EUPHRESCO DNA Barcoding

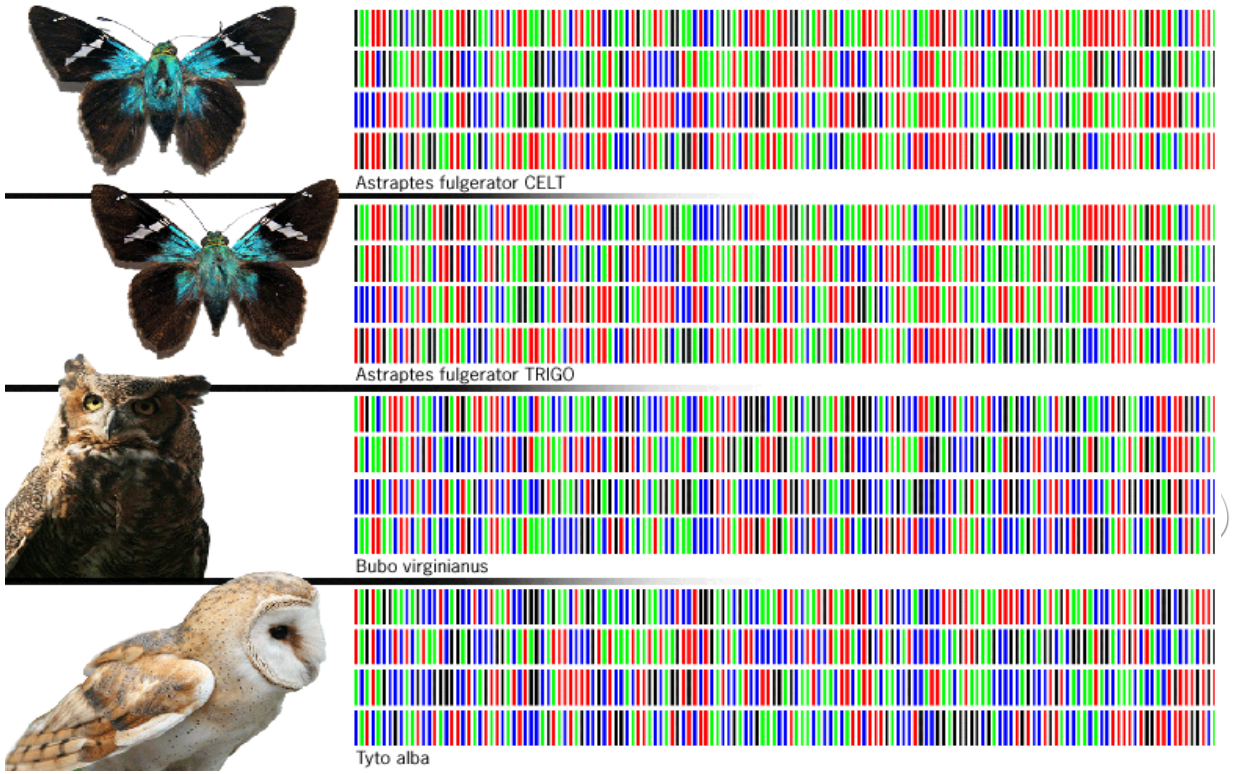
Optimising and validating DNA barcoding protocols for plant pests

BTLH van de Vossenberg[1], M. Westenberg[1], M. Botermans[1], J. Hodgetts[2], M. Maes[3]

1. NPPO-NL (NL), 2. FERA Ltd. (UK), 3. ILVO (BE)

27 July 2016, EEC-EPPO Workshop on Euphresco

DNA barcoding



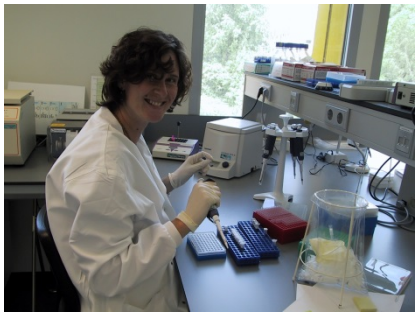
- Short DNA sequence data of a standardised genetic marker to aid species identification
- Mitochondrial partial COI gene (500 bp) first DNA Barcode (Hebert *et al*. 2003)

