union (2009–12). As part of the EUPHRESCO II DNA Barcoding Project (2013–14), the protocols were further optimized by the Dutch NPPO.

1.3 A combination of three tests is used to identify selected regulated nematodes: the 18S rDNA (small subunit, SSU), the 28S rDNA (large subunit, LSU) and the mitochondrial *COI* gene (see Fig. 5). Table 5 gives an overview of the loci needed for the selected regulated nematodes.

1.4 Primer sequences, amplicon sizes and thermocycler settings are provided in the test-specific sections. HPLC-purified primers should be ordered to avoid nonspecific PCR amplification.

1.5 Reaction mixes are based on the Phusion® High-Fidelity (New England Biolabs) reagents (cat. no. M0530).

1.6 Molecular-grade water (MGW) is used to set up reaction mixes; this should be purified (deionized or distilled), sterile (autoclaved or 0.22 µm filtered) and nuclease-free.

1.7 Amplification is performed in a Peltier-type thermocycler with heated lid, e.g. C1000 (Bio-Rad).

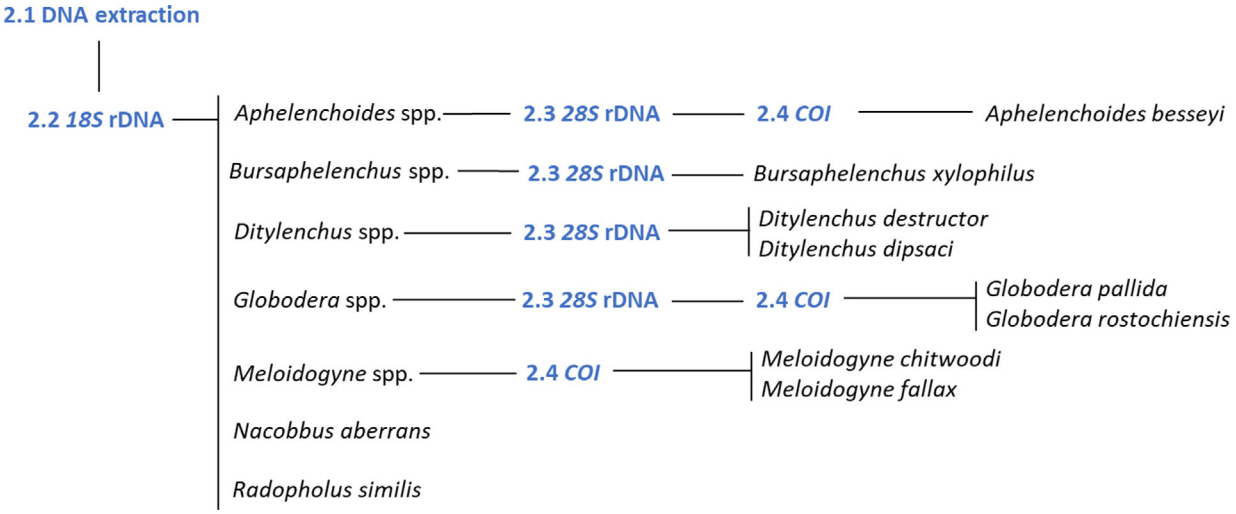


Fig. 5 Diagnostic testing scheme for identification of selected regulated nematodes using DNA barcodes. The steps shown refer to the sections in this appendix which should be followed to reach reliable identification of the corresponding taxa. When sequence data of multiple loci are generated, the MLSA tools in EPPO-Q-bank need to be used.

Table 5. Regulated nematodes successfully identified with barcoding protocols

	Tests*		
Regulated organism	2.2 18S rDNA	2.3 28S rDNA	2.4 *COI*
Aphelenchoides besseyi	x	x	x
Bursaphelenchus xylophilus	x	x	
Ditylenchus destructor	x	x	
Ditylenchus dipsaci	x	x	
Globodera pallida	x	x	x
Globodera rostochiensis	x	x	x
Meloidogyne chitwoodi	x		x
Meloidogyne fallax	x		x
Nacobbus aberrans	x		
Radopholus similis	x		

*Tests marked with 'x' need to be performed to reach reliable identification of the corresponding taxa. When multiple loci are indicated in the table, the MLSA tools in EPPO-Q-bank should be used.

2. Methods

2.1 Nucleic acid extraction

2.1.1 Single nematodes or cysts in 25 µL of molecular-grade water are used as input for DNA extraction.

2.1.2 DNA is extracted using the Single Worm Lysis kit (ClearDetections) following the manufacturer's instructions.

2.1.3 Lysates should be used immediately or stored at −20°C until use.

2.2 Conventional PCR 18S rDNA (SSU) – nematodes

2.2.1 PCR sequencing of approximately 1730 bp of the small subunit 18S ribosomal DNA (18S rDNA (SSU)) is adapted from Holterman *et al.* (2006) using two separate reactions: 988F/1912R (amplicon size including primers approximately 980 bp) and 1813F/2646R (amplicon size including primers approximately 880 bp).

2.2.2 Primer sequences and their application are described in the table below.

Reaction	Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
			PCR	Sequencing
1	988F	CTCAAAGATTA AGCCATGC	X	X
	1912R	TTTACGGTCAG AACTAGGG	X	X
2	1813F	CTGCGTGAGAG GTGAAAT	X	X
	2646R	GCTACCTTGTTA CGACTTTT	X	X

2.2.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL) Reaction 1	Volume per reaction (µL) Reaction 2	Final concentration
Molecular-grade water	NA	16.05	16.05	NA
Phusion HF Buffer (NEB)*	5×	5.0	5.0	1×
dNTPs (NEB)	10 mM	0.5	0.5	200 µM
988F	10 µM	0.6	–	0.24 µM
1912R	10 µM	0.6	–	0.24 µM
1813F	10 µM	–	0.6	0.24 µM
2646R	10 µM	–	0.6	0.24 µM
Phusion DNA polymerase (NEB)	2 Units μL^{-1}	0.25	0.25	0.5 Unit
Subtotal		23.0	23.0	
Genomic DNA extract		2.0	2.0	
Total		25.0	25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.4 Thermocycler profile: 1 min at 98°C, 5× (10 s at 98°C, 20 s at 45°C, 60 s at 72°C), 35× (10 s at 98°C, 20 s at 54°C, 60 s at 72°C), 10 min at 72°C.

2.2.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.2.6 18S rDNA (SSU) is a noncoding but conserved locus that is transcribed in 18S ribosomal RNA. Translation tables do not apply to *18S rDNA* (SSU).

2.3 *Conventional PCR 28S rDNA (LSU) – nematodes*

2.3.1 PCR sequencing of approximately 1000 bp (amplicon size including primers) of the large subunit *28S* ribosomal DNA (*28S* rDNA (LSU)) is adapted from Holterman *et al*. (2008).

2.3.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
28-81for	TTAAGCATATCA TTTAGCGGAGGAA	X	X
28-1006rev	GTTCGATTAGTCTTTCGCCCCT	X	X

2.3.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	16.05	NA
Phusion HF Buffer (NEB)*	5×	5.0	1×
dNTPs (NEB)	10 mM	0.5	200 µM
28-81for	10 µM	0.6	0.24 µM
28-1006rev	10 µM	0.6	0.24 µM
Phusion DNA polymerase (NEB)	2 Units μL^{-1}	0.25	0.5 Unit
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.3.4 Thermocycler profile: 1 min 98°C, 5× (10 s 98°C, 20 s 45°C, 30 s 72°C), 35× (10 s 98°C, 20s 54°C, 30 s 72°C), 10 min 72°C.

2.3.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.3.6 *28S* rDNA (LSU) is a noncoding but conserved locus that is transcribed in *28S* ribosomal RNA. Translation tables do not apply to *28S* rDNA (LSU).

2.4 *Conventional PCR* COI – *nematodes*

2.4.1 PCR sequencing of 447 bp (amplicon size including primers) of the mitochondrial cytochrome c oxidase subunit I (*COI*) gene is adapted from Hu *et al*. (2002).

2.4.2 Primer sequences and their application are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
JB3	TTTTTTGGGCAT CCTGAGGTTTAT	X	X
JB5	AGCACCTAAACTT AAAACATAATGAAAATG	X	X

2.4.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	16.05	NA
Phusion HF Buffer (NEB)*	5×	5.0	1×
dNTPs (NEB)	10 mM	0.5	200 µM
JB3	10 µM	0.6	0.24 µM
JB5	10 µM	0.6	0.24 µM
Phusion DNA polymerase (NEB)	2 Units μL^{-1}	0.25	0.5 Unit
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.4.4 Thermocycler profile: 1 min 98°C, 40× (10 s 98°C, 20 s 41°C, 30 s 72°C), 10 min 72°C.

2.4.5 Cycle sequencing reactions are performed using the obtained PCR products with primers used for amplification in separate reactions.

2.4.6 The mitochondrial *COI* is a protein coding region. Translation Table 5 (Invertebrate Mitochondrial Code) applies to the mitochondrial *COI* gene.

2.4.7 Primer pair JB3/JB5 results in a *COI* sequence with codon start in reading frame 1 of the primer-trimmed consensus sequence.

3. Essential procedural information

3.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each series of nucleic acid extraction and amplification of the target organism and target nucleic acid, respectively.

- Negative isolation control (NIC) to monitor contamination during nucleic acid extraction: DNA extraction of an Eppendorf tube containing 25 µL of molecular-grade water.
- Negative amplification control (NAC) to rule out false positives due to contamination during the preparation of the reaction mix: amplification of molecular-grade water that was used to prepare the reaction mix.
- Positive amplification control (PAC) to monitor the efficiency of the amplification: amplification of gBlock EPPO_PAC_Nematodes_1 (0.1 ng μL^{-1}; see Appendix 9) or genomic DNA of a relevant target organism (see Table 5).

3.2 Interpretation of results

Verification of the controls

- NIC and NAC: no band is visualized.
- PAC: a band of the expected size is visualized.

