

# The organization of proficiency testing in plant pathology (qualitative methods of analysis) according to the ISO/IEC 17043: example of the French national reference laboratory

A. Chabirand · G. Anthoine ·  
O. Pierson · B. Hostachy

Received: 28 October 2013 / Accepted: 9 January 2014 / Published online: 2 March 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** The use of methods of analysis capable of producing reliable analytical results is a prerequisite to the effective control of quarantine plant pathogens. Proficiency testing is considered to be one of the most reliable ways to verify and coordinate analytical results. As a French national reference laboratory in plant pathology, the Anses Plant Health Laboratory organizes proficiency tests in order to ensure that officially approved laboratories (certified by government services) are capable of producing reliable analytical results for the detection of plant pathogens. Proficiency tests in plant pathology have a number of notable features including the processing of qualitative results. This paper presents the experience of the Anses Plant Health Laboratory's Unit for Tropical Pests and Diseases (LSV-RAPT) as an organizer of proficiency tests in plant pathology. The LSV-RAPT has gained recognition for the methodology it has developed in the form of accreditation as a proficiency testing provider according to the ISO/IEC 17043. The methodology can be applied to

many other disciplines that use qualitative detection methods.

**Keywords** Interlaboratory comparisons · Proficiency testing · ISO/IEC 17043 · Plant pathology · Descriptive statistics · Qualitative method

## Introduction

Interlaboratory comparisons are increasingly used internationally for a number of purposes. Proficiency testing (PT) is a method of checking laboratory testing performance by means of an interlaboratory comparison. The need for ongoing confidence in laboratory performance is essential for laboratories and their customers, as well as for other interested parties. These include the regulatory authorities that issue an approval to certain laboratories to perform official analyses (the officially approved laboratories), and the accreditation bodies, which take account of the laboratories' participation and performance in proficiency testing.

The reference laboratories have to supervise and coordinate the activity conducted by officially approved laboratories (i.e. laboratories certified by government services) in order to ensure that the latter are capable of producing reliable analytical results.

The Anses Plant Health Laboratory's Unit for Tropical Pests and Diseases (LSV-RAPT) is the French national reference laboratory for the detection of viruses and bacteria on tropical plants. In this context, LSV-RAPT organizes interlaboratory proficiency tests to detect bacteria and viruses on tropical plants. French officially approved laboratories are required to take part in these tests.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00769-014-1034-y) contains supplementary material, which is available to authorized users.

A. Chabirand (✉) · B. Hostachy  
Anses Plant Health Laboratory, 7 Chemin de l'IRAT,  
97410 Saint Pierre, Reunion Island, France  
e-mail: aude.chabirand@anses.fr

G. Anthoine  
Anses Plant Health Laboratory, 7 Rue Dixméras, 49044 Angers,  
France

O. Pierson  
Anses, 27-31 Avenue du Général Leclerc,  
94701 Maisons-Alfort, France

The usual procedure for proficiency tests in plant pathology consists of sending identical sets of samples from an organizing laboratory (the organizer or provider) to a group of participating laboratories (the participants) for detection of one (or more) target plant pathogen(s). The samples are intended to simulate the routine samples usually analyzed. The participants are not informed of the expected results (assigned values) for the samples and are requested to perform the analyses (blind analyses) in the same manner as for routine samples. Results obtained from the samples are returned to the organizer for an evaluation of conformity with the expected results.

Good performance results provide independent and objective evidence that analytical quality assurance is effective [1] and encourage the maintenance of a high level of performance. Conversely, poor performance results can help to detect anomalies in the analytical process. Subsequently, a causal analysis is conducted, and an action plan to improve the performance is implemented.

Nonconforming results in proficiency testing can have important consequences for the laboratory, such as the review of its official approval or accreditation. Consequently, it is essential that the organizer can provide the participant with guarantees of its competence in organizing proficiency testing.

Therefore, LSV-RAPT decided to develop a quality management system following the requirements of ISO/IEC 17043 [2]. Proficiency tests in plant pathology have a number of notable features. This paper presents the specificities of these proficiency tests and the solutions adopted by LSV-RAPT (the organizer) to meet the ISO/IEC 17043 [2] requirements. In particular, key features of the management system are highlighted. Issues relating to the selection and validation of test samples and to the statistical data processing of participants' results are developed and illustrated with actual examples.

There are many guidelines (standards and publications) on the organization of proficiency tests, which concern quantitative data [3–6] or can be adapted for this purpose using guidelines for interlaboratory tests for the validation of quantitative analytical methods or the certification of reference materials [7, 8]. However, very little guidance is available for the organization of proficiency tests that concern qualitative data, such as in plant pathology. Recent publications concerning the processing of qualitative data, based on the probability of detection (POD) and the analysis of ordinal variation (ORDANOVA) [9–12], present promising statistical methods for application in proficiency testing. However, the amount of data needed can limit their use in plant pathology (low availability of reference materials, small number of participants). The methodology developed in this paper was adapted for proficiency tests. It is based on the guidelines for the interlaboratory tests for

the validation of qualitative analytical methods (mainly ISO 16140 [13], AOAC International guidelines for the validation of official methods of qualitative and quantitative food microbiological analysis [14] and a publication of McClure [15]). It consists of descriptive statistics: numbers of positive agreements, negative agreements, positive deviations and negative deviations are calculated from the qualitative results submitted by the laboratories and are used to calculate performance criteria.

The LSV-RAPT obtained accreditation by the Cofrac (the French accreditation body) as proficiency testing organizer in plant pathology in June 2013.

## Challenges for proficiency tests in plant pathology

### Qualitative data

All test results in laboratories performing analyses in plant pathology are given in qualitative terms (“positive”, “negative” or “indeterminate” when applicable). Even if some analytical methods produce intermediate quantitative data (absorbance values for enzyme linked immunosorbent assay (ELISA), crossing threshold (Ct) values for real-time polymerase chain reaction (PCR), number of cells for immunofluorescence, etc.), such data are generally compared to a threshold value to determine the qualitative final result.

For example, for each sample analyzed using ELISA technique (widely used in plant pathology), an absorbance value ( $A$ ) is produced: in the final step of the analysis (based on specific antibody–antigen reactions revealed by colorimetry), light absorbance of a colored solution is measured on a spectrophotometer.

Usually, two threshold values ( $S_1$  and  $S_2$ , respectively the negativity threshold and the positivity threshold) are determined (calculated from the absorbance values of negative control samples). Then, for an unknown sample:

- if  $A < S_1$ , the test result is defined as negative for the detection of the target pathogen;
- if  $S_1 \leq A < S_2$ , the test result is defined as indeterminate;
- if  $A \geq S_2$ , the test result is defined as positive.

The threshold values, as well as the intermediate quantitative data (absorbance values in the previous example), may vary widely between laboratories (depending on the equipment and consumables used). Consequently, the use of scores (widely used to assess laboratories' performance in quantitative proficiency testing) calculated from the intermediate quantitative data can lead to the establishment of incorrect performance results. In particular, the use of scores can lead to declare wrongly as nonconforming a

laboratory with conforming qualitative results, but for which the intermediate quantitative results (but also the threshold values) would be widely higher (or lower) than those of the other participating laboratories. And conversely, the use of scores can lead to declare wrongly as conforming a laboratory with nonconforming qualitative results but for which the intermediate quantitative results would be close to those of the other participating laboratories.