DNA barcoding in plant pest diagnostics

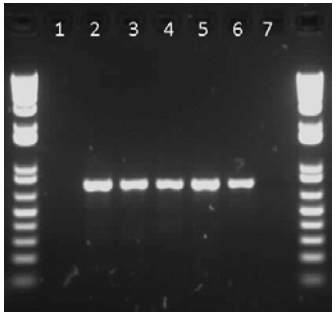
Inspection

Suspected pest species

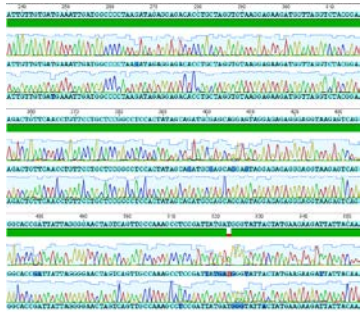
Morphological (putative) identification



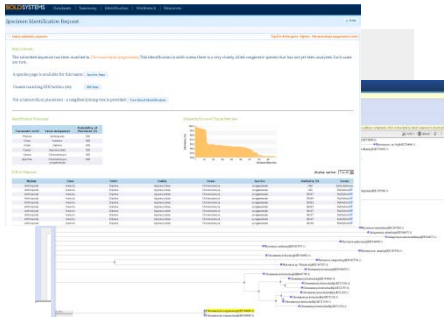
DNA extraction



PCR amplification



Sanger Sequencing

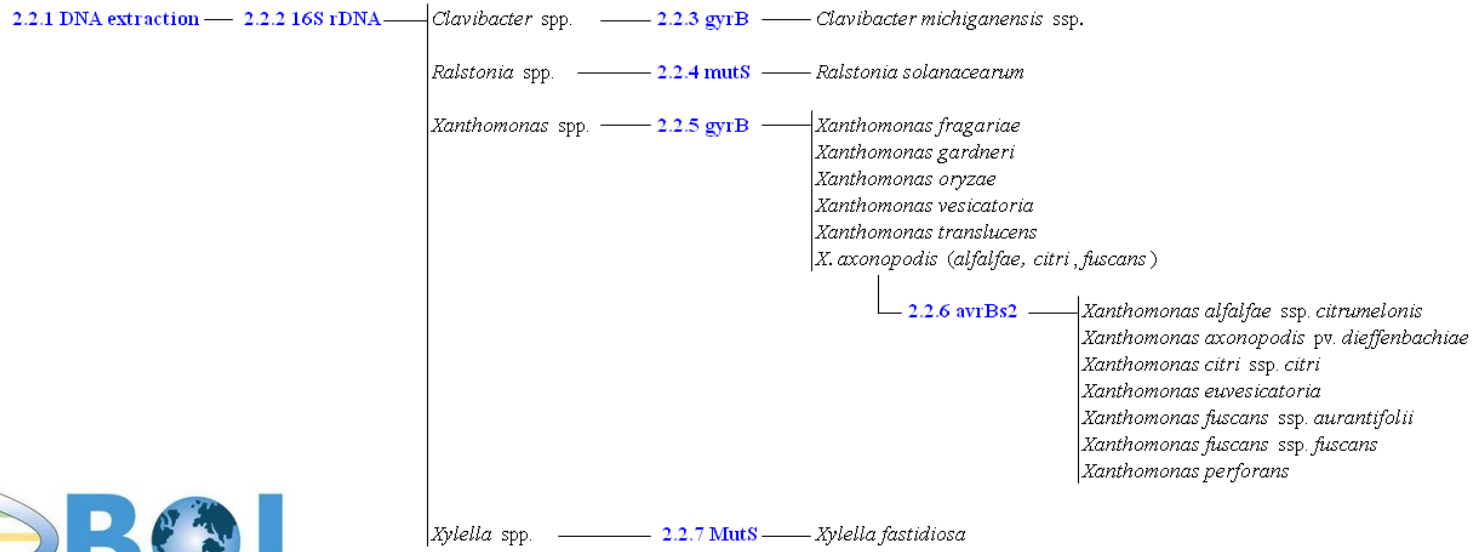


Data-analysis

Final diagnosis

Development of plant pest DNA barcoding

- A single genetic marker is often not sufficient
- Barcoding protocols for selected regulated organisms developed under the EU QBOL project (2009-2011)
  - Arthropods, Bacteria, Fungi, Nematodes, Phytoplasmas



QBOL international test performance study

- Validation of QBOL protocols in an international TPS
  - Protocols fit for purpose

- Needs for improvement were identified during QBOL TPS (van de Vossenberg *et al*., 2013)

  - User-friendliness of protocols
  - Guidance on data-analysis
  - Improved harmonisation between organism groups

  - Test performance study set-up

  - Proficiency participants

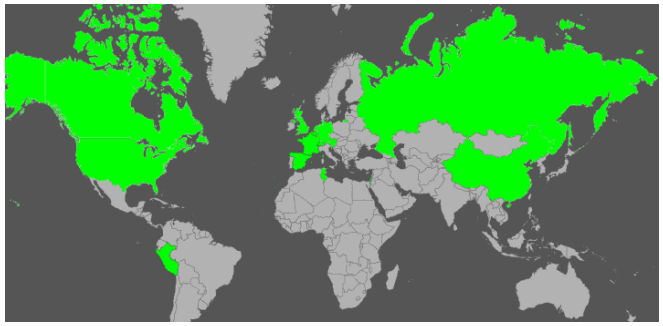
EUPHRESCO DNA barcoding project

- **WP 1: coordination**
  - NPPO-NL

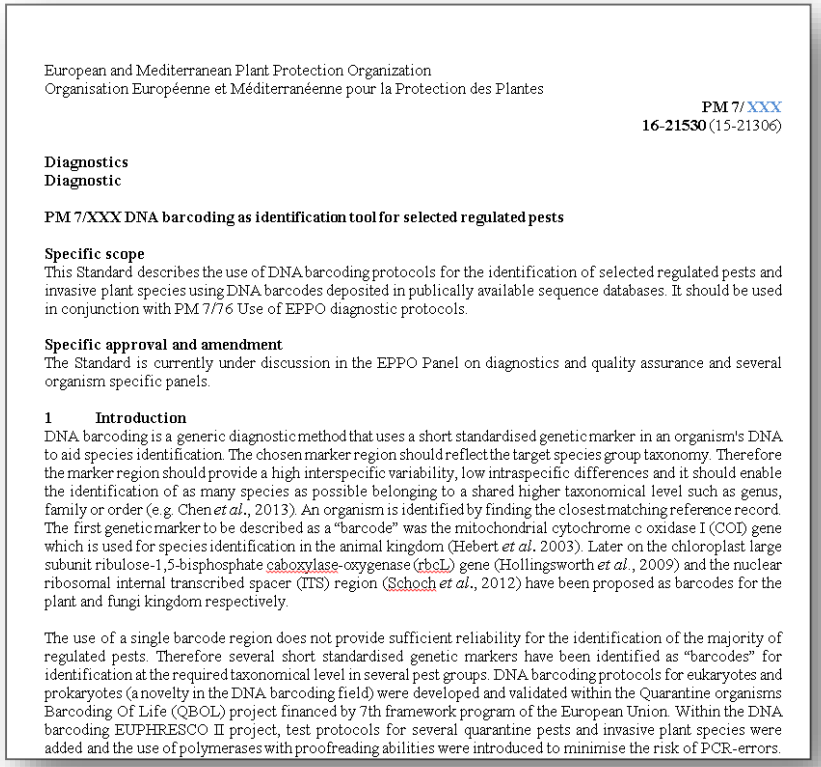
- **WP 2: update protocols**
  - Update test protocols (chemistry used, generic sequencing primers)
  - Include new tests for bacteria, fungi and invasive plant species
  - Produce guidelines on data-analysis