When these conditions are met

- Tests yielding amplicons of the expected size are used for cycle sequencing.
- Tests should be repeated if any contradictory or unclear results are obtained.

4. Performance characteristics available

Performance characteristics for the tests in this appendix were determined under the EUPHRESCO DNA Barcoding project in an international consortium of nine participants (see EPPO validation sheet, https://dc.eppo.int/validation_data/dwvalidation?id=119). Additional data was generated by the Dutch NPPO laboratory.

4.1 Analytical sensitivity

For all protocols DNA purified from a single nematode is sufficient to generate an amplicon that can be sequenced leading to a consensus sequence with a HQ (Phred score > 40) of at least 86%.

4.2 Analytical specificity

The locus or combination of loci indicated in Table 5 possess sufficient interspecies variation to allow for species level identification. Apart from the species listed in Table 5, species from several genera (the number of species is indicated in brackets below) have successfully been amplified and sequenced by the Dutch NPPO using the protocols in this appendix (the number of species tested is indicated in brackets below; for the list of species, see the relevant EPPO validation sheet).

Test 2.2 18S rDNA: *Aphelenchoides* (5), *Bursaphelenchus* (3), *Cactodera* (1), *Ditylenchus* (2), *Globodera* (3), *Heterodera* (4), *Heterorhabditis* (1), *Longidorus* (1), *Meloidogyne* (7), *Nacobbus* (1), *Paratrichodorus* (3), *Pratylenchus* (6), *Radopholus* (1), *Steinernema* (2), *Subanguina* (1), *Trichodorus* (3) and *Xiphinema* (1).

Test 2.3 28S rDNA: *Aphelenchoides* (5), *Bursaphelenchus* (2), *Cactodera* (1), *Ditylenchus* (2), *Globodera* (2), *Heterodera* (4), *Heterorhabditis* (1), *Longidorus* (1), *Meloidogyne* (6), *Nacobbus* (1), *Paratrichodorus* (3), *Pratylenchus* (3), *Radopholus* (1), *Steinernema* (2), *Subanguina* (1), *Trichodorus* (1) and *Xiphinema* (1).

Test 2.4 COI: *Aphelenchoides* (5), *Bursaphelenchus* (3), *Cactodera* (1), *Globodera* (3), *Heterodera* (4), *Heterorhabditis* (1), *Laimaphelenchus* (1), *Longidorus* (1), *Meloidogyne* (8), *Nacobbus* (1), *Pratylenchus* (6), *Radopholus* (1), *Steinernema* (2) and *Xiphinema* (1).

It should be recognized that the potential for amplification and sequencing with the generic primers in this appendix is much larger.

4.3 Selectivity

Selectivity does not apply as individual specimens are used.

4.4 Diagnostic sensitivity

TPS partners In the EUPHRESCO II DNA Barcoding Project analysed five DNA samples of the following species: *Aphelenchoides besseyi*, *Aphelenchoides fragariae*, *Bursaphelenchus xylophilus*, *Ditylenchus dipsaci* and *Meloidogyne chitwoodi*. The overall diagnostic sensitivity obtained was 96% (*A. besseyi* 89%, *A. fragariae* 89%, *B. xylophilus* 100%, *D. dipsaci* 100% and *M. chitwoodi* 100%). One partner incorrectly analysed the sequence data for both *Aphelenchoides* species. Reanalysis of the data provided by partners shows that identification at the required taxonomic level as listed in Table 5 is possible and an overall diagnostic sensitivity of 100% could be obtained.

4.5 Reproducibility

The same DNA samples are analysed by different partners. Therefore, in this situation, the reproducibility is identical to diagnostic sensitivity.

One of the TPS participants reported that they also obtained nonspecific amplicons during PCR. In such cases the PCR product of expected size should be excised from agarose gel (see also Appendix 7, section 2.5).

The outcome of data analysis is dependent on the databases used and relies on a combination of nucleotide similarity, specific clustering in tree views and the ability of end users to recognize sequence data deposited in databases which is likely to be misidentified. The analysis of sequence data using online resources and the interpretation of BLAST and MLSA results heavily depends on the

proficiency of the operators handling the data. All relevant (online) resources should be used to draw a final conclusion for the data analysis. See Appendix 7 for guidance on data analysis.

Appendix 6 – DNA barcoding Phytoplasmas

The tests below are described as they were carried out to generate the validation data provided in section 4 (unless stated otherwise) in combination with the guidance provided in Appendix 7. Other equipment, kits or reagents may be used provided that verification (see PM 7/98) is carried out.

1. General information

1.1 This appendix describes protocols for the identification of selected regulated phytoplasmas using conventional PCR followed by Sanger sequencing analysis. Table 6 shows the selected regulated organisms that have successfully been tested with the protocols described in this appendix. It is very likely that other phytoplasmas can successfully be identified using these protocols, but validation data has not been generated to support this.

1.2 These protocols were developed by Institute of Integrated Pest Management, Aarhus University, Denmark and the University of Bologna, as part of the QBOL Project financed by the 7th Framework Programme of the European Union. As part of the EUPHRESCO II DNA Barcoding Project (2013–14), the protocols were further optimized by the Food and Environment Research Agency (Fera Science Limited), United Kingdom.

1.3 Two tests in parallel are used to identify selected regulated phytoplasmas; elongation factor EF-Tu (*tuf* gene) and *16S* rDNA (see Fig. 6). Table 6 gives an overview of the loci needed for the selected regulated phytoplasmas.

1.4 Primer sequences, amplicon sizes and thermocycler settings are provided in the test-specific sections. HPLC-purified primers should be ordered to avoid nonspecific PCR amplification.

1.5 Reaction mixes are based on the MyFi™ Mix (Bioline) reagents (cat. no. BIO-25049). Tests 2.2 and 2.3 were initially validated using BIO-X-ACT™ Short Mix (Bioline) which is no longer available. Verification performed in NVWA showed that MyFi™ Mix and BIO-X-ACT™ Short Mix performed equally (van de Vossenberg, pers. comm.).

1.6 Molecular-grade water is used to set up reaction mixes; this should be purified (deionized or distilled), sterile (autoclaved or 0.22- μm filtered) and nuclease-free.

1.7 Amplification is performed in a Peltier-type thermocycler with a heated lid, e.g. C1000 (Bio-Rad).

Table 6. Regulated phytoplasmas successfully identified with barcoding protocols

	Tests*	
Regulated organism	2.2 EF-Tu	2.3 16S rDNA
'*Candidatus* Phytoplasma mali'	x	x
'*Candidatus* Phytoplasma pruni'	x	x
'*Candidatus* Phytoplasma prunorum'	x	x
'*Candidatus* Phytoplasma pyri'	x	x
'*Candidatus* Phytoplasma solani'	x	x
Grapevine "flavescence dorée" MLO§	x	x

*Tests marked with 'x' need to be performed to reach reliable identification of the corresponding taxa. When multiple loci are indicated in the table, the MLSA tools in EPPO-Q-bank should be used.

§See comment in section 4.2 regarding the specificity of the protocols for grapevine "flavescence dorée" phytoplasma.

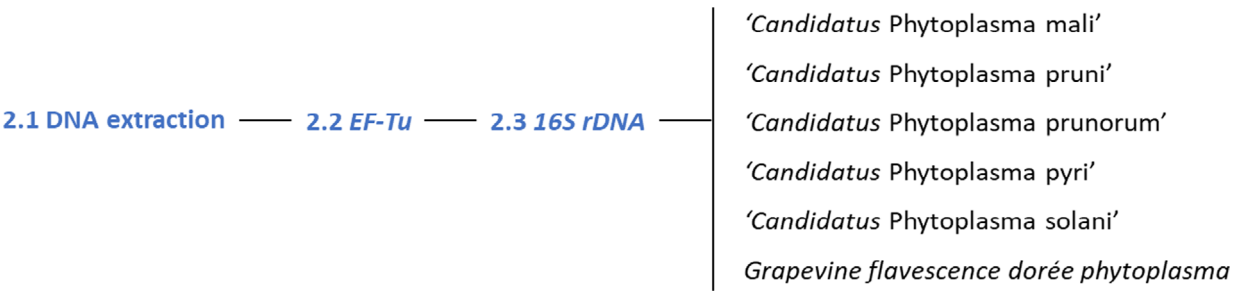


Fig. 6 Diagnostic testing scheme for identification of selected regulated phytoplasmas using DNA barcodes. The steps shown refer to the sections in this appendix which should be followed to reach reliable identification of the corresponding taxa. When sequence data of multiple loci are generated, the MLSA tools in EPPO-Q-bank need to be used.

2. Methods

2.1 Nucleic acid extraction and purification

2.1.1 Place 1 g of fresh or frozen plant tissue in a pre-cooled, sterile and dry mortar and add liquid nitrogen.

2.1.2 Homogenize the plant tissue using a sterile porcelain pestle.

2.1.3 Add 100 mg of the homogenized tissue to a pre-cooled microcentrifuge tube.

2.1.4 Alternatively, 100 µL of plant sap can be used for DNA extraction.

2.1.5 Proceed with DNA extraction using the DNeasy Plant Mini Kit (cat. no. 69104) according to the manufacturer's instructions (Qiagen).

2.1.6 No DNA clean-up is required after DNA extraction.

2.1.7 The extracted DNA should either be used immediately or stored at −20°C or below until use.

2.2 Conventional PCR EF-Tu phytoplasmas

2.2.1 PCR sequencing of 480 bp (amplicon size nested-PCR including primers) of the elongation factor Tu (*EF-Tu*) gene is adapted from Makarova *et al*. (2012).

