

RESEARCH ARTICLE

Interlaboratory evaluation of two Reverse-transcriptase Polymeric Chain Reaction-based methods for detection of four fruit tree viruses

S. Massart¹, Y. Brostaux², L. Barbarossa³, A. Batlle⁴, V. Cesar⁵, O. Dutrecq⁶, F. Fonseca⁷, R. Guillem⁸, B. Komorowska⁹, A. Olmos¹⁰, S. Steyer¹¹, T. Wetzel¹², J. Kummert¹ & M.H. Jijakli¹

1 Plant Pathology Unit, Gembloux Agricultural University (FUSAGx), Passage des Déportés, Gembloux, Belgium

2 Unité de Statistique, Informatique et Mathématiques appliquées (SIMa), FUSAGx, Gembloux, Belgium

3 CNR-Istituto di Virologia Vegetale-Sezione di Bari, Via Amendola, Bari, Italy

4 Department de Protecció Vegetal, Institut de Recerca i Tecnologia Agroalimentaries, Ctra. de Cabrils, Cabrils, Barcelona, Spain

5 Corder ASBL, Unité de Phytopathologie, Croix du Sud, Louvain-La-Neuve, Belgium

6 DNALis sprl, rue de la Croix-Rouge, Gembloux, Belgium

7 University of Algarve, CDCTPV – Molecular Plant Virology Laboratory, Campus de Gambelas, Gambelas, Faro, Portugal

8 Service Régional de la protection des végétaux de la Direction Régionale de 'Agriculture et de la Forêt d'Aquitaine, Bordeaux, France

9 Virology Laboratory, Research Institute of Pomology and Floriculture, Pomologiczna, Skierniewice, Poland

10 Instituto Valenciano de Investigaciones Agrarias (IVIA), Carretera de Moncada a Náquera, Moncada (Valencia), Spain

11 Département Lutte biologique et Ressources phylogénétiques, Centre Wallon de Recherches Agronomiques, Rue de Liroux, Gembloux, Belgium

12 RLP Agrosience, AlPlanta – Institute for Plant Research, Breitenweg, Neustadt an der Weinstrasse, Germany

Keywords

Interlaboratory; plant; ring test; sensitivity; specificity; virus.

Correspondence

M.H. Jijakli, Plant Pathology Unit, Gembloux Agricultural University (FUSAGx), Passage des Déportés, 2, B-5030 Gembloux, Belgium.
Email: phytopat@fsagx.ac.be

Received: 14 December 2007; revised version accepted: 3 July 2008.

doi:10.1111/j.1744-7348.2008.00281.x

Abstract

Recent technological development of molecular methods has led to the proliferation of new rapid PCR or reverse-transcriptase (RT)-PCR-derived diagnostic tests for plant viruses. Nevertheless, for routine use, the reliability of all these new methods is not widely established and there is still an apprehension to adopt them in official diagnostic for certification of plant material. This is partly because of the lack of confidence in the obtained results and the poor knowledge on the reproducibility and limits of the RT-PCR protocols. There is a lack of information on the adequate risk assessment in the use of this new technology. An interlaboratory evaluation of two RT-PCR duplex protocols for the detection of four different fruit tree viruses was performed to address these questions. Identical samples were sent as crude extract preparation to each of the participant laboratories. Samples were coded to ensure a double-blind test. General principles of result analysis are described, for example calculation of parameters such as specificity, sensitivity, repeatability, reproducibility, likelihood ratios and post-test probabilities. These parameters and the integration of the protocols within official certification scheme are discussed. Finally, guidelines for researchers desirous of validating their new plant virus diagnostic protocols through interlaboratory evaluation are suggested.

Introduction

Plant viruses may remain latent and can be present in very low titres. Their detection in seeds, propagative material and other reservoirs is a priority to avoid their widespread distribution and large economic losses. Phytosanitary certification schemes have been established worldwide to

certify the propagation of virus-free plant material. The implementation of these schemes to control disease requires techniques with high sensitivity and specificity, such as polymerase chain reaction (PCR)-based methods. New rapid diagnostic tests based on molecular methods are continuously developed, holding the promise for an improved management and control of infectious diseases

(Banoo *et al.*, 2006). The proliferation of PCR-based methods has also revolutionised molecular diagnosis for plant viruses. Nevertheless, their development is rarely accompanied with appropriate evaluation of reliability and reproducibility (Olmos *et al.*, 2007). Consequently, there are concerns about their routine use, delaying their application in routine diagnostic.

Diagnostic tests are tools that increase the information about the sanitary status of a plant, strengthening or lessening the probability of infection. It must be assumed that there is no perfect method without false-positive and/or false-negative results and that a test gives more accurate probability of presence or absence of infection (post-test probability of infection). Any sample to be tested has a pre-test probability of infection that corresponds to the prevalence of infection within the studied population. The post-test probability of infection represents the probability of infection of the sample after carrying out the test, whatever the result. When a sample is diagnosed as infected, the post-test probability of infection is higher than the pre-test probability of infection. Alternatively, when a sample is diagnosed as healthy, the post-test probability of infection is lower than the pretest probability of infection. It is therefore mandatory to know the extent to which the post-test probability of infection is strengthened or lessened using a particular test. The evaluation of this property is regularly performed through interlaboratory evaluations in human or animal virology (Bootman *et al.*, 1999; Ruelle *et al.*, 2004; Ferris *et al.*, 2006) but is more scarcely documented in plant virology.