- **WP 3: validation**
  - Validate protocols in an international TPS
  - 23 participants – 15 countries worldwide

Euphresco
DNA Barcoding



Work package 2: update protocols



European and Mediterranean Plant Protection Organization
Organisation Européenne et Méditerranéenne pour la Protection des Plantes

PM 7/XXX
16-21530 (15-21306)

Diagnostics
Diagnostic

PM 7/XXX DNA barcoding as identification tool for selected regulated pests

Specific scope
This Standard describes the use of DNA barcoding protocols for the identification of selected regulated pests and invasive plant species using DNA barcodes deposited in publically available sequence databases. It should be used in conjunction with PM 7/76 Use of EPPO diagnostic protocols.

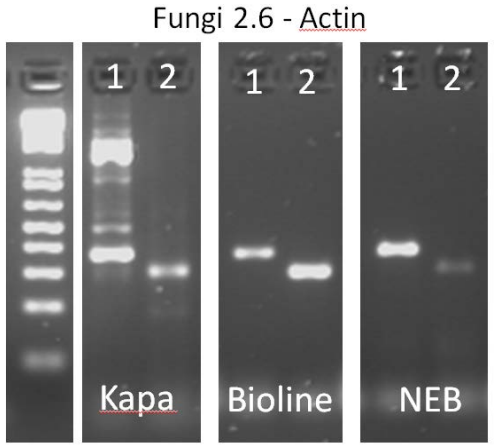
Specific approval and amendment
The Standard is currently under discussion in the EPPO Panel on diagnostics and quality assurance and several organism specific panels.

1 Introduction
DNA barcoding is a generic diagnostic method that uses a short standardised genetic marker in an organism's DNA to aid species identification. The chosen marker region should reflect the target species group taxonomy. Therefore the marker region should provide a high interspecific variability, low intraspecific differences and it 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 on the chloroplast large subunit ribulose-1,5-bisphosphate caboxylase-oxygenase (rbcL) gene (Hollingsworth *et al.*, 2009) and the nuclear ribosomal internal transcribed spacer (ITS) region (Schoch *et al.*, 2012) have been proposed as barcodes for the plant and fungi kingdom respectively.

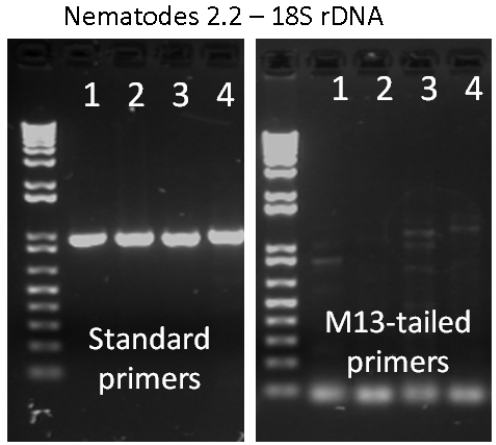
The use of a single barcode region does not provide sufficient reliability for the identification of the majority of regulated pests. Therefore several short standardised genetic markers have been identified as "barcodes" for identification at the required taxonomical level in several pest groups. DNA 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 7th framework program 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 were introduced to minimise the risk of PCR-errors.

WP2 – Improving user-friendliness

- Use as few polymerases as possible in the overall standard
- Selection of 3 widely used proofreading polymerases
- All groups, except nematodes, could be optimised for a single polymerase.

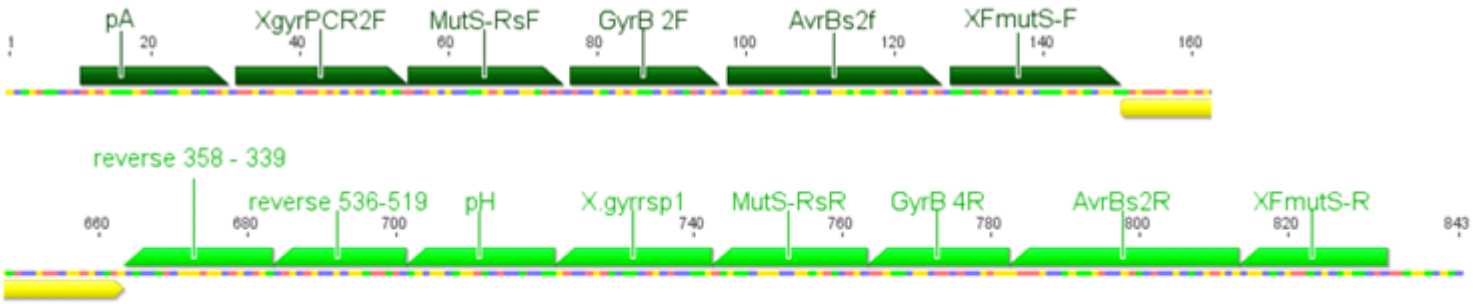


- Tailing of amplification primers to decrease the number of sequencing primers
- Generic M13 primers: M13rev-29 & M13uni-21
- Tailing of amplification primers successful in 9 of 22 tests (41%)