2.2.2 Primer sequences are described in the table below. The Tuf340 PCR primer cocktail is prepared by pooling an equal volume of 10 µM of primers Tuf340a and Tuf 340b. The Tuf890 PCR primer cocktail is prepared by pooling an equal volume of 10 µM of primers Tuf890ra, Tuf890rb and Tuf 890rc. The Tuf400 PCR primer cocktail is prepared by pooling an equal volume of 10 µM of primers Tuf400a, Tuf400b, Tuf400c, Tuf400d and Tuf 400e. The Tuf835 primer cocktail is prepared by pooling an equal volume of 10 µM of primers Tuf835ra, Tuf835rb and Tuf 835rc.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for: PCR	nested-PCR	Sequencing
Tuf340a	GCTCCTGAAGAA ARAGAACGTGG	X		
Tuf340b	ACTAAAGAAGAAA AAGAACGTGG	X		
Tuf890ra	ACTTGDCCTCTTTCK ACTCTACCAGT	X		
Tuf890rb	ATTTGTCCTCTTTCW ACACGTCCTGT	X		
Tuf890rc	ACCATTCCTCTTTCA ACACGTCCAGT	X		
Tuf400a (M13-tagged)	caggaaacagctatgaccGA AACAGAAAAAC GTCAYTATGCTCA*		X	
Tuf400b (M13-tagged)	caggaaacagctatgacc GAAACTTCTAAAAGA CATTACGCTCA		X	
Tuf400c (M13-tagged)	caggaaacagctatgaccGA AACATCAAAA AGACAYTATGCTCA		X	
Tuf400d (M13-tagged)	caggaaacagctatgacc GAAACAGAAAAA AGACAYTATGCTCA		X	
Tuf400e (M13-tagged)	caggaaacagctatgaccCAAA CAGCTAAAAGACA TTATYCTCA		X	
Tuf835ra (M13-tagged)	tgtaaaacgacggccagtA ACATCTTCWACH GGCATTAAGAAAGG		X	
Tuf835rb (M13-tagged)	tgtaaaacgacgg ccagtAACACCTTCAAT AGGCATTAAAAAWGG		X	
Tuf835rc (M13-tagged)	tgtaaaacgacggcc agtAACATCTTCTATA GGTAATAAAAAAGG		X	
M13rev-29	caggaaacagctatgacc			X
M13uni-21	tgtaaaacgacggccagt			X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.2.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
Tuf340 primer cocktail	10 µM total	0.5	0.2 µM
Tuf890 primer cocktail	10 µM total	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.4 Thermocycler profile: 5 min 95°C, 35× (30 s 94°C, 30 s 54°C, 60 s 72°C), 10 min 72°C.

2.2.5 The PCR test results in a 550 -bp PCR product.

2.2.6 Two microliters of 1/30 diluted PCR product should be used as input for the nested-PCR test.

2.2.7 Master mixes for the nested-PCR are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
Tuf400 primer cocktail	10 µM total	0.5	0.2 µM total
Tuf835 primer cocktail	10 µM total	0.5	0.2 µM total
Subtotal		23.0	
1/30 diluted PCR product		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.2.8 Thermocycler profile: 5 min 95°C, 35× (30 s 94°C, 30 s 54°C, 60 s 72°C), 10 min 72°C.

2.2.9 Cycle sequencing reactions are performed using the primers targeting the respective M13 tags in separate reactions.

2.2.10 The *EF-Tu* gene is a protein coding region. Translation Table 11 (Bacterial, Archaeal and Plant Plastid Code) applies to the *EF-Tu* gene.

2.2.11 The M13-tailed primer cocktail Tuf400/Tuf835 results in an *EF-Tu* sequence with a codon starting in reading frame 2 of the primer-trimmed consensus sequence.

2.3 Conventional PCR 16S rDNA – phytoplasmas

2.3.1 PCR sequencing of approximately 600 bp (amplicon size including primers) of the *16S* ribosomal DNA (*16S* rDNA) is adapted from Makarova *et al*. (2012).

2.3.2 Primer sequences are described in the table below.

Primer name	Primer sequence (5′–3′ orientation)	Primer used for	
		PCR	Sequencing
P1-ATT (M13-tagged)	caggaaacagctatgacc AAGAGTTTGAT CCTGGCTCAGG*	X	
P625r (M13-tagged)	tgtaaaacgacggccagtA CTTAYTAAACCG CCTACRCACC	X	
M13rev-29	caggaaacagctatgacc		X
M13uni-21	tgtaaaacgacggccagt		X

*Lower case characters indicate the universal M13 tails. These tails play no role in amplification of the target but are used for generating cycle sequence products.

2.3.3 Master mixes are prepared according to the table below.

Reagent	Working concentration	Volume per reaction (µL)	Final concentration
Molecular-grade water	NA	9.5	NA
MyFi™ mix (Bioline)*	2×	12.5	1×
P1-ATT (M13-tagged)	10 µM	0.5	0.2 µM
P625r (M13-tagged)	10 µM	0.5	0.2 µM
Subtotal		23.0	
Genomic DNA extract		2.0	
Total		25.0	

*Or other verified PCR master mixes containing a polymerase with proofreading activity.

2.3.4 PCR cycling parameters: 5 min 95°C, 35× (30 s 94°C, 30 s 54°C, 60 s 72°C), 10 min 72°C.

2.3.5 Cycle sequencing reactions are performed using the primers targeting the respective M13-tags in separate reactions.

2.3.6 *16S* rDNA is a non-coding but conserved locus that is transcribed in *16S* ribosomal RNA. Translation tables do not apply to *16S* rDNA.

3. Essential procedural information

3.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each series of nucleic acid extraction and amplification of the target organism and target nucleic acid, respectively.

- Negative isolation control (NIC) to monitor contamination during nucleic acid extraction: DNA extraction of an Eppendorf tube containing 100 µL of molecular-grade water.
- Negative amplification control (NAC) to rule out false positives due to contamination during the preparation of the reaction mix: amplification of molecular grade water that was used to prepare the reaction mix.
- Positive amplification control (PAC) to monitor the efficiency of the amplification: amplification of gBlock EPPO_PAC_Phytoplasmas_1 (0.1 ng μL^{-1}) (see Appendix 9) or genomic DNA of a relevant target organism (see Table 6).

3.2 Interpretation of results

Verification of the controls

- NIC and NAC: no band is visualized.
- PAC: a band of the expected size is visualized.

When these conditions are met

- Tests yielding amplicons of the expected size are used for cycle sequencing.
- Tests should be repeated if any contradictory or unclear results are obtained.

4. Performance characteristics available

Performance characteristics for the tests in this appendix were determined under the EUPHRESCO DNA Barcoding Project in an international consortium of 10 participants (see EPPO validation sheet, https://dc.eppo.int/validation_data/dwvalidation?id=118). Additional data was generated by the Dutch NPPO laboratory and Fera Science Limited, GB.

4.1 Analytical sensitivity

Concentration of the DNA extract was determined with Qubit dsDNA Assay (Invitrogen). For all protocols a DNA extract dilution of approximately 30 ng μL^{-1} and a relative infection grade of 10% (i.e. 10× dilution) is sufficient to generate an amplicon that can be sequenced, leading to a consensus sequence with a HQ (Phred > 40) of at least 98%. For more information, see the relevant EPPO validation sheet.

4.2 Analytical specificity

The locus or combination of loci indicated in Table 6 possess sufficient interspecies variation to allow the identification to the '*Candidatus* Phytoplasma' species level. In addition to the '*Candidatus* Phytoplasma' species listed in Table 6, phytoplasma strains related to the following have successfully been amplified and sequenced using the protocols in this appendix by the Dutch NPPO: '*Ca*. P. asteris', '*Ca*. P. aurantifolia', '*Ca.* P. phoenicium' and '*Ca*. P. trifolii'.

The tests described in sections 2.2 and 2.3 allow the discrimination of subgroups of ribosomal group 16SrV but not between strains within a subgroup. Strains of "flavescence dorée" phytoplasma belong to subgroups 16SrV-C and -D. Since alder yellows phytoplasma and Palatinate grapevine yellows phytoplasmas are also members of subgroup 16SrV-C, the protocol is not specific for grapevine "flavescence dorée" phytoplasma strains of subgroup 16SrV-C.

4.3 Selectivity

'*Ca.* P. mali', '*Ca.* P. prunorum', '*Ca.* P. pyri' and two strains of '*Ca.* P. solani' have been tested from *Malus*, *Prunus domestica* 'St Julien', *Pyrus* and *Catharanthus roseus*, respectively. Other matrices might apply and need to be verified by end-users before implementing the tests described in this appendix.

4.4 Diagnostic sensitivity

TPS partners in the EUPHRESCO II DNA Barcoding Project analysed five DNA samples of the following '*Candidatus* Phytoplasma' species: '*Ca.* P. mali', '*Ca.* P. prunorum', '*Ca.* P. pyri' and two strains of '*Ca.* P. solani'. The overall diagnostic sensitivity obtained was 96% ('*Ca.* P. mali' 100%, '*Ca.* P. prunorum' 90%, '*Ca.* P. pyri' 100% and '*Ca.* P. solani' 90% and 100%). Reanalysis of the data provided by partners shows that identification at the required taxonomic level as listed in Table 6 is possible and an overall diagnostic sensitivity of 98% could be obtained.

4.5 Reproducibility

The same DNA samples are analysed by different partners. Therefore, in this situation the reproducibility is identical to diagnostic sensitivity.

The outcome of data analysis is dependent on the databases used and relies on a combination of nucleotide similarity, specific clustering in tree views and the ability of end users to recognize sequence data deposited in databases which is likely to be misidentified. The analysis of sequence data using online resources and the interpretation of BLAST and MLSA results heavily depends on the proficiency of the operators handling the data. All relevant (online) resources should be used to draw a final conclusion for the data analysis. See Appendix 7 for guidance on data analysis.

Appendix 7 – Sanger sequencing, consensus preparation and data analysis

1. General information

1.1 This appendix describes how to generate sequence data, how to create a consensus sequence and how to analyse data using online resources. This appendix may also contain useful information for the analysis of sequences of viruses and viroids (although they do not have DNA barcodes).