Quantitative data should not be used as such to assess participants' performance. The performance assessment must be based on the result as it is transmitted to the customers of the laboratory (i.e. the qualitative result). So, specific statistical procedures appropriate for qualitative variable must be applied to assess proficiency of laboratories performing analyses in plant pathology.

However, the participants can be asked to report their intermediate quantitative data to ensure that their interpretation is correct (conversion into a qualitative result). Poor interpretation can lead to a declaration of nonconformity with regard to the implementation of the proficiency test and also potentially in relation to several technical performance criteria if the resulting qualitative result is incorrect.

For the homogeneity and stability studies, intermediate quantitative data can be used, because produced in the same laboratory (usually the organizer).

#### Large number of test samples and low availability of reference materials

The number of test samples for a qualitative evaluation is generally higher than that required for a quantitative evaluation. This is due to the fact that less information is provided by a qualitative result compared to a quantitative result. The number of different test samples needed to evaluate the detection of one plant pathogen is often higher than 10. In addition, the availability of reference materials used to produce the positive test samples is often low. Consequently, the protocol generally used to validate the test samples (homogeneity and stability studies) can be difficult to combine with the practical constraints.

#### Small number of participants

The number of officially approved laboratories, which perform analyses in plant pathology, is often limited, compared to other disciplines. This is all the more true for analyses that concern tropical plant pathogens. Consequently, some data processing methods that are widely used in other disciplines (e.g. calculations of a z-score on the

basis of a consensus value and a standard deviation derived from the result reported by the participants) cannot be used for proficiency testing in plant pathology. And, as already mentioned, specific statistical procedures must be applied.

Generally, in order to increase the number of participants and, consequently, to improve the level of confidentiality, the proficiency tests, which are initially designed for officially approved laboratories, are also opened to laboratories that are partners of the organizer (e.g. research laboratories or other laboratories that perform analyses in plant pathology). Assessment procedures are identical for all the categories of participants, although the conditions for participation may differ on several points (invoicing, use of alternative methods, etc.).

### Management system

#### General management system and technical skills

The organization of proficiency tests is a highly technical activity [1] which must be supported by a solid quality management system and technical skills. The ISO/IEC 17025 [16] accreditation of the organizer guarantees both.

Each PT organization managed as a project based on a PDCA model

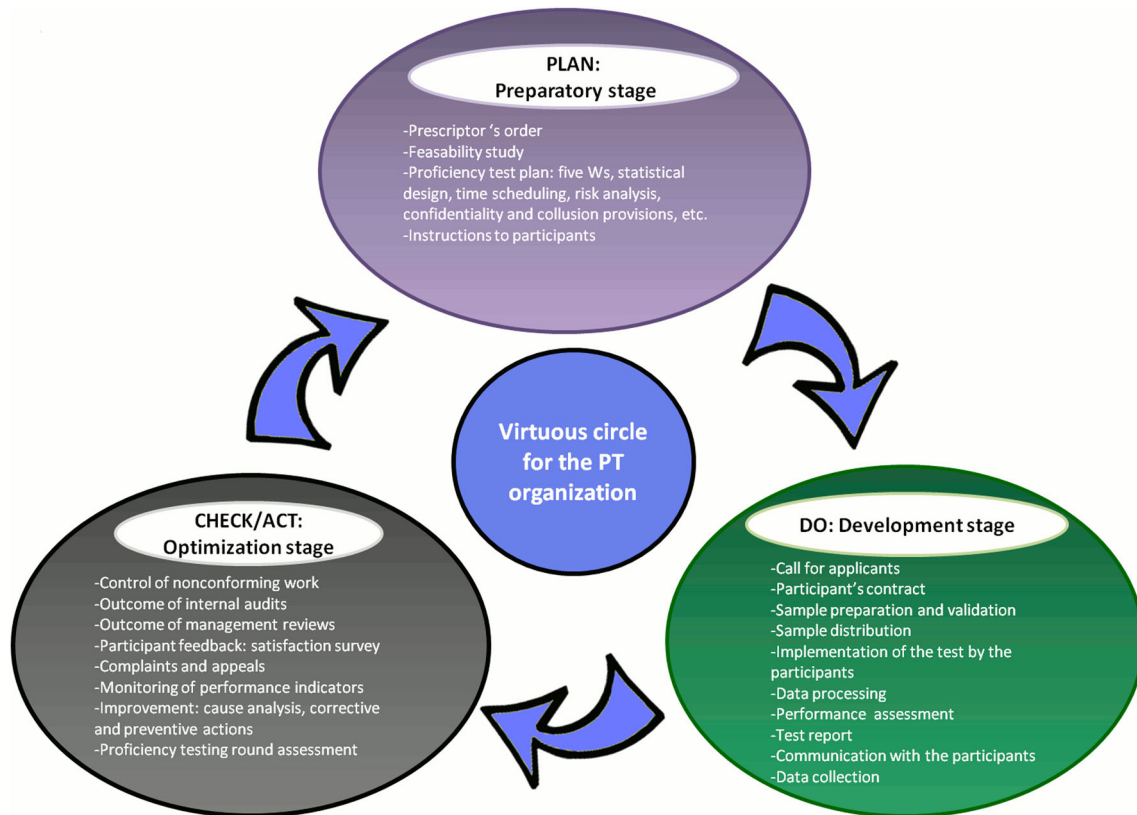
To meet the requirements specific to ISO/IEC 17043 standard [2], the LSV-RAPT (the organizer) manages the organization of each proficiency test as a project based on the PDCA (Plan Do Check Act) continuous improvement model, as shown in Fig. 1. The proficiency test plan which sets out the design of the proficiency test and provides a risk assessment connected to the PT organization, is a key document in this process.

When the PDCA approach is applied to each production cycle of a proficiency test, the production steps can be standardized and levers can be found so that there is continuous improvement from one cycle to the next (virtuous circle). This improves the general performance of the PT organization process.

#### Confidentiality

The identity of participants and all the information provided by a participant to the organizer remain confidential.

In addition to the confidentiality ensured by the organizer's ISO/IEC 17025 [16] quality management system, specific provisions are implemented. Throughout the pro-



**Fig. 1** Project management based on a PDCA model applied to the organization of proficiency tests

proficiency testing round, the participating laboratories are identified by a code, which consists of the letter L followed by a randomly assigned two-digit number (e.g. “L07”). Results obtained from the participants are presented anonymously in order to ensure confidentiality with regard to the other participants. Then each laboratory is informed of its own laboratory identification code for the exploitation of results.

#### Collusion and falsification

The proficiency test is designed to minimize the opportunity for collusion and falsification of results. Each test sample is identified according to the following code components: PT code—set code—sample code, e.g. 05XD-L01-05: sample No. 5 of the set L01 of the proficiency test 05XD. The coding of the sets is assigned randomly to each participant. The individual coding of the samples is randomly chosen and is different for each set (and consequently for each participant). Randomizations are performed using R statistical software (R Development Core Team, Vienna, Austria).

In addition, lures are included in each set of samples. The lures are samples similar to other samples, randomly chosen among a pool of extra (positive and negative) test

samples, but which are not taken into account for evaluating the laboratories. Lures are only used to vary the proportion of positive and negative results among the sets of samples.