WP2 – Process control

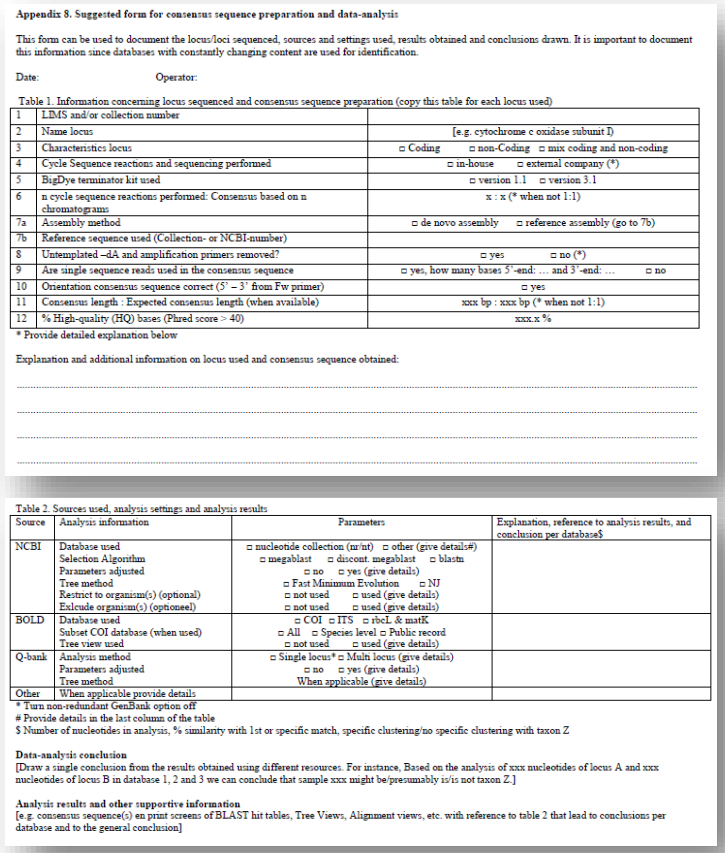
- Introduction of one process control per organism group
- Synthetic DNA contstruct
- Assessment of amplification success, Sanger sequencing, data-analysis, and proficiency of technicians



gBlock name: EPPO_PAC_Bacteria_1
version: 1
length : 843
NCBI accession: KT429643

WP2 – Standardised reporting form

- Introduction of a standardised reporting form of DNA Barcoding results
- Information on consensus sequence preparation and data-analysis
- Online databases change from day to day which can influence the repeatability of the data-analysis



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.

Date: Operator:

Table 1. 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 locus	◻ Coding ◻ non-Coding ◻ mix coding and non-coding
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 2. 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)	
	Exlcude organism(s) (optioneel)	◻ 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)	
Q-bank	Analysis method	◻ Single locus* ◻ Multi locus (give details)	
	Parameters adjusted	◻ no ◻ yes (give details)	
	Tree method	When applicable (give details)	
Other	When applicable provide details		

* Turn non-redundant 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) en print screens of BLAST hit tables, Tree Views, Alignment views, etc. with reference to table 2 that lead to conclusions per database and to the general conclusion]

Work Package 3 – Test performance study

WP3 – TPS set-up

- Call for participants in EPPO, Euphresco and IPPC networks
- Participation in at least 2 organism groups
- > 2 years experience with PCR, sequencing and (on-line) data-analysis
- All items, except polymerases, provided to TPS partners
- TPS partners had to:
  - Select the appropriate protocols
  - Amplify, sequence and analyse relevant loci
  - Report conclusion per sample, and provide (raw) sequence data

WP3 – Sample sets

Appendix	Sample	Scientific Name	Test needed for identification					
			2.2	2.3	2.4	2.5	2.6	2.7
1. Arthropods	1	*Vespa crabro*	x					
	2	*Bemisia tabaci*	x					
	3	*Liriomyza huidobrensis*	x					
	4	*Spodoptera eridania*	x					
	5	*Anoplophora glabripennis*	x					
2. Bacteria	1	*Xanthomonas axonopodis* pv *dieffenbachiae*	x			x	x	
	2	*Clavibacter michiganensis* subsp *michiganensis*	x	x				
	3	*Ralstonia solanacearum*	x		x			
	4	*Xylella fastidiosa*	x					x
	5	*Xanthomonas axonopodis* pv *begoniae*	x			x	x	
3. Fungi	1	*Phytophthora ramorum*	x					x
	2	*Lecanosticta acicola*	x		x			
	3	*Stagonosporopsis chrysanthemi*	x				x	
	4	*Verticillium dahliae*	x			x		
	5	*Ceratocystis fimbriata* f. sp. *platani*	x	x				
4. Invasive Plants	1	*Ludwigia peploides*	x	x				
	2	*Ludwigia grandiflora*	x	x				
	3	*Hydrocotyle ranunculoides*	x	x				
	4	*Myriophyllum heterophyllum*	x	x				
	5	*Hydrocotyle vulgaris*	x	x				
5. Nematodes	1	*Meloidogyne chitwoodi*	x		x			
	2	*Bursaphelenchus xylophilus*	x	x				
	3	*Aphelenchoides besseyi*	x	x	x			
	4	*Ditylenchus dipsaci*	x	x				
	5	*Aphelenchoides fragariae*	x	x	x			
6. Phytoplasmas	1	*Candidatus* Phytoplasma solani	x	x				
	2	*Candidatus* Phytoplasma solani	x	x				
	3	*Candidatus* Phytoplasma mali	x	x				
	4	*Candidatus* Phytoplasma pyri	x	x				
	5	*Candidatus* Phytoplasma prunorum	x	x				