1.2 Sequence data files containing chromatograms (also referred to as electropherograms or trace data, e.g. *.ab1, *.abi or *.scf) and quality scores (Phred scores) are used as input for consensus sequence preparation and data analysis. The sequence data files are sometimes referred to as reads.

1.3 The use of sequence data files without chromatograms (e.g. *.seq, *.fas or *.txt) for consensus sequence preparation might lead to unreliable results.

1.4 Sequencing analysis software that allows alignment and editing of sequence data containing chromatograms with Phred scores is essential for the creation of reliable consensus sequences [e.g. the Lasergene software package (DNAstar), CLC genomic workbench (CLC bio) or Geneious (Biomatters)].

1.5 Access to the Internet is needed to access online databases such as NCBI GenBank, BOLD and EPPO-Q-bank.

2. Sanger sequencing

Other equipment, kits or reagents may be used provided that validation or verification (see PM 7/98) is carried out.

2.1 PCR products, together with the primers used for the sequencing reaction, can be sent to commercial companies for Sanger sequencing.

2.2 All of the indicated marker regions should be sequenced in forward and reverse directions as indicated under the specific test sections.

2.3 Sequencing primers indicated in the primer tables (Appendices 1–6) should be provided to the commercial company.

2.4 If multiple PCR products (>100 bp) are visible after amplification, the PCR product of expected size (see test-specific sections in Appendices 1–6) should be excised from the agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen) before sending it for sequencing.

Below an example is provided of the steps that could be taken when PCR products are sequenced in-house:

2.5 Purify PCR products using a QIAquick PCR Purification Kit (Qiagen). Purified PCR product is eluted in 30–50 µL of elution buffer (provided). If multiple PCR products are visible on agarose gel after amplification, the PCR product of expected size (see test-specific sections in Appendices 1–6) should be excised from the agarose gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

2.6 Separate cycle sequencing reactions are performed for each primer (see specific protocols) using BigDye Terminator v. 1.1 or v. 3.1 Cycle Sequencing Kits (Life Technologies) according to the manufacturer's instructions.

2.7 Cycle sequence products are purified using Sephadex G50 columns in 96-well multiscreen HV plates (Millipore) or the DyeEx 2.0 spin kit (Qiagen).

2.8 An equal volume of HiDi formamide (Life Technologies) should be added to the purified cycle sequence product.

2.9 Analyse the purified cycle sequence product: HiDi formamide on a Sanger sequence platform (e.g. 3500 Genetic Analyzer, Life Technologies).

2.10 Generated chromatograms are used to create a single consensus file.

3. Consensus sequence preparation

In general, overlapping sections are used to generate consensus sequences. When needed (e.g. when discriminatory sequences are located in parts of the consensus sequence covered by a single Sanger sequence read), sections that are covered only once can be included in the consensus sequence. Visual inspection of the assembly is an important part of the creation of a consensus sequence. Phred scores can be used to aid consensus sequence creation as they indicate the reliability of base-calling: a Phred score of 10 = 90%, 20 = 99%, 30 = 99.9%, 40 = 99.99% and 50 = 99.999% reliability for the selected base. Phred scores> 40 are regarded as high-quality (HQ) data.

3.1 Upload the chromatograms in the sequencing analysis software.

3.2 Select the chromatograms (at least two) needed for the preparation of consensus sequences. Chromatograms can be generated using, for instance, a forward and reverse primer (e.g. *COI* gene arthropods) or two reverse primers (e.g. *16S* rRNA gene, bacteria). In some cases, multiple PCR products are used to generate a single consensus sequence (e.g. *18S* rRNA gene, nematodes).

3.3 Assemble the chromatograms so that an alignment is obtained that shows the electropherograms of the individual reads.

3.4 Trim 3′ untemplated –dA from the consensus sequence.

3.5 Trim amplification primers from the consensus sequence. Internal sequence primer sequences can be retained. Appendix 8 shows a suggested form for preparation of consensus sequences and data analysis.

3.6 Assess the assembly visually and edit where needed. Check the entire sequence in order to detect any errors in the assembly and consensus sequence. The following

Table 7. IUPAC ambiguity codes

Code	Represents	Complement
A	Adenine	T
G	Guanine	C
C	Cytosine	G
T	Thymine	A
Y	Pyrimidine (C or T)	R
R	Purine (A or G)	Y
W	Weak (A or T)	W
S	Strong (G or C)	S
K	Keto (T or G)	M
M	Amino (C or A)	K
D	A, G, T (not C)	H
V	A, C, G (not T)	B
H	A, C, T (not G)	D
B	C, G, T (not A)	V
N	Any base	N
-	Gap	-

rules are used as a guide. Visual inspection of the assembly might lead to different decisions.

- Trim the low quality ends of the consensus sequence to prevent an unreliable consensus sequence because of low-quality bases: (i) for 1× coverage the Phred score should be at least 30 for the individual read; (ii) for 2× or more coverage it should be at least 20 for the individual reads.
- Bases in the consensus sequence with a Phred score <20 should be noted as N.
- Make sure that the consensus sequence is shown in the right direction (5′–3′ from the forward primer; see primer tables in Appendices 1–6). This is particularly important when using the BOLD database for data analysis. When using a consensus sequence that has the wrong direction, BOLD will not be able to match the sequence to other sequences in the database.
- When polymorphisms (double peaks) are observed in good-quality data, IUPAC ambiguity codes should be used (see Table 7).
- When insertions or deletions (InDels) are present in coding sequences (the presence of InDels can be inferred by analysing the BLAST hit alignment), the consensus sequence can be converted to amino acids in order to check that there are no unexpected stop codons in the coding sequence (note that the correct reading frame should be used; see test-specific sections in Appendices 1–6).

3.7 Generate a consensus sequence from the assembly.

4. (Online) data analysis

Relevant resources should be used to draw a final conclusion for the data analysis. There are several online resources available that can be used for the analysis of the consensus sequence obtained. A detailed description of the different resources and the interpretation of BLAST results are shown in section 5.

4.1 Document all (online) resources consulted, the settings used, results and conclusions per source. Appendix 8 shows a suggested form for preparation of consensus sequences and data analysis.

4.2 Document the results per resource used (e.g. by providing screenshots or pdf-files of BLAST hits, MLSA results, tree views, alignments, etc.).

4.3 Draw a general conclusion from the conclusions per source, making use of conservative terms (e.g. Sample X possibly is/isn't taxon Z, or it is (very) likely/unlikely that Sample X is taxon Z) avoid using absolute terms (e.g. Sample X is taxon Z).

4.4 When a misidentification of an accession in the online databases is suspected, end users can BLAST the sequence of the presumed misidentified organism against 'NCBI + organism' (see section 5.3) to determine the reliability of the identification.

4.5 It has to be noted that PCR sequencing is used in support of species identification. Origin, host plant and other characteristics (e.g. morphological, biochemical, reactions on indicator plants) are typically needed to complete the diagnosis.

5. Essential procedural information

5.1 Controls

For a reliable test result to be obtained, the following external controls should be included for each sequencing run and derived consensus sequence(s) generation and sequence analysis:

Positive cycle sequence control (PCC) to monitor the efficiency of cycle sequence reactions, the generation of sequence data and consensus sequence preparation: amplicon of a sample with known identity and sequence analysis as a sequencing process control (e.g. amplicons obtained with synthetic PACs or DNA with a known sequence). The percentage of high-quality bases and the sequence length obtained from this sample are indicative for cycle sequence reactions and the generation efficiency for sequence data. Alignment of the PCC consensus sequence with the known reference sequence (should be 100%) is indicative of the success of consensus sequence preparation.

Generating consensus sequences heavily depends on the proficiency of the operators handling raw data. The same applies to the interpretation of BLAST results. Synthetic PACs are standardized controls that can be used to unambiguously monitor success from cycle sequence reaction to sequence analysis. Between-run repeatability for individual operators and the overall reproducibility within a laboratory can be used to monitor trends in sequence analysis success.

In addition, the proficiency of operators working with sequencing analysis can be monitored using blind samples with known sequences or by participation in proficiency tests.

When unclear results are obtained, sequence data is analysed by a second operator or the test is repeated.

5.2 Validation

Determining performance criteria for DNA barcoding is performed in two separate steps: (1) PCR reactions (all performance criteria described in PM 7/98 apply unless stated otherwise in Appendices 1–6) and (2) creating consensus sequences and sequence analysis (only the performance criteria analytical specificity, diagnostic sensitivity and reproducibility are relevant).

The analytical specificity of the locus (or combination of loci) used can change over time because of the use of (online) databases with constantly changing content. Changes made to the content of (online) databases might influence the usability of generated sequence data for the identification on the required taxonomic level. Instead of determining performance criteria for the sequence data analysis step, the usability of generated data (i.e. analytical specificity) is evaluated each time an analysis is performed by determining if the generated data provides sufficient resolution between taxa (e.g. no overlap in inter- and intraspecific variation, or taxon-specific clustering). The validation status of a species–locus (loci) combination relies on the last time that combination was assessed. The protocols in this Standard have proven to be fit for purpose for the selected regulated pests. Only selected regulated pests that were previously tested by the authors of this Standard have been included in the Standard, but it should be noted that these protocols can be used for a much broader range of (nonregulated) organisms. Laboratories implementing these protocols have to verify each time that an analysis is performed that the resolution of the generated sequence(s) still allows species identification.

Synthetic PACs can be used to determine the repeatability and reproducibility of the sequence analysis steps (see Appendix 7, section 5.1).

5.3 Background information on online resources

The most commonly used online databases and their application are described in the table below. Terms used in the table are explained in a glossary.