Participants are also asked to submit any relevant documents as proofs of their results (analysis data sheet, photos of PCR gel for PCR analysis, etc.).

These precautions are intended to prevent the participants to compare their results and potentially to falsify them (codification and number of positive and negative results are different for each participant).

#### Performance assessment

##### Performance criteria

Laboratories which participated to proficiency tests organized by LSV-RAPT are evaluated on their capacity to produce accurate results for the detection of one (or more) plant pathogen(s). The accuracy of results is a global criterion which includes both trueness and precision:

- The trueness is evaluated through the capacity to obtain positive results from positive samples (sensitivity) and negative results from negative samples (specificity).

**Table 1** Definitions of the parameters of positive agreement (PA), negative agreement (NA), positive deviation (PD) and negative deviation (ND) (definitions adapted from ISO 16140 [12])

Laboratory result	Assigned value	
	Positive	Negative
Positive	PA = positive agreement	PD = positive deviation
Negative	ND = negative deviation	NA = negative agreement
Indeterminate	ND = negative deviation	PD = positive deviation

- The precision is evaluated through the capacity to obtain the same qualitative result from identical samples analyzed under conditions of repeatability. In other words, the intra-assay precision is evaluated and subsequently named “repeatability.”

The participating laboratories make the commitment to perform the analyses in their laboratory according to the organizer’s instructions and in their usual conditions of work. They must use the standardized method of analysis (usually the French official method of analysis) as a priority. They also have to inform the organizer of any modification or deviation from this standardized method of analysis. If the organizer can ensure that the conditions of participation are comparable to other participants, the participant is authorized to use an alternative method.

The performance assessment is conducted in the same manner whatever the method used by the participants.

The qualitative results submitted by the participants are interpreted for each laboratory by calculating the number of positive agreements (PA), negative agreements (NA), positive deviations (PD) and negative deviations (ND), according to Table 1.

These parameters are used to calculate the performance criteria summarized in Table 2.

### Required performance levels

The laboratories’ performance is evaluated in terms of the minimum levels required for the performance criteria. The minimum levels are defined independently of the proficiency test results. They are shown in Table 3.

### Conformity of the results

The analysis of the results of a participating laboratory leads to a declaration of conformity (or nonconformity):

- “Conforming results” indicate that the participant’s results correspond to the expected results. The indication “conforming results” can be associated with one or several “comment(s)”. In this case, the results correspond to the

**Table 2** Definitions and calculations of the performance criteria

Performance criteria	Definition	Calculation
Accuracy (AC)	Closeness of agreement between the laboratory result and the assigned value (definition adapted from ISO 5725-1 [8])  Comments: the accuracy is a global criterion which can be subdivided, to refine the analysis, into three criteria: sensitivity, specificity and repeatability. This subdivision helps to direct investigations in case of nonconformities	$AC = (N_{PA} + N_{NA})/N$
Sensitivity (SE)	Closeness of agreement between the laboratory result and the assigned value for samples for which the assigned value is positive (definition adapted from ISO 16140 [12])  Comments: as far as possible, the evaluation of sensitivity should include the evaluation of inclusivity (the laboratory must be able to detect a wide range of target strains; the definition has been adapted from ISO 16140) and detectability (the laboratory must be able to obtain a positive response from a low contaminated sample; the definition has been adapted from NF U 47-400 [15])	$SE = N_{PA}/N^+$  Comments: the result of the calculation $(1-SE)$ gives the rate of false negatives obtained by the laboratory
Specificity (SP)	Closeness of agreement between the laboratory result and the assigned value for samples for which the assigned value is negative (definition adapted from ISO 16140 [12])  Comments: as far as possible, the evaluation of specificity should include <ul style="list-style-type: none"> <li>• healthy samples, i.e. symptomless samples not contaminated by any particular nontarget organism. The healthy samples are included in order to check the absence of interference with the plant matrix</li> <li>• samples contaminated by nontarget organisms that can be found in routine samples and can interfere (nontarget plant pathogens which attack the same host plant or saprophytic organisms naturally present in the plant)</li> </ul>	$SP = N_{NA}/N^-$  Comments: the result of the calculation $(1-SP)$ gives the rate of false positives obtained by the laboratory



**Table 2** continued

Performance criteria	Definition	Calculation
Repeatability ( <i>DA</i> ) (or accordance)	Closeness of agreement between independent test results obtained under conditions of repeatability, i.e. conditions under which independent test results are obtained by the same method, on identical test samples in the same laboratory, by the same operator, using the same equipment, within a short period of time (ISO 5725-1 [8])	<i>DA</i> denotes the percentage chance of obtaining the same result (positive, negative or indeterminate) from two identical samples analyzed in the same laboratory Calculation adapted from ISO 16140 [12]

$N_{PA}$  = number of positive agreements;  $N_{NA}$  = number of negative agreements;  $N$  = total number of samples;  $N^+$  = number of samples for which the assigned value is positive;  $N^-$  = number of samples for which the assigned value is negative

expected results; however, the organizer makes a comment, which calls for an improvement from the participant.

- “Nonconforming results” indicate that the participant’s results do not correspond to the expected results. In this case, the participant is encouraged to perform a causal analysis to explain the nonconformities and to implement appropriate corrective and preventive actions. Participation in the next proficiency test round concerning the same pathogen and method is the best way to evaluate the effectiveness of the corrective and preventive actions.

The declaration of conformity or nonconformity is primarily based on the laboratories’ technical performance. However, other items related to the implementation of the proficiency test are also taken into consideration (respect of the deadline for performing the analyses, correct interpretation of the intermediate quantitative data (conversion into the correct qualitative result), etc.).

#### Minimum number of participants

The methodology used to evaluate the participants’ performance is not based on consensus values calculated from the participants’ results. Furthermore, statistical tests that require a minimum amount of data are not used. Consequently, there is no statistical requirement concerning the minimum number of participants.

### Sample selection, preparation and validation

Main steps of the production and validation of proficiency test samples are shown in Fig. 2.

**Table 3** Performance levels required for validating the participants’ performance as defined for LSV-RAPT proficiency tests

Performance criteria	Required level (%)	Comments
Sensitivity ( <i>SE</i> )	100	For positive samples, the requirement is that 100 % of the positive samples contaminated above the required detection level give a positive result
Specificity ( <i>SP</i> )	100	The requirement is that 100 % of negative samples give negative results
Repeatability ( <i>DA</i> )	100	Qualitative results obtained from identical positive samples contaminated above the required detection level must be identical. Qualitative results obtained from identical negative samples must be identical. This corresponds to a repeatability of 100 %
Accuracy ( <i>AC</i> )	100	Accuracy summarizes the three other performance criteria, and therefore, failure to achieve one of the three performance criteria will lead to failure in accuracy