WP3 – Analysis of TPS data

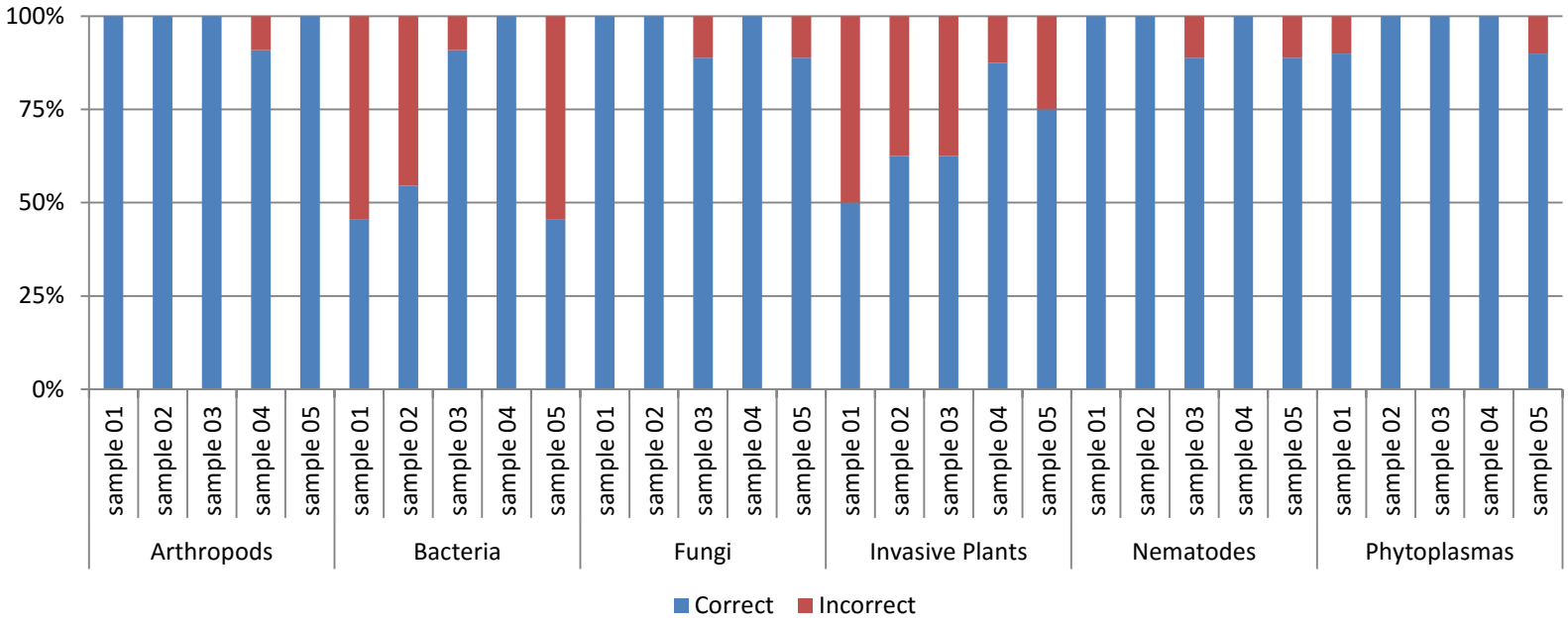
- Diagnostic sensitivity was calculated with TPS participant data
  - Indication of protocol success rate

- Robustness criteria were checked if they influence diagnostic sensitivity:
  - Use of correct tests
  - deviations from test protocol
  - trimming of consensus sequence
  - Sequence quality
  - assembly methods
  - databases used

- Re-analysis of data provided by TPS participants when robustness criteria could not explain false negative results

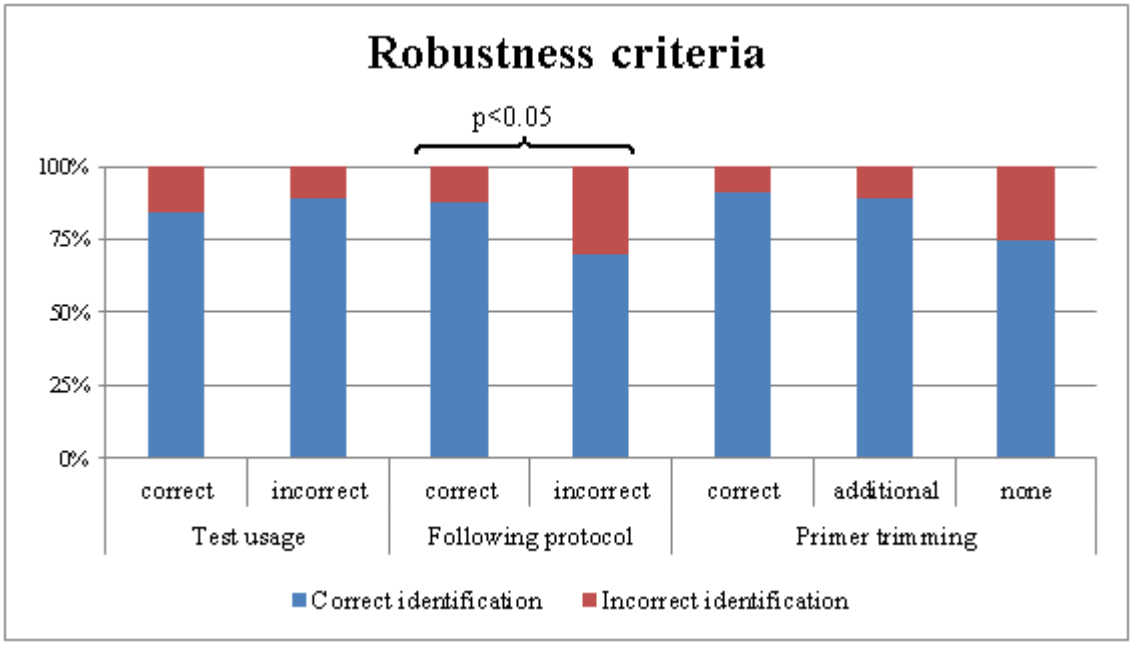
WP3 – Diagnostic sensitivity from TPS data