5.3.1 Glossary.

BLAST	In a BLAST search, a sequence is broken into small pieces (word size) that are matched with the data in the database (seeds). Rewards and penalties for matching and mismatching bases are awarded. Changing the scoring settings of the algorithm parameters can greatly influence the BLAST (especially the gap penalty) output which consists of hit names, accession numbers, max score, total score, E-value, coverage and similarity.
Max score	Highest alignment score (bit score) between the query sequence and the database sequence segment. The scores of different alignments cannot be compared, nor can they be used to select the best alignment because their scale depends on the gap penalty.
Total score	Sum of alignment scores of all segments from the same database sequence that match the query sequence (calculated over all segments). This score is different from the max score if several parts of the database sequence match different parts of the query sequence. The scores of different alignments cannot be compared, nor can they be used to select the best alignment because their scale depends on the gap penalty.
E-value	The E-value (Expect value) indicates the reliability of the hit, and the closer it is to zero the more 'significant' a hit is (note: the hit, not the identity of the specimen!). BLAST hits are typically sorted on E-value (low to high). The first BLAST hit (lowest E-value) is not necessarily the most likely species identity. Particularly when sequence data with large changes in query coverage are present in the database the E-value can be unreliable to identify the best match. Because of this, tree views of the obtained BLAST hits are used to further determine the identity of the sequenced specimen.
Consensus	A theoretical representative sequence in which each nucleotide is the one which occurs most frequently at that site in the different sequences (e.g. sequences generated with the forward primer and reverse primer of a given amplicon in separate reactions). It is the results of multiple sequence alignments in which related sequences are compared to each other.
Coverage	Percentage of the query length that is included in the aligned segments. This coverage is calculated over all segments.
Similarity	Percentage of identical bases in the alignment. The percentage is calculated over all segments.
MLSA	In multilocus sequence analysis (MLSA) or multilocus sequence typing (MLST), sequence data of more than one locus is analysed simultaneously.
Gap penalty	If the gap penalty is too large, gaps are avoided and the sequences cannot be properly aligned. If the gap penalty is too low, gaps are inserted everywhere to prevent mismatches. This does not produce any informative alignment. The 'best' alignment is obtained for an intermediate gap penalty.

5.3.2 Commonly used online databases

	NCBI GenBank	BOLD	EPPO-Q-bank
Hyperlink	http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome	http://www.boldsystems.org/index.php/IDS_OpenIdEngine	https://qbank.eppo.int/
Database description	NCBI GenBank sequence database is a publicly accessible database containing sequence data of more than 260 000 formally described species (Benson *et al*., 2013). The sequence data in the NCBI database consists of a many loci from all organisms groups that are relevant to the plant health field (bacteria, fungi, oomycetes, insects, invasive plant species, nematodes, phytoplasmas, viruses and viroids). Many quarantine and quality organisms, phylogenetically related species and look-alikes are represented in this database. Data in NCBI is checked on various technical aspects before publication. Through the taxonomy database (select "Taxonomy" in the dropdown menu on the NCBI website), it is possible to see which organisms are present in the NCBI database.	The BOLD database (Ratnasingham & Hebert, 2007) is the DNA BARCODE sequence database for the identification of animalia, fungi and plants. The database includes *COI* for animalia, *ITS* for fungi and *rbcL* and *matK* for plants. Sequence data in BOLD have to meet strict requirements to ensure species identity of the specimens in the database, but errors remain. Specimens and strains used to generate sequence data are vouchered. The *COI* database can be used for arthropods and nematodes identification. Although the main focus of BOLD lies with *COI* sequences for animalia, the *ITS*, and the *rbcL* and *matK* databases can be useful for fungi and invasive plants, respectively.	EPPO-Q-bank is a scientifically curated database that focuses specifically on regulated pests (including pathogens, and invasive plants) and related species. Sequence data of most pest 'barcodes' that are generated with the protocols described in this standard are available. Specimens and strains used to generate the EPPO-Q-bank sequence data are vouchered and can often be acquired via the curator of a database section.

(continued)

Table *(continued)*

	NCBI GenBank	BOLD	EPPO-Q-bank
Database subsets	The NCBI database knows many subsets such as: - Nucleotide collection (nr/nt): "nr" stands for "non-redundant," but it isn't - Reference genomic sequences (refseq_genomic): Comprehensive, integrated, non-redundant, well-annotated set of sequences - NCBI Genomes (chromosome): Complete genomes and chromosomes from Reference Sequences Typically the nr/nt database is used. End users have to be aware that this database contains misidentified or not yet published sequences, and especially sequences from nonvouchered materials. Additional analyses can be performed to determine if a sequence is derived from a misidentified specimen (e.g. analysis in other databases, BLAST of putative misidentified sequence to nr/nt database restricted to species identity).	Within the *COI* database (Animalia) several subsets of the database can be used: - All Barcode Records on BOLD - Species level Barcode Records - Public Record Barcode Database - Full-Length Record Barcode Database The first three options require a *COI* fragment of at least 500 bp for identification, while the 'Full-length record barcode database' needs at least 640 bp. The first-mentioned database ('All barcode records') also contains sequence data from specimens which are not identified to species level and is less suitable for species identification. By default, 'Barcode species level records' is selected. The *ITS* database does not have subsets in the database and requires a fragment of at least 100 bp in order to perform a BLAST search. The database contains *ITS* sequence data from specimens which are not identified to species level and therefore does not have the same status as the 'Species level barcode records' *COI* database. The *rbcL* and *matK* database does not have subsets in the database, and requires a fragment of at least 500 bp to perform a BLAST search. The *rbcL* and *matK* database contains sequence data from specimens which are not identified to species level and therefore does not have the same status as the 'Species level barcode records' *COI* database. There are very few *rbcL* and *matK* records on BOLD. End users have to be aware that this database contains misidentified or not yet published sequences, and especially sequences from nonvouchered materials.	The EPPO-Q-bank database is composed of seven subsets: arthropods, bacteria, fungi, invasive plants, nematodes, phytoplasmas, and viruses and viroids. The BLAST algorithm can be used to query all sequences in the entire database while the MLSA tools (named Multi-locus ID) are accessed through the discipline-specific subset of the database.
Frequently used analysis tools	Single locus Basic Local Alignment Search Tool (BLAST)	Single locus BLAST	Single locus BLAST Multilocus BLAST

(continued)

Table *(continued)*

	NCBI GenBank	BOLD	EPPO-Q-bank
BLAST and MLSA parameters	In NCBI three BLAST pre-sets are available: megablast, discontinuous megablast and blastn. Megablast is designed for the comparison of sequences with high similarity (>95%) and is in those cases very quick. Megablast utilizes a large word size (n = 28). Discontinuous megablast makes use of a smaller word size (n = 11) in which mismatches are allowed. Genbank indicates that this is particularly useful for comparison across species. Blastn is the slowest algorithm, and also makes use of a word size n = 11, but if desired, this can be adjusted to 7. Megablast is used by default, but if this does not yield useful hits other algorithms can be used. Under the heading 'Algorithm parameters' settings as (e.g.) number of hits to be shown, word size, match/mismatch scores, number of displayed results can be changed. It is possible to restrict BLAST to a specific taxon or taxa (e.g. genus, species, subspecies), or to exclude certain taxa. To do so, type the name of the desired taxon or taxa in the "Organism" field on the BLAST page. It should be noted that not all sequences in NCBI have a taxonomic name assigned to it and could be missed in the selection you made. Also, synonyms are not taken into account. It has to be noted that BLAST results restricted to a specific taxon sometimes show different similarity percentages in the hit table compared to the alignment. Usually the latter shows the correct percentage. It is also possible to compare the query sequence only to one or a restricted number of genbank accessions, e.g. those from reference materials. To do so, type the genbank accession number of the reference material into the "Entrez Query" field.	It is not possible to adjust the BLAST settings in BOLD.	**BLAST** *Note:* From the EPPO-Q-bank homepage, the BLAST search can be accessed through '*Blast against all Q-bank sequences'*. It can also be accessed from the discipline-related sections of the database. Under 'show advanced parameters', different BLAST settings such as word size, maximum alignments to display, and cut-off settings for minimum similarity and overlap can be adjusted. In general, the default settings are appropriate, but it is important to check which databases are selected for the search. **MLSA (Multi-locus ID tool)** *Note:* MLSA is accessed under ID in the discipline-specific subsets of the database. Sequences of different loci can be submitted. Under 'show advanced parameters', different BLAST settings such as word size, maximum alignments to display, and cut-off settings for minimum similarity and overlap can be adjusted.

(continued)