#### Sample selection

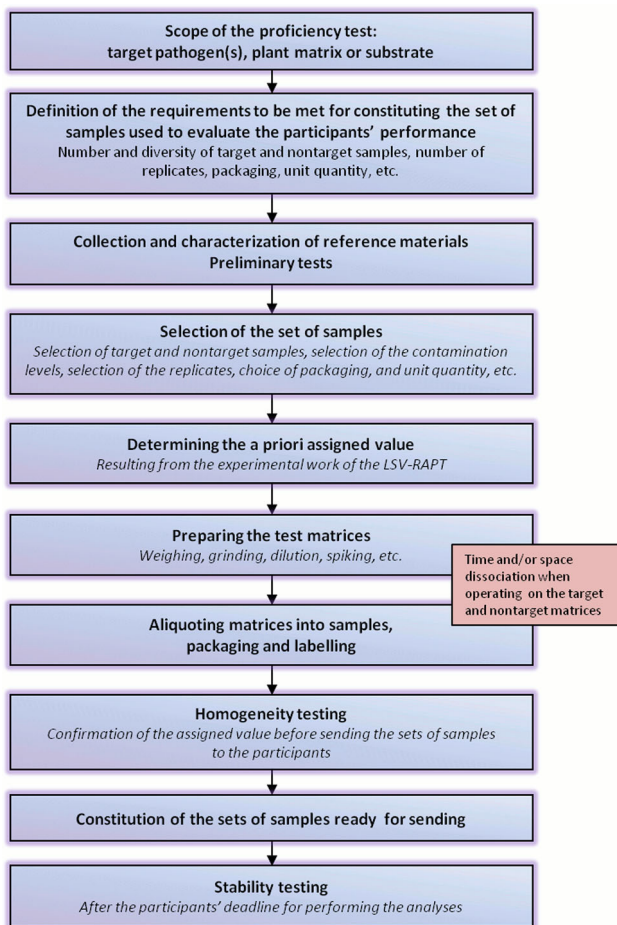
##### General requirements

The test samples are chosen to match (as closely as possible) with the materials encountered in routine testing (plant matrices, contamination levels).

The selection of samples is also linked to the statistical requirements needed to evaluate the different performance criteria. A minimum of five samples per criterion evaluated is a good compromise between the quality of evaluation and the practical constraints [17]. If a criterion is evaluated from five different samples, the minimum detectable error rate at 95 % is only 46 % [result obtained using the R statistical software (version 2.15.3; R Development Core Team, Vienna, Austria) with  $\text{binom.test}(0, 5, p = 0.46, \text{alternative} = \text{“less”})$ ].

This leads to a rough performance evaluation. Indeed, 46 % for over-detections (false positives) or under-detections (false negatives) is not acceptable for a laboratory performing official analyses in plant pathology. However, the selection of particular positive and negative samples representative of the diversity of situations that may be encountered in routine analyses significantly improves the quality of evaluation, although it is not quantified properly (e.g. the selection of positive samples representative of the target diversity, positive samples with low contamination levels, nontarget samples known to interfere or with high background noise).

Finally, the selection of samples is conditioned by knowledge of the pathogen (conservation, detectability) and technical difficulties (availability of the reference materials).



**Fig. 2** Main steps of the production and validation of proficiency test samples

When designing the proficiency test, the set of samples used to evaluate the participants' performance (number of target and nontarget samples, their nature, concentration, number of replicates, packaging and unit quantity of samples) is defined with the utmost care.

The definitive selection generally occurs after a preliminary stage, which is required to collect and characterize the reference materials and perform initial tests.

Once selected, the reference materials are used to prepare test matrices, which are aliquoted and packaged into test samples. Maximum care is taken to avoid any cross-contamination (time and/or space dissociation for preparing/aliquoting/packaging target and nontarget matrices).

### Example

Table 4 shows the detailed composition of the set of samples sent to each participant for the proficiency test 05XD organized in 2013 by LSV-RAPT. The proficiency test concerned the detection of the bacterium *Xanthomonas*

*axonopodis* pv. *dieffenbachiae* (Xad) in *Anthurium* spp. using the French official method MOA 028 [18] based on the PCR technique. This technique produces only qualitative results. A set of 17 coded samples was sent to each participant.

The composition of the set of samples was chosen to allow the evaluation, from qualitative results, of the performance criteria defined in Table 2:

*The sensitivity*—presence of ten target samples (samples A–F2).

- The inclusivity (as defined in Table 2 under the sensitivity section) was evaluated from anthurium samples artificially contaminated with target strains corresponding to a genetic and geographic diversity (representative of the diversity of French overseas departments and territories): strains from Martinique, Guadeloupe, Reunion Island and New-Caledonia (different years of isolation, over a period ranging from 1992 to 2004), and strains belonging to four different genetic subgroups (coded G2a/d/e/g).
- The detectability (as defined in Table 2 under the sensitivity section) was evaluated from anthurium samples artificially contaminated until a level of  $2 \times 10^5$  CFU  $\text{mL}^{-1}$  (where CFU is the number of colony forming units). This dilution level was chosen above the required detection level (RDL) of the official method of analysis (MOA 028), which is between  $5 \times 10^3$  CFU  $\text{mL}^{-1}$  and  $5 \times 10^4$  CFU  $\text{mL}^{-1}$ .

*The specificity*—presence of five nontarget samples (samples K–O), which included:

- a healthy anthurium sample, i.e. a symptomless sample not contaminated with any particular nontarget organism;
- anthurium samples artificially contaminated with nontarget organisms that can be found in routine samples and can interfere: saprophytes (*Curtobacterium flaccumfaciens*) and other anthurium pathogens (*Ralstonia solanacearum* and *Acidovorax anthurii*);
- an anthurium sample artificially contaminated with a genetically related nontarget organism (Xad not pathogenic to anthurium).

In addition, different varieties of *Anthurium* spp. were used for the nontarget samples: ACROPOLIS<sup>®</sup> (white flowering variety), LUCARDI<sup>®</sup> (pink-green flowering variety), TROPICAL<sup>®</sup> (red flowering variety), TROPIC NIGHT<sup>®</sup> (red-chocolate flowering variety), in order to check the absence of interference with the plant matrix.

*The repeatability*—presence of four target samples in duplicates (samples C, D, E and F) giving a total of eight repeated samples.

**Table 4** Description of the set of samples sent to each participant in the proficiency test 05XD concerning the detection of Xad in *Anthurium* spp. using the French official method MOA 028 [16]