- Diagnostic sensitivity values obtained with identities provided by partners
  - Individual samples: 45%-100%
  - Organism groups: 67%-98%
- Bacteria and Invasive plants scored lower compared to other organism groups



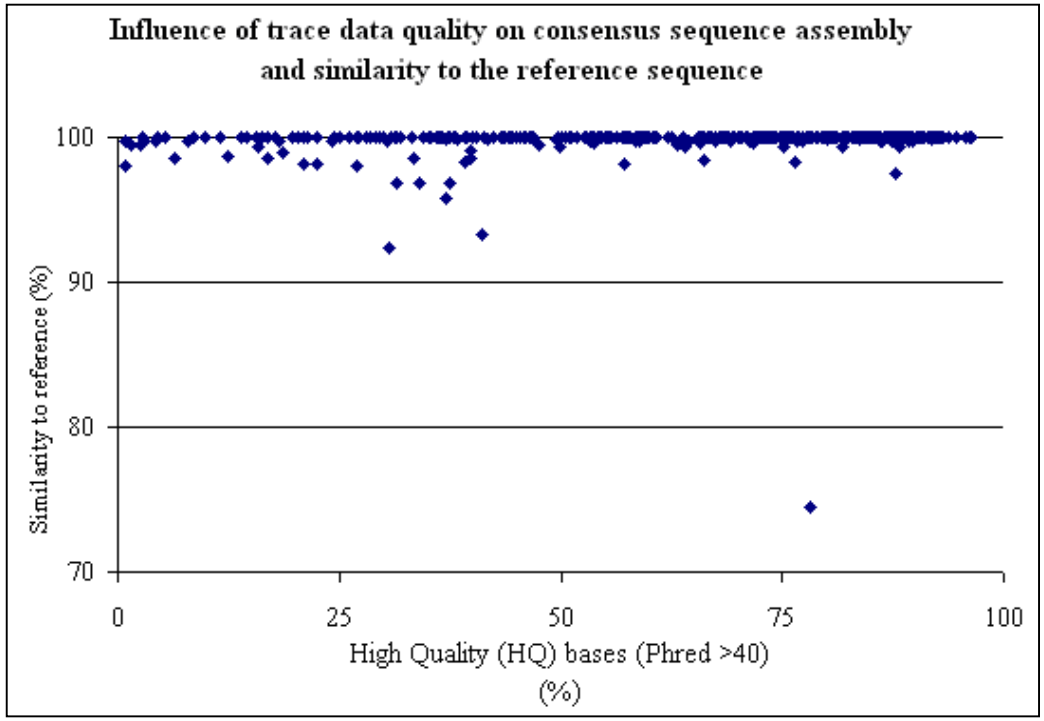
WP3 – Robustness criteria

- Incorrect test usage or primer trimming did not result in significant lower percentages of correctly identified samples
- Not following protocol did have a significant negative effect
  - No correlation with organism groups scoring low: participant effect



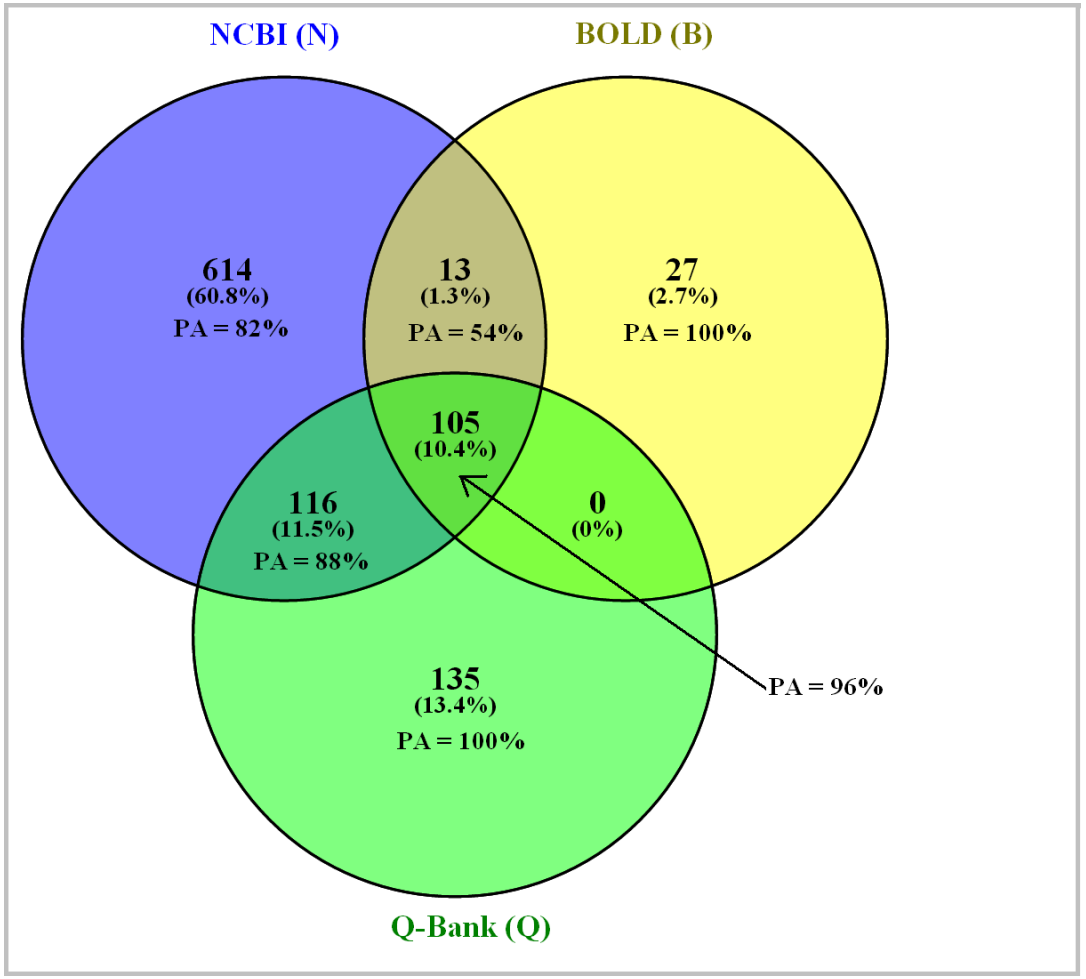
WP3 – Robustness criteria

- Sequence data quality and assembly methods do not have an effect on the percentage similarity to the consensus sequence.