Table *(continued)*

	NCBI GenBank	BOLD	EPPO-Q-bank
BLAST and MLSA output	BLAST results are by default displayed in three different ways: Graphic summary, a BLAST hit table (Descriptions) and a detailed overview per hit (Alignments). The Graphic summary shows the length of the query sequence (Sbjct) and the hit lengths and their position relative to the query sequence. The hit table shows, among others, the name of the hits, their accession number, the coverage with respect to the query sequence, the percentage similarity, and the E-value (Expect value). The detailed overview per hit gives information about the percentage of agreement, overlap, an alignment between query and Sbjct and information relating to the accession number (e.g. locus). Simultaneous BLAST of multiple sequence items is possible to increase the sequence analysis throughput.	Apart from the 'All barcode database records', the BLAST results of *COI* sequence data will be displayed as a hit table with similarity percentages, a graph showing the similarity scores, and a probability that the sequence belongs to a particular taxonomic level (Identification summary). The 'All records barcode identification database' gives no Identification summary. BOLD does not account for synonyms, so it is possible that the identification summary states that a certain sequence belongs to either species A or B, while A and B are synonyms. The *ITS* and *rbcL* and *matK* databases show BLAST results largely in the same way as NCBI. Additionally, graphs with similarity scores and E-values are given. Simultaneous search of multiple sequence items is possible after registration.	**BLAST** BLAST results are displayed as a hit table showing, among others, the name of the hits, the coverage with respect to the query sequence (% overlap) and the percentage similarity. Furthermore, the orientation of inputted sequence with respect to the hit is displayed under Direction (+/+ or +/–). In EPPO-Q-bank, The E-value is referred to as probability. In addition, the score is shown. Alignments can be accessed by expanding the hit results (click on the '+' sign next to the hit). Simultaneous BLAST of more than one sequence is not possible. **MLSA (Multi-locus ID tool)** In the MLSA results, EPPO-Q bank shows the number of loci that are included in the analysis ("Accounted"). Also, the degree of similarity is displayed. Alignments of different loci can be accessed by expanding the hit (click the '+' sign next to the hit).
Tree views (see also 5.4 for the interpretation of tree views)	BLAST hit results can be displayed as a Fast Minimum Evolution (FME) tree or Neighbour Joining (NJ) tree view by selecting "Distance tree of results" on the BLAST results page. Selecting "show all" under "collapse Mode" will allow to assess if a query sequence (highlighted in yellow) falls in a species-specific clade.	*COI* BLAST hit results can be displayed as a Neighbour Joining (NJ) tree view by selecting "Tree based identification" on the BLAST results page. Tree settings cannot be adjusted. The query sequence is highlighted in red. *ITS* and *rbcL* and *matK* BLAST hits cannot be shown in a tree view.	Single BLAST hit results can be displayed using different tree views by selecting 'Tree' on the BLAST results page and then by clicking on 'Neighbour joining' or 'UPGMA'. The query sequence is indicated with 'My_query'. Apart from choosing the tree algorithm, the display of the trees can be adjusted.
Species included	Through the taxonomy database (select 'Taxonomy' in the dropdown menu on the NCBI homepage), it is possible to see which organisms are represented in the NCBI database.	Through the taxonomy database (select the 'Taxonomy' tab on the BOLD homepage) it is possible to see which species are present in the BOLD database.	Overviews of the number of species (specimen and sequences) included in the EPPO-Q-bank database are provided for the entire database (General page) and for the discipline-related subsets of the database (home page of the specific subset). The lists of species included in each discipline-related subset of the database are available in the "organisms included" pages.

5.4 Interpretation of tree views

Tree views obtained from BLAST and MLSA results are used in addition to BLAST hits for reliable species identification. It should be noted that the usefulness of tree views is, similar to the interpretation of BLAST and MLSA hits, highly dependent on the availability of relevant loci and taxa in the database consulted. Furthermore, the implemented algorithms for multiple sequence alignments (ClustalW) and tree construction (fast minimum evolution, neighbour joining) do in some cases not show/optimally reflect the species position within the tree depending on the genetic variation of the chosen loci and the number of taxonomic differences from the reference sequences available in the database. In principle, an unknown sequence can be assigned to a particular taxon when it falls within a taxon-specific cluster.

It is important to realize that trees generated from (partial) gene sequences or sequence data from noncoding regions only show the relationship between these (partial) genes or regions and do not necessarily show a phylogenetic relationship among the taxa. To infer phylogenetic relationships more in-depth analyses are necessary (for a practical handbook see Lemey *et al*., 2009).

A tree consists of a root, branches, nodes and leaves (= external nodes) (see Fig. 7A). The external nodes show the taxa that are used. These taxa can be species, genera or families, but also subspecies or pathovars. The nodes of the tree represent the (hypothetical) ancestors or, better, represent sequences of the (hypothetical) ancestors. Groups of taxa with the same (hypothetical) ancestors form clades or clusters. When determining phylogenetic relationships, an outgroup is chosen to root the tree (= outgroup rooting) (Fig. 7A). However, when BLAST results are used to draw a tree, there is no outgroup and trees are typically midpoint rooted, which is indicated with a node on the branch between the specimens with the lowest homology (Fig. 7B). In Fig. 7A,B, all specimens of species 2 form a clade, but also all specimens of species 1 + the unknown sequence + species 2 and 3. Species 4 and 5 together form a nonspecies-specific clade. Based on the gene or region used to draw this tree, there is no resolution between species 4 and 5. If an unknown sequence would cluster in clade 4/5, identification on the basis of this tree is not possible. In this case, it can be said that the unknown sequence possibly belongs to species 4 or 5.

Different terms are used to indicate the relationship between external nodes. In Fig. 7A,B, species 2 is a sister group of species 1 + unknown sequence (and vice versa). Species 3 is again a sister group of species 1 + unknown sequence + species 2, and so on. In general, a branch splits into two branches after a node (= dichotomous). Specimens with a common (hypothetical) ancestor form a monophyletic group (e.g. all the specimens in species 2 in Figure 7A–C). A polyphyletic group consists of specimens with different (hypothetical) ancestors (e.g. species 1 in Fig. 7C). The latter can sometimes occur in trees obtained from BLAST results. Specimens of the same species may be found at different places in the tree and form a polyphyletic group. Identification is then still possible, provided that the unknown sequence clusters with a species-specific clade or with a sequence from a relevant reference material. For instance: in Figure 7(A–C) an unknown sequence is included in the analysis. In Figure 7A,B the sequence clusters with a species-specific clade which contains all specimens of this species available in the database (no overlap with other species). In Figure 7C the sequence falls in one

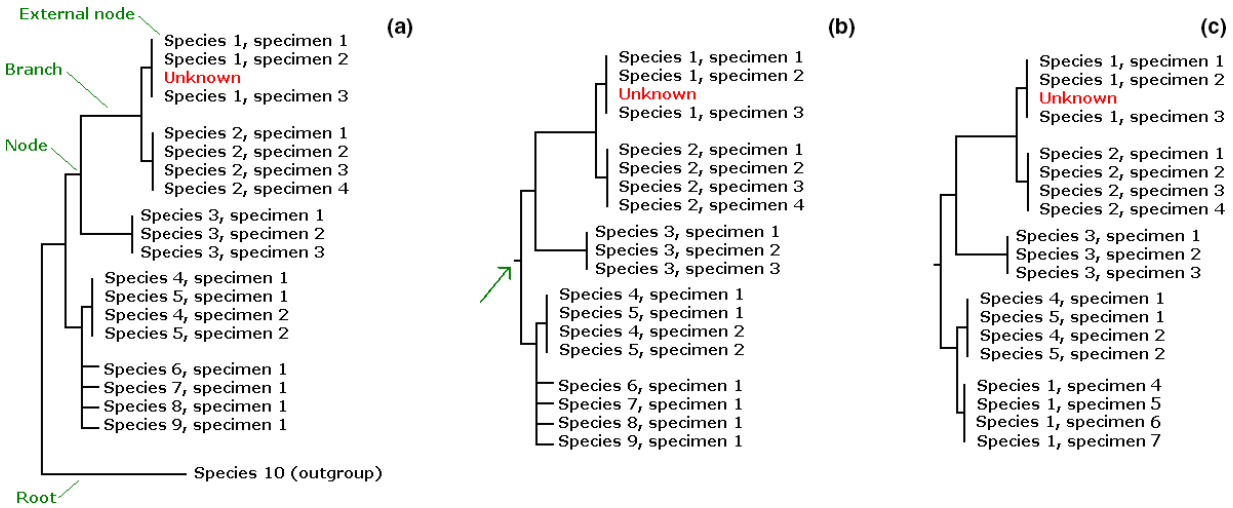


Fig. 7 (A) Outgroup rooted tree with species 1–10. Species 1, 2 and 3 form monophyletic groups, species 4 and 5 form a nonspecies-specific cluster and species 6–9 represent a polytomy. Species 10 is the outgroup in this cladogram. (B) Midpoint rooted tree. The same cladogram as in (A) but without an outgroup. This tree is rooted on the branch between the specimens with the lowest homology. (C) Midpoint rooted tree in which species 1 represents a polyphyletic group.

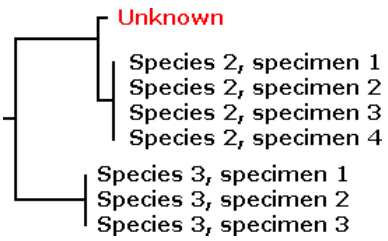


Fig. 8 The same midpoint rooted tree as Figure 7B without species 1 and 4–10.

of the species-specific clusters from the polyphyletic species 1. In both cases this provides a reasonably strong indication that the unknown sequence probably belongs to species 1. Sometimes it is not possible to determine the relationship between the different taxa (see species 6, 7, 8 and 9). This is called a polytomy. If a tree obtained from BLAST results shows a polytomy, this often indicates that the diagnostic resolution of the analysed locus or loci is not sufficient.

The usefulness of tree views is highly dependent on the sampling of the relevant taxa. If some taxa are not represented it is difficult to interpret the tree. In Figure 8, species 1 and species 4–10 (relative to Fig. 7B) are not included. It is impossible to see that the unknown sequence clusters with species 1 and might be misidentified as variation of species 2. When an unknown sequence clusters as sister to a species- specific cluster or as a single branch in a tree special caution is needed, since this could either be a result of variation within a species that has not been sequenced before or lack of sampling of other related species.

Appendix 8. – Suggested form for consensus sequence preparation and data analysis

This form can be used to document the locus/loci sequenced, sources and settings used, results obtained and conclusions drawn. It is important to document this information since databases with constantly changing content are used for identification. This appendix may also contain useful information for the analysis of sequences of viruses and viroids (although they do not have DNA barcodes).

Date: Operator:

Table 8. Information concerning locus sequenced and consensus sequence preparation (copy this table for each locus used)

1	LIMS and/or collection number	
2	Name locus	(e.g. cytochrome c oxidase subunit I)
3	Characteristics of the locus	☐ Coding ☐ noncoding ☐ mix of coding and noncoding
4	Cycle sequence reactions and sequencing performed	☐ in-house ☐ external company (*)
5	BigDye terminator kit used	☐ version 1.1 ☐ version 3.1
6	*n* cycle sequence reactions performed: consensus based on *n* chromatograms	x: x (* when not 1:1)
7a	Assembly method	☐ de novo assembly ☐ reference assembly (go to 7b)
7b	Reference sequence used (Collection- or NCBI-number)	
8	Untemplated –dA and amplification primers removed?	☐ yes ☐ no (*)
9	Are single sequence reads used in the consensus sequence	☐ yes, how many bases 5′ end: . . . and 3′ end: . . .☐ no
10	Orientation consensus sequence correct (5′–3′ from Fw primer)	☐ yes
11	Consensus length: Expected consensus length (when available)	xxx bp: xxx bp (* when not 1:1)
12	% High-quality (HQ) bases (Phred score > 40)	xxx.x %

*Provide detailed explanation below.