Composition of the set of samples	Type	Nature	Final bacterial concentration (CFU mL <sup>-1</sup> )	Packaging	Assigned value	Used for performance assessment		
						Sensitivity	Specificity	Repeatability
Sample_A	Target samples	Anthurium leaves (var. ROSA <sup>®</sup> ) + bacterial suspension of Xad (G2a <sup>a</sup> , Martinique, 1992)	1 × 10 <sup>8</sup>	Leaves in crushing bag	Positive	X		
Sample_B		Anthurium leaves (var. ROSA <sup>®</sup> ) + bacterial suspension of Xad (G2d <sup>a</sup> , Guadeloupe, 1992)	3 × 10 <sup>8</sup>	Leaves in crushing bag	Positive	X		
Sample_C1		DNA extracts obtained from ground anthurium leaves (var. ROSA <sup>®</sup> ) spiked with Xad (G2g <sup>a</sup> , N. Caledonia, 2004)	2 × 10 <sup>7</sup>	DNA extracts in tube	Positive	X		X
Sample_C2		Identical to sample_C1			Positive	X		X
Sample_D1		DNA extracts obtained from ground anthurium leaves (var. ROSA <sup>®</sup> ) spiked with Xad (G2e <sup>a</sup> , N. Caledonia, 2004)	5 × 10 <sup>7</sup>	DNA extracts in tube	Positive	X		X
Sample_D2		Identical to sample_D1			Positive	X		X
Sample_E1		DNA extracts obtained from ground anthurium leaves (var. ROSA <sup>®</sup> ) spiked with Xad (G2a, Reunion Island, 1999)	2 × 10 <sup>6</sup>	DNA extracts in tube	Positive	X		X
Sample_E2		Identical to sample_E1			Positive	X		X
Sample_F1		DNA extracts obtained from ground anthurium leaves (var. ROSA <sup>®</sup> ) spiked with Xad (G2a, Reunion Island, 1999)	2 × 10 <sup>5</sup>	DNA extracts in tube	Positive	X		X
Sample_F2		Identical to sample_F1			Positive	X		X
Sample_K	Nontarget samples	Healthy anthurium leaves (var. LUCARDI <sup>®</sup> )	0	Leaves in crushing bag	Negative		X	
Sample_L		Anthurium leaves (var. ACROPOLIS <sup>®</sup> ) + bacterial suspension of Xad G3 not pathogenic to anthurium, ( <i>Dieffenbachia</i> , USA, 1950)	1 × 10 <sup>8</sup>	Leaves in crushing bag	Negative		X	
Sample_M		Anthurium leaves (var. TROPIC NIGHT <sup>®</sup> ) + bacterial suspension of <i>Curtobacterium flaccumfaciens</i> ( <i>Anthurium</i> , Reunion Island, 2001)	2 × 10 <sup>8</sup>	Leaves in crushing bag	Negative		X	
Sample_N		DNA extracts obtained from ground anthurium leaves (var. TROPICAL <sup>®</sup> ) spiked with <i>Ralstonia solanacearum</i> Phylotype II—sequevar 4 NP ( <i>Anthurium</i> , Martinique, 2001)	6 × 10 <sup>7</sup>	DNA extracts in tube	Negative		X	



**Table 4** continued

Composition of the set of samples	Type	Nature	Final bacterial concentration (CFU mL <sup>-1</sup> )	Packaging	Assigned value	Used for performance assessment		
						Sensitivity	Specificity	Repeatability
Sample_O		DNA extracts obtained from ground anthurium leaves (var. TROPICAL <sup>®</sup> ) spiked with <i>Acidovorax anthurii</i> ( <i>Anthurium</i> , Martinique, 1991)	8 × 10 <sup>7</sup>	DNA extracts in tube	Negative		X	
Sample_P	Lures	Randomly chosen among a pool of extra samples			None	No	No	No
Sample_Q					None	No	No	No

<sup>a</sup> G2a/d/e/g: codes for the genetic groups of Xad pathogenic to anthurium

This is one of the different ways of evaluating repeatability. It is also possible to use several replicates (e.g. five replicates) of only one sample contaminated by the target organism, preferably at a low contamination level (but above the required detection level).

*The accuracy*—which summarizes the three above criteria.

Two lures were also included in each set of samples. Results of the lures were not taken into account for evaluating the laboratories.

The samples were packaged in two forms: fresh anthurium leaves in crushing bags and DNA extracts in tubes, according to the samples. The packaging “Anthurium leaves in crushing bag” was chosen in addition to the packaging “DNA extracts in tube” although known to be less homogeneous, in order to evaluate the laboratories from the grinding step of the official method of analysis.

#### Sample validation

To ensure that the interlaboratory comparison is reliable, samples used to evaluate the participants’ performance are validated in terms of assigned value (status), homogeneity and stability.

#### *Assigned value (or accepted reference value)*

The assigned value is the value attributed to a particular property of a proficiency test sample. The proficiency tests organized by LSV-RAPT concern qualitative methods. Therefore, the assigned value is qualitative and corresponds to the status attributed to the sample: “positive” or “negative” (the target samples are chosen above the required detection level, and consequently, the indeterminate status is inappropriate).

The samples are assigned a value as a result of the LSV-RAPT’s experimental work. This value corresponds to the accepted reference value as defined in point 3.5b of ISO-

5725-1 [7]. It is defined independently of the participants’ results.

The LSV-RAPT’s experimental work consists of:

- establishing an a priori assigned value, based on the knowledge and/or analytical characterization of the reference materials used to produce the test matrices, as well as on the formulation of the test matrices;
- confirming the a priori assigned value during the homogeneity study by repeated analyses on the samples packaged in their final form.

The detailed procedure for the determination of the assigned value by LSV-RAPT is provided in Table S1 of the electronic supplementary material (ESM).

#### *Homogeneity and stability testing*

The homogeneity and the stability are assessed for all test matrices (positives or negatives), except for the lures.

The organizer assesses the homogeneity and stability by using the same standardized analytical method and measuring the same characteristic of the samples as defined in the PT design. The result of each repeat for each sample is interpreted according to the recommendations of the standardized method used to analyze the samples.

When possible, principles similar to quantitative methods are used to process the homogeneity and stability data produced with qualitative methods.

Consequently, when a qualitative method generates quantitative data, both quantitative and qualitative data are processed. It is also possible to use an indirect quantitative method (e.g. bacterial enumeration) when there is a proven link to the qualitative detection method (e.g. conventional PCR).

When quantitative data are unavailable (such as, for example, in the proficiency test 05XD), the statistical analysis is limited to the qualitative results (status homogeneity and stability).

When quantitative data are available, it is processed according to the ISO 13528 [3] and IUPAC Harmonized Protocol [4] using a standard deviation for the proficiency assessment defined by perception. The final conclusion for homogeneity or stability takes into account both quantitative and qualitative results. A decision matrix has been established to cover all situations (Table S2 of ESM).

It is worth noting that this decision matrix can be qualified as “fit for purpose.” The organizer must guarantee the qualitative results of the samples. Consequently, it is possible to consider samples as sufficiently homogeneous or stable even if they are not strictly homogeneous or stable in terms of the quantitative results.

In the same way, any sample with a qualitative homogeneity or a qualitative stability defect (i.e. a qualitative result of a repeat different from the assigned value) is automatically excluded from the evaluation of the participants’ performance.

### Homogeneity

The assessment of homogeneity is performed after the samples have been packaged in their final form and before distribution to participants.

For each test matrix, ISO 13528 [3] recommends testing a number  $g$  of samples (aliquots) in duplicate (where  $g \geq 10$ ).

However, this number is hardly compatible with practical constraints, given the number of samples used to evaluate the laboratories’ performance in plant pathology. This number can be reduced after a relevant risk analysis. In accordance with the ILAC discussion paper [19], LSV-RAPT decided to reduce the number of tested samples to five for negative samples and heavily contaminated positive samples (no more than 50 samples are prepared per test matrix). This number may be further reduced but always subjected to a relevant risk analysis (e.g. if suitable data are available from previous homogeneity testing on similar samples prepared by the same procedures and/or according to the organizer’s expertise and/or if very few samples are prepared per test matrix). The use of 10 samples is maintained for matrices for which a lack of homogeneity could have the most important consequences, for example a change in the assigned value, such as samples E and F for the proficiency test 05XD (cf. Table 4).