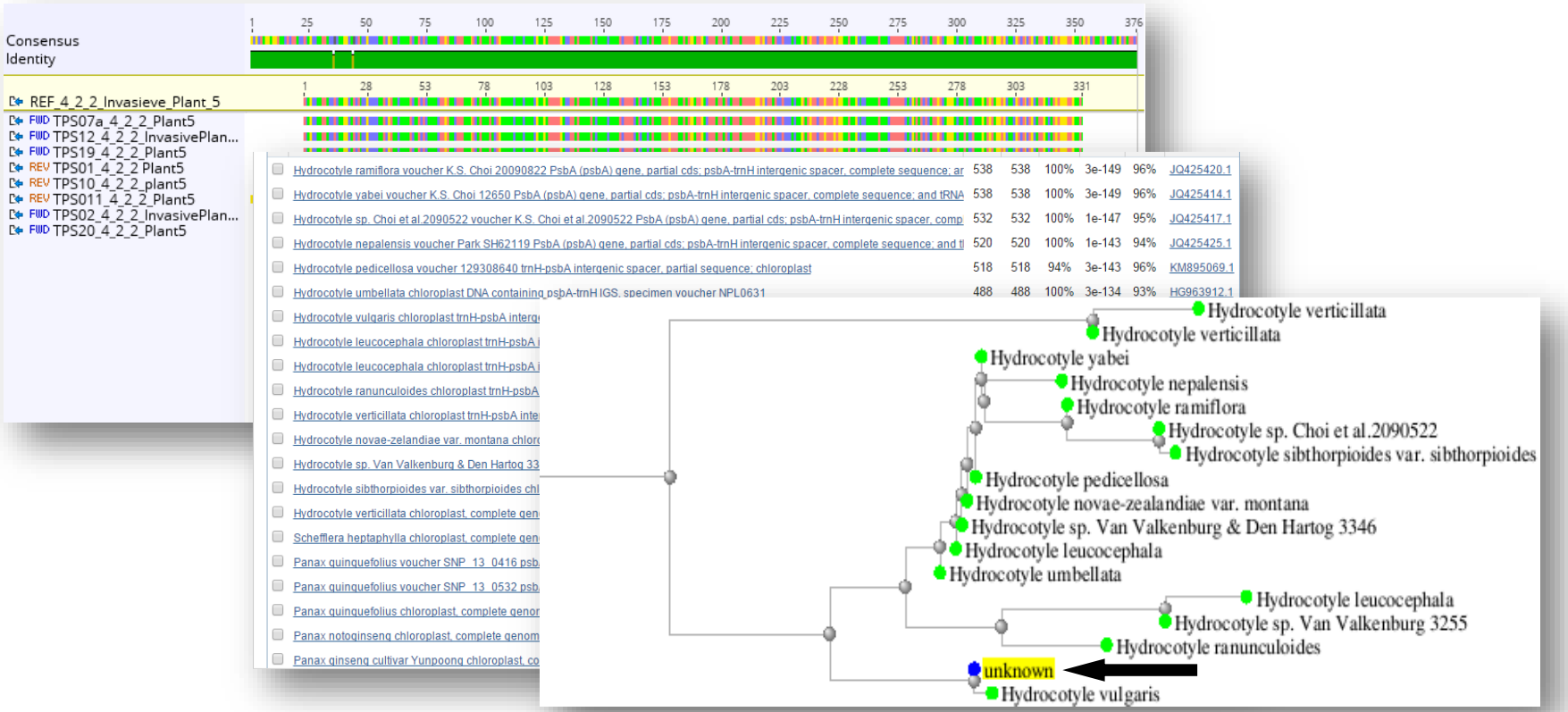
WP3 – Database use

- Top 3 used database combinations
  - NCBI (61%)
  - Q-Bank (13%)
  - NCBI + Q-Bank (12%)
- Optimal database (combination) is organism group and sample dependent
- Exclusive use of a single database is not considered best practice



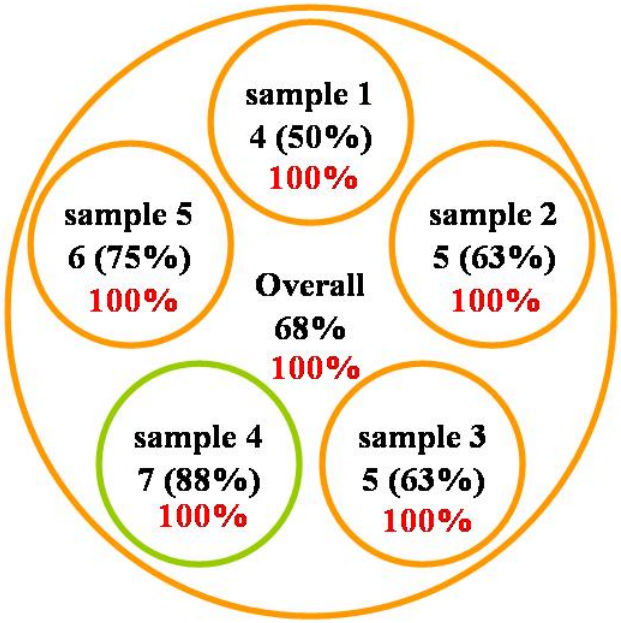
WP3 – Re-analysis of TPS data

- Robustness criteria could not explain false negative results
- Re-analysis of consensus sequences provided by TPS participants