Explanation and additional information on locus used and consensus sequence obtained:

Table 9. Sources used, analysis settings and analysis results

Source	Analysis information	Parameters	Explanation, reference to analysis results and conclusion per database‡
NCBI	Database used	☐ Nucleotide collection (nr/nt) ☐ other (give details†)	
	Selection algorithm	☐ Megablast ☐ Discont. megablast ☐ Blastn	
	Parameters adjusted	☐ No ☐ Yes (give details)	
	Tree method	☐ Fast minimum evolution ☐ NJ	
	Restrict to organism(s) (optional)	☐ Not used ☐ Used (give details)	
	Exclude organism(s) (optional)	☐ Not used ☐ Used (give details)	
BOLD	Database used	☐ *COI* ☐ *ITS* ☐ *rbcL & matK*	
	Subset COI database (when used)	☐ All ☐ Species level ☐ Public record	
	Tree view used	☐ Not used ☐ Used (give details)	
EPPO-Q-bank	Analysis method	☐ Single locus* ☐ Multilocus (give details)	
	Parameters adjusted	☐ No ☐ Yes (give details)	
	Tree method	When applicable (give details)	
Other	When applicable provide details		

*Turn nonredundant GenBank option off.

†Provide details in the last column of the table.

‡Number of nucleotides in analysis, % similarity with 1st or specific match, specific clustering/no specific clustering with taxon Z.

Data analysis conclusion

(Draw a single conclusion from the results obtained using different resources. For instance, based on the analysis of xxx nucleotides of locus A and xxx nucleotides of locus B in database 1, 2 and 3 we can conclude that sample xxx might be/presumably is/is not taxon Z.)

Analysis results and other supportive information

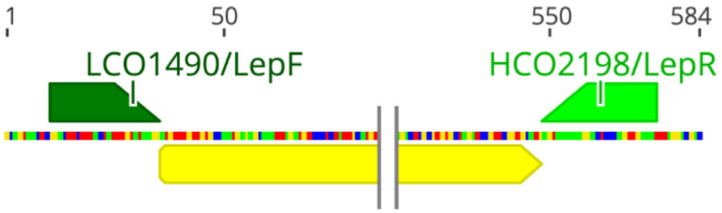
(e.g. consensus sequence(s) and print screens of BLAST hit tables, Tree Views, Alignment views, etc. with reference to Table 9 that lead to conclusions per database and to the general conclusion.)

Appendix 9. – gBlocks

The sections below provide graphical representation and background information of the gBlocks that can be used as PAC for the DNA barcoding tests. gBlocks were designed by NPPO-NL in such a way that they can be used for all tests in a single organism group (or appendix) except for the test 2.4 Conventional PCR *egl Ralstonia solanacearum* species complex (Appendix 2) and for the test 2.8 *TEF1* (Appendix 3). Dark green annotated sequences indicate annealing sites for forward primers, whereas light-green annotated sequences indicate annealing sites for reverse primers. The 513 nt reference sequence phrase is indicated in yellow and will result after translation (reading frame 1, standard code) in the following amino acid sequence twice: *KEEP*CALM*THIS*IS*-MERELY*A*VERY*STRANGE*REFERENCE*PHRA-SE*WITH*EIGHTY*FIVE*CHARACTERS (stop codons are indicated as *).

[Colour figures can be viewed at wileyonlinelibrary.com]

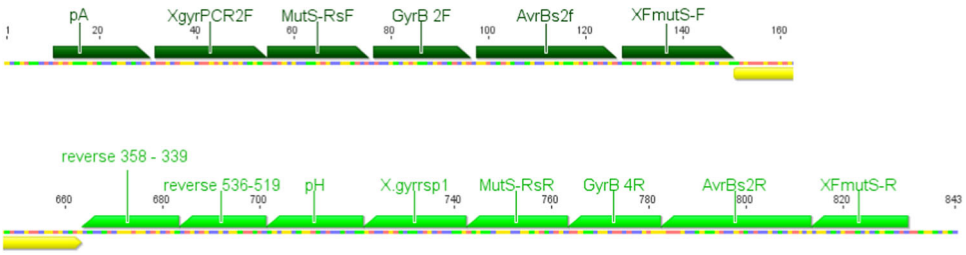
1. Arthropod tests



Note: The primers LepF, LCO1490puc-t1, LCO1490Hem1-t1 and LepR, HCO2198puc-t1, HCO2198Hem1-t1 and HCO2198Hem2-t1 are located at the same positions as the LCO1490 and HCO2198 primers, respectively, but include either mismatched (LepF/LepR) or degenerated nucleotides (LCO1490puc-t1, LCO1490Hem1-t1, HCO2198puc-t1, HCO2198-Hem1-t1 and HCO2198Hem2-t1). The mismatches in the LepF and LepR primers are mainly at the 5′ end of the primer so it is very likely they have no effect on annealing and sequencing.

gBlock name: EPPO_PAC_Arthropods_1
version: 1 [Colour figures can be viewed at wileyonlinelibrary.com]
length: 584 nt
NCBI accession: KT429638
Sequence: GCTGATTCACGGTCAACAAATCATAAAGATATTGGTAGAAAGAAGAGCCTTAGTGTGCTTTAATGTA
GACCCACATAAGCTAGATATCGTAGATGGAGCGCGAATTATACTAGGCGTAGGTTGAACGCTATTAGTCAACT
AGAGCGAATGGCGAATAGAGAGAATTTGAGCGGGAAAACTGTGAGTAGCCGCATAGAGCTAGCGAGTAGTGG
ATTACTCATTAGGAAATCGGACATACCTACTAGTTCATCGTAGAGTAGTGCCATGCACGGGCTTGCACAGAGA
GATCGTGAAAGGAGGAACCATGATGCGCACTTATGTGAACACATATTAGTTGAATATCATGAATGGAAAGAG
AGCTCTATTGAGCCTGAGTCGAGAGGTACTGAAGTACGCGTGCAAACGGAGAGTGACGTGAGTTCGAAAGAG
AGAATTGCGAATGACCTCACCGAGCATCCGAATGATGGATAACCCACTGAGAGATAGGGCATACATATTGATT
TATTGTGGAATGATGTCACGCGAGAGCATGTACCGAACGGAGCTAGTGATTTTTTGGTCACCCTGAAGTTTAA
ATGGTCGTC

2. Bacterial tests

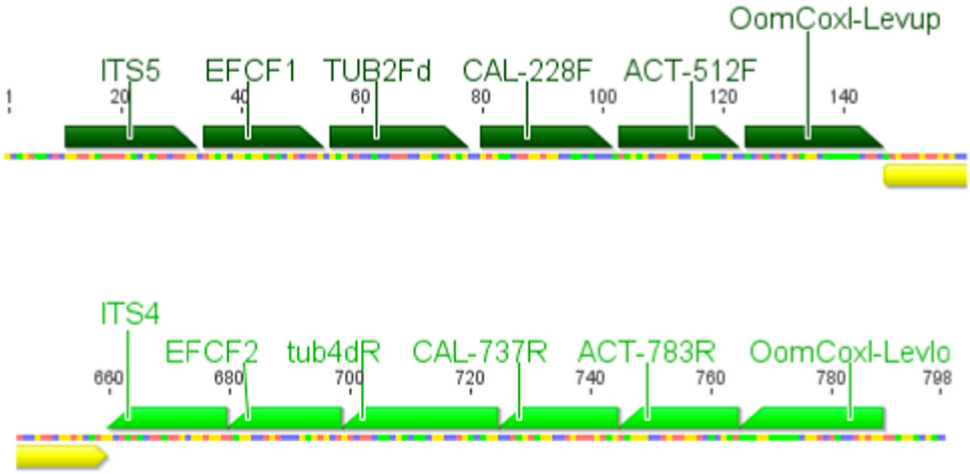


Note: The use of the MutS is no longer recommended for barcoding. The gBlock can still be used for the other tests.

gBlock name: EPPO_PAC_Bacteria_1
version: 1 [Colour figures can be viewed at wileyonlinelibrary.com]
length: 843
NCBI accession: KT429643
Sequence: GCTGATTCACAGAGTTTGATCCTGGCTCAGCAAGCAGGGCAAGAGCGAGCTGTAACAGCGCCTTGA
GCCGGTACACACCGTCGAGTTCGACTACGACGGACTAGTCCTGCCGGTGTTGATGCACGACTTATAGCAGCGC
TTTGAGTCGGTTAGAAAGAAGAGCCTTAGTGTGCTTTAATGTAGACCCACATAAGCTAGATATCGTAGATGGA
GCGCGAATTATACTAGGCGTAGGTTGAACGCTATTAGTCAACTAGAGCGAATGGCGAATAGAGAGAATTTGA
GCGGGAAAACTGTGAGTAGCCGCATAGAGCTAGCGAGTAGTGGATTACTCATTAGGAAATCGGACATACCTA
CTAGTTCATCGTAGAGTAGTGCCATGCACGGGCTTGCACAGAGAGATCGTGAAAGGAGGAACCATGATGCGC
ACTTATGTGAACACATATTAGTTGAATATCATGAATGGAAAGAGAGCTCTATTGAGCCTGAGTCGAGAGGTA
CTGAAGTACGCGTGCAAACGGAGAGTGACGTGAGTTCGAAAGAGAGAATTGCGAATGACCTCACCGAGCAT
CCGAATGATGGATAACCCACTGAGAGATAGGGCATACATATTGATTTATTGTGGAATGATGTCACGCGAGAGC
ATGTACCGAACGGAGCTAGCTCCTACGGGAGGCAGCAGTCAGCAGCCGCGGTAATACTGCGGCTGGATCACC