The  $g$  samples are randomly chosen, and the  $2g$  test portions are analyzed in a random order under repeatability conditions to minimize the differences between samples. The general average ( $\bar{x}$ ), standard deviation within samples ( $s_w$ ), standard deviation of sample averages ( $s_x$ ), standard deviation between samples ( $s_s$ ) and the standard deviation for proficiency assessment ( $\sigma_{PT}$ ) are calculated as shown in Table 5.

According to the ISO 13528 [3], the samples are considered to be adequately homogeneous if  $s_s \leq 0.3\sigma_{PT}$ . According to the IUPAC Harmonized Protocol [4], the samples are considered to be sufficiently homogenous if

$$s_s^2 \leq F_1 \cdot (0.3\sigma_{PT})^2 + F_2 \cdot s_w^2$$

where  $F_1$  and  $F_2$  are constants that are derived from standard statistical tables as  $F_1 = \chi_{m-1,0.95}^2 / (m-1)$  where  $\chi_{m-1,0.95}^2$  is the value exceeded with probability 0.05 by a chi-squared random variable with  $m-1$  degrees of freedom, and  $F_2 = 1/2(F_{m-1, m, 0.95} - 1)$  where  $F_{m-1, m, 0.95}$  is the value exceeded with probability 0.05 by a random variable with a Fisher distribution with  $m-1$  and  $m$  degrees of freedom.

### Stability

The samples used for the stability study are subjected to identical conditions as the samples sent to the participants. They are prepared and packaged in the same way, and they are subjected to simulated travel conditions (they are kept at ambient temperature in a room without air-conditioning system for the longest delivery time for the participants’ parcels).

Then, they are stored as recommended until the stability test is performed. The stability test is performed just after the participants’ deadline for performing analyses. As recommended by ISO 13528 [3], a number  $g$  of samples (aliquots) are analyzed in duplicate (where  $g \geq 3$ ). The  $g$  samples are randomly chosen, and the  $2g$  test portions are analyzed in a random order under repeatability conditions to minimize the differences between samples.

The general average of the measurements obtained in the homogeneity study ( $\bar{x}$ ) is compared to the general average of the results obtained in the stability study ( $\bar{y}$ ). According to the ISO 13528 [3], the samples are considered to be adequately stable if  $|\bar{x} - \bar{y}| \leq 0.3\sigma_{PT}$ .

### Example

The following example illustrates the homogeneity and stability calculations and the decision associated with the results from four samples.

The four samples were produced for proficiency tests concerning the detection of a banana virus (the cucumber mosaic virus or CMV) in banana leaves using the French official method MOA 009 [20] based on the ELISA technique. As previously explained in the section “Qualitative data”, the ELISA technique produces a quantitative result (absorbance value), which is compared to threshold values to give a qualitative result. For the detection of CMV, the absorbance value ( $A$ ) is photometrically measured at a

**Table 5** Example of homogeneity and stability calculations for a positive sample and a negative sample analyzed using the ELISA technique

	Example of data processing for a positive sample (AA)					Example of data processing for a negative sample (AP)				
	Sample number	Quantitative results ( $A^a$ )		Qualitative results		Sample number	Quantitative results ( $A$ )		Qualitative results	
		Repeat 1	Repeat 2	Repeat 1	Repeat 2		Repeat 1	Repeat 2	Repeat 1	Repeat 2
Homogeneity results	AA1.1	2.635	2.685	Positive	Positive	AP1.1	0.015	0.013	Negative	Negative
	AA1.2	2.578	2.537	Positive	Positive	AP1.2	0.019	0.016	Negative	Negative
	AA1.3	2.103	2.062	Positive	Positive	AP1.3	0.023	0.014	Negative	Negative
	AA1.4	2.078	1.954	Positive	Positive	AP1.4	0.022	0.020	Negative	Negative
	AA1.5	2.240	2.266	Positive	Positive	AP1.5	0.019	0.014	Negative	Negative
Stability results	AA2.1	1.429	1.415	Positive	Positive	AP2.1	0.014	0.011	Negative	Negative
	AA2.2	2.911	2.859	Positive	Positive	AP2.2	0.013	0.019	Negative	Negative
	AA2.3	2.444	2.428	Positive	Positive	AP2.3	0.020	0.021	Negative	Negative
$\sigma_{PT}$	Equal to 15 % of the mean absorbance value obtained from the sample during the homogeneity study					Equal to the mean absorbance value obtained from the negative control samples during the homogeneity study (0.019)				
Calculation for quantitative homogeneity	$\bar{x} = 2.314$ $\sigma_{PT} = 2.314 \times 0.15 = 0.347$ $s_x = 0.285$ $s_w = s_{an} = 0.047$ $s_s = s_{sam} = 0.283$ ISO 13528 [4]: $\frac{s_s}{\sigma_{PT}} = 0.816 > 0.3$ $F_1 \cdot (0.3\sigma_{PT})^2 + F_2 \cdot s_{an}^2 = 0.030^b$ IUPAC [5]: $s_{sam}^2 = 0.080 > 0.030$					$\bar{x} = 0.017$ $\sigma_{PT} = 0.019$ $s_x = 0.003$ $s_w = s_{an} = 0.003$ $s_s = s_{sam} = 0.000$ ISO 13528 [4]: $\frac{s_s}{\sigma_{PT}} = 0.036 \leq 0.3$ $F_1 \cdot (0.3\sigma_{PT})^2 + F_2 \cdot s_{an}^2 = 0.0001$ IUPAC [5]: $s_{sam}^2 = 0.0000 \leq 0.0001$				
Calculation for quantitative stability	$\bar{y} = 2.247$ $0.3\sigma_{PT} = 0.3 \times 0.347 = 0.104$ $ \bar{x} - \bar{y}  = 2.314 - 2.247 = 0.067 \leq 0.104$					$\bar{y} = 0.016$ $0.3\sigma_{PT} = 0.3 \times 0.019 = 0.006$ $ \bar{x} - \bar{y}  = 0.017 - 0.016 = 0.001 \leq 0.006$				
Conclusion	No quantitative homogeneity Quantitative stability					Quantitative homogeneity Quantitative stability				

<sup>a</sup> A: absorbance value photometrically measured at a wavelength of 405 nm using Sunrise™ Microplate Absorbance Reader (Tecan, Lyon, France)

<sup>b</sup>  $F_1$  and  $F_2$  are constants that are derived from standard statistical Tables [5]

wavelength of 405 nm, and the threshold values are determined as follows:  $S_1 = 2A_c$  and  $S_2 = 3A_c$ , where  $A_c$  denotes the mean absorbance of the negative control samples.

The positivity threshold ( $S_2$ ) obtained during the homogeneity and stability studies of the four samples used as examples here, was on average 0.060 (with a standard deviation of 0.008).

The description of the four samples is summarized in Table 6.

The calculations of homogeneity and stability are illustrated from a positive sample (sample AA) and a negative sample (sample AP) in Table 5. The standard deviation for proficiency assessment ( $\sigma_{PT}$ ) is determined by expert perception according to the following principle.

For positive samples,  $\sigma_{PT}$  is equal to 15 % of the mean absorbance value obtained from the samples during the homogeneity study, up to a minimum value equal to the mean absorbance value obtained from the negative control samples during the homogeneity study (0.019 in this example).

For negative samples,  $\sigma_{PT}$  is equal to the mean absorbance value obtained from the negative control samples during the homogeneity study (0.019 in this example).