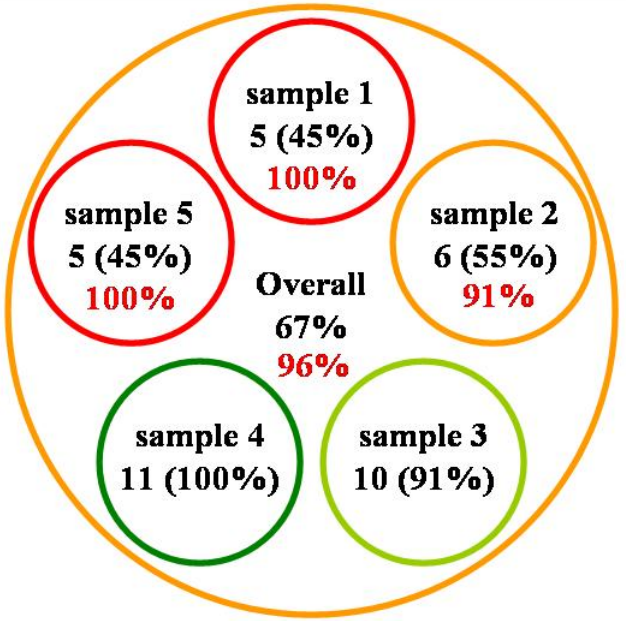


WP3 – Re-analysis of TPS data

- Re-analysis of TPS data generated by participants shows that an overall diagnostic sensitivity of 99% can be reached.



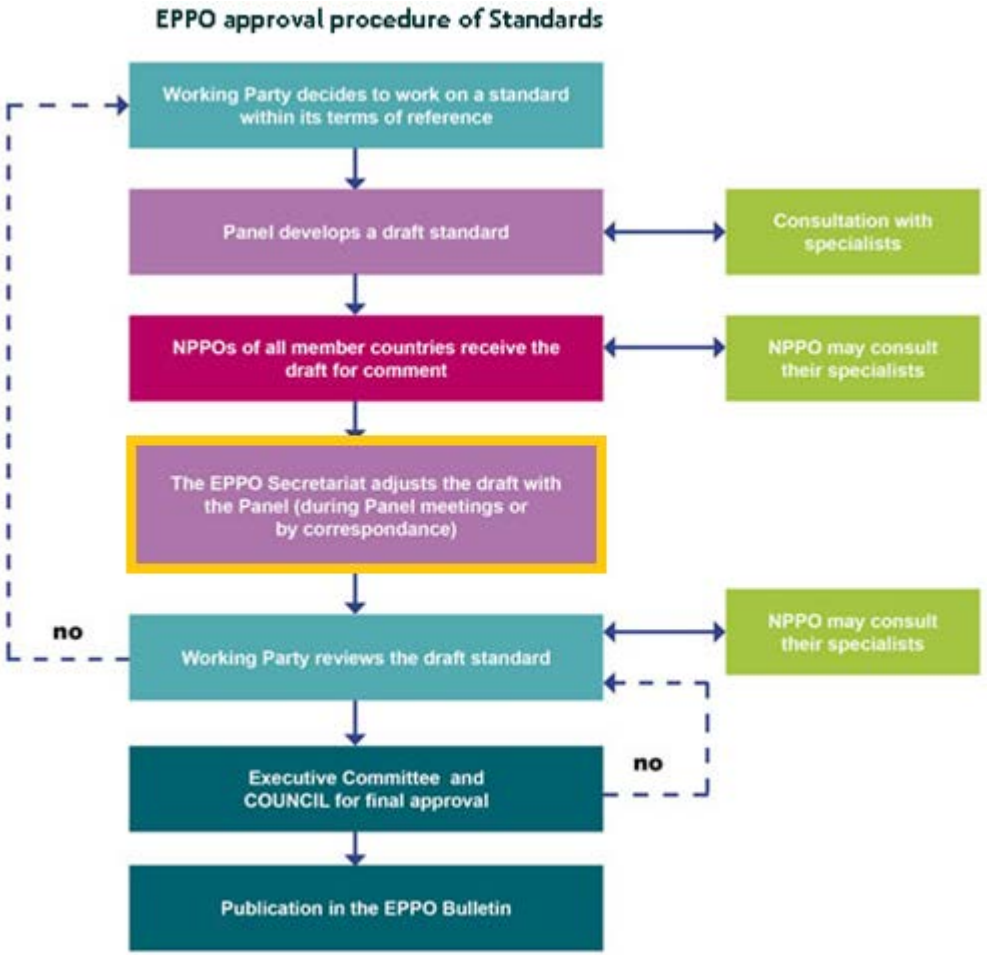
Invasive Plants (8)



Bacteria (11)

EPPO DNA barcoding standard

- Finalisation of standard using TPS validation results
- Country consultation round finalised
- Presenting standard to council Sept 2016 for final approval



Conclusions

Work package 2 – update protocols

- New tests for bacteria, fungi and invasive plants were added
- Improved user-friendliness
- End-users were actively involved in the update of the standard

Work package 3 – Test performance study

- 23 participants representing 15 countries worldwide
- Data generated by TPS partners show that 99% diagnostic sensitivity can be reached
- Tests are robust
- Data interpretation proved to be challenging for some participants

Acknowledgements

- Work package coordinators
- Standard drafting team
- TPS participants
- EPPO, IPPC, EUPHRESCO
- NPPO-NL researchers and technicians