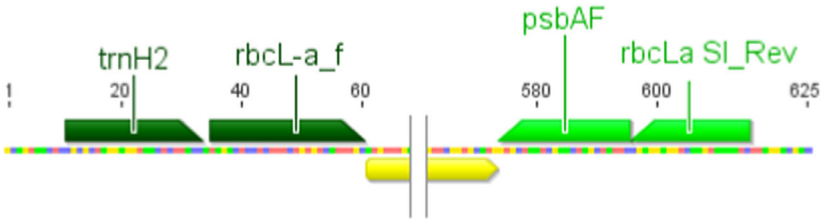
TCCTTGACCAGATCTTCAGCACCTTGATGTTCGGGCCGGTGATCAGCAAGTTCGGCAACACCGAGGGAAAGCC
TGTTGACCGATCACCGCTCGAGCGCGGCTCGAATCGCTGTTCACAATGGTCGTC

3. Fungal and Oomycete tests



gBlock name: EPPO_PAC_Fungi_1
version: 1 [Colour figure can be viewed at wileyonlinelibrary.com]
Length: 798
NCBI accession: KT429642
Sequence: GCTGATTCACGGAAGTAAAAGTCGTAACAAGGCAGTGCGGTGGTATCGACAAGCGTGCACCTCCAA
ACCGGTCAGTGCCGAGTTCAAGGAGGCCTTCTCCCTATGTGCAAGGCCGGTTTCGCCTCATCACGATGGCTTTT
TTCAACTAGAAAGAAGAGCCTTAGTGTGCTTTAATGTAGACCCACATAAGCTAGATATCGTAGATGGAGCGCG
AATTATACTAGGCGTAGGTTGAACGCTATTAGTCAACTAGAGCGAATGGCGAATAGAGAGAATTTGAGCGGG
AAAACTGTGAGTAGCCGCATAGAGCTAGCGAGTAGTGGATTACTCATTAGGAAATCGGACATACCTACTAGT
TCATCGTAGAGTAGTGCCATGCACGGGCTTGCACAGAGAGATCGTGAAAGGAGGAACCATGATGCGCACTTA
TGTGAACACATATTAGTTGAATATCATGAATGGAAAGAGAGCTCTATTGAGCCTGAGTCGAGAGGTACTGAA
GTACGCGTGCAAACGGAGAGTGACGTGAGTTCGAAAGAGAGAATTGCGAATGACCTCACCGAGCATCCGAAT
GATGGATAACCCACTGAGAGATAGGGCATACATATTGATTTATTGTGGAATGATGTCACGCGAGAGCATGTAC
CGAACGGAGCTAGGCATATCAATAAGCGGAGGAATGGCCAGACCCGTGAGCAGACAACTTCGTCTTCGGCCA
GTCTGGCCATGATGGCCAGAAAGATGATGGGCCAGAAGGACTCGTATTTGGTTTTTCGGACATCCAGAGGAAT
GGTCGTC

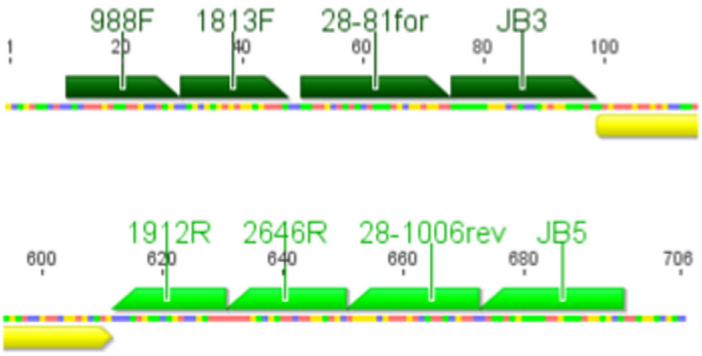
4. Invasive plant species tests



gBlock name: EPPO_PAC_Invasive_Plants_1
version: 1 [Colour figures can be viewed at wileyonlinelibrary.com]
length: 625
NCBI accession: KT429639
Sequence: GCTGATTCACCGCGCATGGTGGATTCACAATCCTATGTCACCACAAACAGAGACTAAAGCTAGAAA
GAAGAGCCTTAGTGTGCTTTAATGTAGACCCACATAAGCTAGATATCGTAGATGGAGCGCGAATTATACTAG

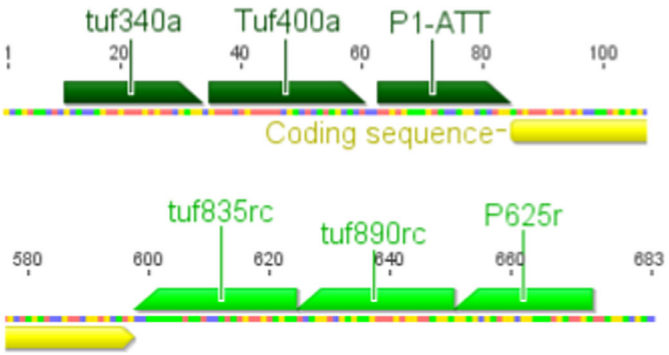
GCGTAGGTTGAACGCTATTAGTCAACTAGAGCGAATGGCGAATAGAGAGAATTTGAGCGGGAAAACTGTGAG
TAGCCGCATAGAGCTAGCGAGTAGTGGATTACTCATTAGGAAATCGGACATACCTACTAGTTCATCGTAGAGT
AGTGCCATGCACGGGCTTGCACAGAGAGATCGTGAAAGGAGGAACCATGATGCGCACTTATGTGAACACATA
TTAGTTGAATATCATGAATGGAAAGAGAGCTCTATTGAGCCTGAGTCGAGAGGTACTGAAGTACGCGTGCAA
ACGGAGAGTGACGTGAGTTCGAAAGAGAGAATTGCGAATGACCTCACCGAGCATCCGAATGATGGATAACCC
ACTGAGAGATAGGGCATACATATTGATTTATTGTGGAATGATGTCACGCGAGAGCATGTACCGAACGGAGCTA
GGAGCATTACGTTCATGCATAACCGTGGTGGACTTGATTTTACAATGGTCGTC

5. Nematological tests



gBlock name: EPPO_PAC_Nematodes_1
version: 1 [Colour figures can be viewed at wileyonlinelibrary.com]
Length: 706
NCBI accession: KT429641
Sequence: GCTGATTCACCTCAAAGATTAAGCCATGCCTGCGTGAGAGGTGAAATCCTTAAGCATATCATTT
AGCGGAGGAATTTTTTGGGCATCCTGAGGTTTATTAGAAAGAAGAGCCTTAGTGTGCTTTAATGTAGACCCAC
ATAAGCTAGATATCGTAGATGGAGCGCGAATTATACTAGGCGTAGGTTGAACGCTATTAGTCAACTAGAGCGA
ATGGCGAATAGAGAGAATTTGAGCGGGAAAACTGTGAGTAGCCGCATAGAGCTAGCGAGTAGTGGATTACTC
ATTAGGAAATCGGACATACCTACTAGTTCATCGTAGAGTAGTGCCATGCACGGGCTTGCACAGAGAGATCGTG
AAAGGAGGAACCATGATGCGCACTTATGTGAACACATATTAGTTGAATATCATGAATGGAAAGAGAGCTCTA
TTGAGCCTGAGTCGAGAGGTACTGAAGTACGCGTGCAAACGGAGAGTGACGTGAGTTCGAAAGAGAGAATT
GCGAATGACCTCACCGAGCATCCGAATGATGGATAACCCACTGAGAGATAGGGCATACATATTGATTTATTG
TGGAATGATGTCACGCGAGAGCATGTACCGAACGGAGCTAGCCCTAGTTCTGACCGTAAAAAAAGTCGTAAC
AAGGTAGCAGGGGCGAAAGACTAATCGAACCATTATGTTTTAAGTTTAGGTGCTAATGGTCGTC

6. Phytoplasma tests



gBlock name: EPPO_PAC_Phytoplasmas_1
version: 1 [Colour figures can be viewed at wileyonlinelibrary.com]

Length: 683
NCBI accession: KT429640
Sequence: GCTGATTCACGCTCCTGAAGAAAGAGAACGTGGCGAAACAGAAAAACGTCACTATGCTCACCAAGA
GTTTGATCCTGGCTCAGGTAGAAAGAAGAGCCTTAGTGTGCTTTAATGTAGACCCACATAAGCTAGATATCGT
AGATGGAGCGCGAATTATACTAGGCGTAGGTTGAACGCTATTAGTCAACTAGAGCGAATGGCGAATAGAGAG
AATTTGAGCGGGAAAACTGTGAGTAGCCGCATAGAGCTAGCGAGTAGTGGATTACTCATTAGGAAATCGGAC
ATACCTACTAGTTCATCGTAGAGTAGTGCCATGCACGGGCTTGCACAGAGAGATCGTGAAAGGAGGAACCAT
GATGCGCACTTATGTGAACACATATTAGTTGAATATCATGAATGGAAAGAGAGCTCTATTGAGCCTGAGTCGA
GAGGTACTGAAGTACGCGTGCAAACGGAGAGTGACGTGAGTTCGAAAGAGAGAATTGCGAATGACCTCACCG
AGCATCCGAATGATGGATAACCCACTGAGAGATAGGGCATACATATTGATTTATTGTGGAATGATGTCACGCG
AGAGCATGTACCGAACGGAGCTAGCCTTTTTTATTACCTATAGAAGATGTTACTGGACGTGTTGAAAGAGGAA
TGGTGGTGCGTAGGCGGTTTAGTAAGTAATGGTCGTC