The same calculations applied to the four samples give the final results summarized in Table 7.

All samples are homogeneous and stable as far as the qualitative data are concerned.

Sample AA is not homogenous but stable in relation to the quantitative results. The degree of confidence that can be placed in the stability results is low due to the lack of

**Table 6** Description of the samples used as examples for homogeneity and stability calculations

Sample code	Nature	Contamination level	Packaging	Assigned value	Number of samples <sup>a</sup> for homogeneity study	Number of samples <sup>a</sup> for stability study
Sample_AA	Banana leaves contaminated by the target virus	High $A^b > 1$	Freeze-dried leaves in crushing bag	Positive	5	3
Sample_AX	Banana leaves contaminated by the target virus	Low $A < 0.3$	Freeze-dried leaves in crushing bag	Positive	10	3
Sample_AD	Banana leaves contaminated by the target virus	High $A > 1$	Freeze-dried ground leaves in tube	Positive	5	3
Sample_AP	Banana leaves contaminated by a nontarget virus	0	Freeze-dried leaves in crushing bag	Negative	5	3

<sup>a</sup> Analyzed in duplicate

<sup>b</sup> A: absorbance value photometrically measured at a wavelength of 405 nm using Sunrise<sup>TM</sup> Microplate Absorbance Reader (Tecan, Lyon, France)

**Table 7** Homogeneity and stability final results for samples AA, AX, AD and AP

Sample number	Results of homogeneity			Results of stability			Uncertainty <sup>b</sup>	Used for performance assessment
	Mean absorbance	Quantitative homogeneity	Qualitative homogeneity	Mean absorbance	Quantitative stability	Qualitative stability		
Sample_AA	2.314 (Min <sup>a</sup> = 1.954)	No	Yes	2.247 (Min = 1.415)	Yes	Yes	Low	Yes
Sample_AX	0.253 (Min = 0.215)	No	Yes	0.244 (Min = 0.221)	Yes	Yes	High	No
Sample_AD	1.262 (Min = 1.139)	Yes	Yes	0.977 (Min = 0.947)	No	Yes	Low	Yes
Sample_AP	0.017	Yes	Yes	0.016	Yes	Yes	Low	Yes

<sup>a</sup> Min: minimum

<sup>b</sup> The uncertainty takes into account both the degree of confidence that can be placed in the stability results and the risk linked to the contamination level of the target pathogen

quantitative homogeneity in the sample. Indeed, the sampling for stability testing is low (only three samples). Consequently, if it is known that the population is not homogeneous, there is a risk that the results of stability obtained from three samples may not be representative of the population. Thus, the population could be wrongly identified as stable. However, as the minimum absorbance value obtained during both homogeneity and stability testing (1.415) is much higher than the positivity threshold (0.060), this sample, which is heavily contaminated by the target virus, can be considered as sufficiently homogeneous and stable and is, therefore, suitable for use to evaluate participants.

For sample AX, the reasoning is similar, except that the sample is slightly contaminated. Given the samples' lack of quantitative homogeneity, there is a risk that the results of stability are not representative of the population. In addition, the minimum absorbance value (0.215) is close to the positivity threshold. Consequently, this sample is excluded from the evaluation of participants. Results of stability are presented, although the sample could be excluded on the

basis of the homogeneity results. The sample is not sufficiently homogeneous to be used to evaluate participants.

Sample AD is homogeneous but not stable for the quantitative results. As it is homogeneous, the degree of confidence that can be placed in the stability results is high. Therefore, even if it is not quantitatively stable, the minimum absorbance value (0.947) is much higher than the positivity threshold and the sample can be considered as stable and homogeneous enough to be used to evaluate the participants' performance.

Sample AP presents no problems of homogeneity or stability, and therefore, it can be used to evaluate the laboratories' performance.

#### Generalization

The previous example is based on samples analyzed using the ELISA technique. The same approach can be used for samples analyzed using real-time PCR (with the Ct values), but only for positive samples (negative samples do not

**Table 8** Descriptive analysis of the results in the proficiency test 05XD per laboratory (only for laboratories L06 and L07)

Laboratory		L06			L07		
Samples	Assigned value	Decoding	Result submitted	Agreement/deviation	Decoding	Result submitted	Agreement/deviation
Sample_A	Positive	11	Positive	PA	7	Positive	PA
Sample_B	Positive	16	Positive	PA	14	Positive	PA
Sample_C1	Positive	1	Positive	PA	3	Positive	PA
Sample_C2	Positive	4	Positive	PA	12	Positive	PA
Sample_D1	Positive	5	Positive	PA	9	Positive	PA
Sample_D2	Positive	15	Positive	PA	1	Positive	PA
Sample_E1	Positive	14	Positive	PA	17	Positive	PA
Sample_E2	Positive	12	Positive	PA	15	Positive	PA
Sample_F1	Positive	13	Positive	PA	10	Positive	PA
Sample_F2	Positive	17	Positive	PA	6	Positive	PA
Sample_K	Negative	8	Indeterminate	<b>PD</b>	4	Negative	NA
Sample_L	Negative	9	Indeterminate	<b>PD</b>	2	Negative	NA
Sample_M	Negative	3	Negative	NA	8	Negative	NA
Sample_N	Negative	10	Negative	NA	16	Negative	NA
Sample_O	Negative	6	Negative	NA	5	Negative	NA
Sample_P	None	2	Positive	Not evaluated	13	Positive	Not evaluated
Sample_Q	None	7	Positive	Not evaluated	11	Positive	Not evaluated

PA, positive agreement; PD, positive deviation; NA, negative agreement

The bold font is used to indicate deviations (from the assigned value) in participants' results

produce a Ct value for the target pathogen, which means that only qualitative results can be used).

For samples analyzed by counting (e.g. bacterial enumeration), the quantitative homogeneity data are processed using  $T_1$ – $T_2$  test, as detailed in ISO/TS 22117 [21].

### Examples for data processing

The data processing is illustrated from the results of the laboratories that participated in the proficiency test 05XD (six participants) for the detection of Xad in *Anthurium* spp. using the French official method MOA 028 [18]. The samples used to evaluate the participants' performance are described in Table 4.

#### Results submitted by the participants

The participants' performance was evaluated as a function of the qualitative results submitted by the participants, according to the data processing method explained in Table 1.

For example, details of the processing of the results submitted by laboratories L06 and L07 are shown in Table 8.

#### Descriptive and performance statistics

The criteria presented in Table 2 are applied to each laboratory participating in the proficiency test 05XD, to give the results shown in Table 9.

The reproducibility of the results under the conditions of this proficiency test was 97 % (details on reproducibility calculations are provided in Data S2 of ESM).

#### Analysis of the performance statistics

In the given example, only the results of laboratory L06 were nonconforming for the global criterion of accuracy (87 % instead of 100 %, i.e. two discordant results in fifteen results), which can be explained by a problem of specificity (60 % instead of 100 %). The two discordant results were false positive results.

The review of this participant's analysis data sheets suggested a problem of cross-contaminations between positive and negative test samples.

#### Performance monitoring

When the data are available (existence of previous proficiency tests concerning the same pathogen and method), a graph to show the monitoring of performance can be presented for the laboratories that participated in the previous proficiency tests.

For example, Fig. 3 represents the evolution of the performance of four laboratories that participated in the three proficiency testing rounds concerning the detection of CMV using the French official method MOA 009 [20] based on the ELISA technique. In this example, there has



**Table 9** Evaluation of performance criteria for each laboratory in the proficiency test 05XD

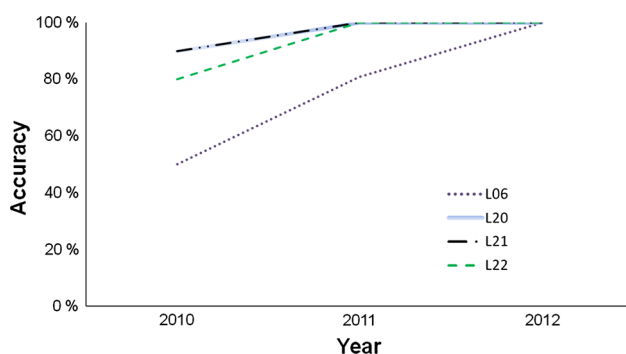
Laboratory	L06	L07	L09	L14	L19	L20
Number of PA	10	10	10	10	10	10
Number of NA	3	5	5	5	5	5
Number of ND	0	0	0	0	0	0
Number of PD	2	0	0	0	0	0
Sensitivity ( <i>SE</i> , %) <sup>a</sup>	100	100	100	100	100	100
Specificity ( <i>SP</i> , %) <sup>a</sup>	<b>60</b>	100	100	100	100	100
Repeatability ( <i>DA</i> , %) <sup>a,b</sup>	100	100	100	100	100	100
Accuracy ( <i>AC</i> , %) <sup>a</sup>	<b>87</b>	100	100	100	100	100

PA, positive agreement; NA, negative agreement; ND, negative deviation; PD, positive deviation

<sup>a</sup> Performance criteria defined in Table 2

<sup>b</sup> Details on repeatability calculations are provided in Data S1 of ESM

The bold font is used to indicate nonoptimal results



**Fig. 3** Monitoring of the performance for the CMV detection: Accuracy results of laboratories that participated in the three PT rounds for the detection of CMV

been a gradual improvement in the results since 2010. By 2012, all the laboratories had conforming results. The effectiveness of corrective and preventive actions taken by the participants following the nonconformities reported by the organizer may explain this improvement.

## Conclusion

With the growing importance of proficiency testing, there is a greater need for additional guidance on the organization of proficiency tests in specific sectors because the available guidelines are not totally adapted.

The Anses Plant Health Laboratory developed a methodology to organize proficiency tests in plant pathology in

accordance with the general requirements of the ISO/IEC 17043 standard [2].

In the same way that laboratories have to demonstrate their capacity to produce accurate analytical results, proficiency test organizers also have to demonstrate their competence in organizing proficiency tests in their area of expertise. Accreditation as proficiency testing organizer provides the PT participants a reasonable degree of confidence in their decision-making process based on proficiency testing results. In this way, each tier of the analytical pyramid contributes to improving the reliability of analytical results, which is essential in plant pathology in order to guarantee the import/export of pest-free plant material. This issue is of such importance in plant pathology that the European Plant Protection Organization (the intergovernmental organization responsible for European cooperation in plant protection in the European and Mediterranean region) is preparing guidelines for the organization of interlaboratory comparisons by plant pest diagnostic laboratories.

The methodology developed by LSV-RAPT provides a basis that can be applied to many other disciplines, which also use qualitative detection methods (e.g. medical, food and veterinary sciences).

**Acknowledgments** We would like to thank all the people involved in organizing the test at LSV-RAPT and all the laboratories that took part in the proficiency tests organized by LSV-RAPT.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. Stuart M, Squirell A (2001) Accreditation of providers of proficiency testing schemes. *Accred Qual Assur* 6:203–205
2. International Organization for Standardization (2010) Conformity assessment—general requirements for proficiency testing. ISO/IEC 17043:2010. ISO, Geneva
3. International Organization for Standardization (2005) Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO 13528:2005. ISO, Geneva
4. Thompson M, Ellison E, Wood R (2006) The international harmonized protocol for proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). *Pure Appl Chem* 78(4):145–196
5. Tholen DW (2001) Statistical treatment of proficiency testing data. *Accred Qual Assur* 3:362–366
6. Fearn T, Thompson M (2001) A new test for ‘sufficient homogeneity’. *Analyst* 126(8):1414–1417
7. International Organization for Standardization (2005) Accuracy (trueness and precision) of measurement methods and results—part 1: general principles and definitions. ISO 5725-1:1994. ISO, Geneva

8. International Organization for Standardization (2006) Reference materials—general and statistical principles for certification. ISO Guide 35:2006. ISO, Geneva
9. Wehling P, LaBudde RA, Brunelle SL, Nelson MT (2011) Probability of detection (POD) as a statistical model for the validation of qualitative methods. *J AOAC Int* 94(1):335–347
10. Bashkansky E, Gadrich T, Kuselman I (2012) Interlaboratory comparison of test results of an ordinal or nominal binary property: analysis of variation. *Accred Qual Assur* 17:239–243
11. Gadrich T, Bashkansky E, Kuselman I (2013) Comparison of biased and unbiased estimators of variances of qualitative and semi-quantitative results of testing. *Accred Qual Assur* 18:85–90
12. Uhlig S, Krügener S, Gowik P (2013) A new profile likelihood confidence interval for the mean probability of detection in collaborative studies of binary test methods. *Accred Qual Assur* 18:367–372
13. International Organization for Standardization (2003) Microbiology of foods and animal feeding stuffs-protocol for the validation of alternative methods ISO 16140:2003. ISO, Geneva, Switzerland
14. Association of Official Agricultural Chemists International Methods Committee Guidelines for Validation of Microbiological Methods for Food and Environmental Surfaces. AOAC Pre-Publication Draft (2012). AOAC Standard Development
15. McClure FD (1990) Design and analysis of qualitative collaborative studies: minimum collaborative program. *J Assoc Off Anal Chem* 73(6):953–960
16. International Organization for Standardization (2005) General requirements for the competence of testing and calibration laboratories. ISO/IEC 17025:2005. ISO, Geneva
17. Association française de normalisation (2002) Animal health analysis methods—guide for the organization of ring tests suitable for animal immuno-serology. NF U 47-400:2002. AFNOR, Saint-Denis La Plaine
18. French Ministry of Agriculture (2013) Detection of *Xanthomonas axonopodis* pv. *dieffenbachiae* (Xad) in *Anthurium* spp. MOA 028 version 1a:2013. French Ministry of Agriculture, Paris
19. Tholen D, Boley N, Gun-Monro J, Gust J, Korth W, Noble M, Petinos P, Wruck D (2006) ILAC discussion paper on homogeneity and stability testing
20. French Ministry of Agriculture (2010) Detection of *Cucumber Mosaic Virus* (CMV) in Banana leaves (*Musa* spp.) using ELISA serological technique. MOA 009 version 1a:2010. French Ministry of Agriculture, Paris
21. International Organization for Standardization (2010) Microbiology of food and animal feeding stuffs—specific requirements and guidance for proficiency testing by interlaboratory comparison. ISO/TS 22117:2010. ISO, Geneva