The interlaboratory evaluation of a new diagnostic technique provides essential information on (a) the test performance, for example its specificity, sensitivity, repeatability and reproducibility, (b) the risk management, for example the likelihood ratios and the post-test probabilities of infection, (c) the ease of implementation, (d) the conditions of use and (f) the storage conditions of samples and reagents. A standard protocol can be subsequently elaborated based on results and suggestions from participants. For all these reasons, multicentre trials are an essential step in the adequate evaluation of a diagnostic test, ensuring its adequate integration within diagnostic schemes, a correct interpretation of results and the design of optimal risk management strategies. Finally, this information will improve the management of plant diseases in nurseries, facilitating the decision-making for growers and regulatory agencies.

In this study, two duplex reverse-transcriptase (RT)-PCR methods, optimised (Massart *et al.*, 2008) with previously published primers (Kummert *et al.*, 2000), were evaluated. These methods were designed to detect four different fruit tree viruses belonging to the genus *Ilarvirus* (*Apple mosaic virus*, ApMV), *Capillovirus* (*Apple stem grooving*

virus, ASGV), *Trichovirus* (*Apple chlorotic leafspot virus*, ACLSV) and *Foveavirus* (*Apple stem pitting virus*, ASPV). We report the interlaboratory evaluation of these methods for the simultaneous detection of ACLSV/ApMV and ASGV/ASPV. General principles of result analysis are detailed and could provide useful guidelines for researchers desirous of validating their plant pathogen diagnostic protocols through interlaboratory evaluation.

Materials and methods

Sample preparation

Shoots from apple, plum and cherry trees, 1 year old, were collected in spring in a reference orchard of FUSAGx and CRA-W in Gembloux. Tree status (healthy, single infection or multiple infection) in the orchard has been assessed by biological indexing and enzyme-linked immunosorbent assay (ELISA) test. Additionally, ELISA tests were carried on during sample preparation and confirmed the tree status. Plant material (0.2 g of vascular tissue) was placed in nylon mesh bags (Bioreba, Reinach, Switzerland) and 2 mL of cold KAJI buffer (DNAlis, Gembloux, Belgium) was added. After grinding plant material (Homex and Bioreba, Reinach, Switzerland), the crude extracts were conserved at -20°C . Subsequently, crude extracts were diluted 10 times in fresh ultrapure water (Invitrogen, La Jolla, CA, USA) and divided into identical aliquots of 50 μL .

Sample distribution

For each duplex, 40 tubes, containing 50 μL of 10 \times diluted crude extract, were sent to each partner, together with a positive and a negative control. These tubes represented 20 samples sent in two repetitions. The samples were coded to ensure a double-blind test, that is neither the sample status nor the correspondence between repetitions was known by the participants. The frozen samples were sent in solid CO_2 and arrived still frozen in the nine laboratories (randomly numbered A to I). The laboratories involved in the assay corresponded to official certification laboratories, diagnostic laboratories, research centres or universities involved in molecular test development and validation. All the reagents for PCR amplification were provided by FUSAGx to each laboratory in the same package as the samples.

RT-PCR

All the laboratories followed the same detailed protocol. The One-step RT-PCR kit (Qiagen, Hilden, Germany) was used as follows: RT-PCR buffer 1 \times , 400 nM of each

primer (Eurogentec, Liège, Belgium), 400 nM dNTPs, 1 µL of enzyme mix, 2 µL of RNase-free water and 5 µL of 100× diluted crude extract in a total volume of 25 µL. The primers (Eurogentec) and amplicon sizes are listed in Table 1. Hundred times diluted crude extracts were prepared by each partner by adding 450 µL of fresh ultrapure water to the received samples. The thermal cycle consisted in a first step of 30 min at 50°C (retrotranscription), followed by 15 min at 95°C (polymerase activation) and 40 cycles at 95°C for 45 s, 55°C for 45 s and 72°C for 1 min followed by an additional step at 72°C for 10 min. PCR products were subjected to electrophoresis in agarose gel and revealed with ethidium bromide. The laboratories used one of the following thermocyclers: PTC 200 (MJ Research, Bio-Rad, Hercules, CA, USA), GeneAmp® PCR System 2700 and 2720 (Applied Biosystem, Foster City, CA, USA), Mastercycler® 5341 and Gradient (Eppendorf, Hamburg, Germany) and I-cycler (Bio-Rad).

Documentation of the results

The participating laboratories mailed the results directly to FUSAGx. This included the gel pictures and their analysis made in each laboratory. The gel pictures were carefully examined at FUSAGx, and the results were approved for inclusion in statistical analysis. The possible discordances observed between the documented results of the participant laboratory and the agarose gel observations at FUSAGx were thoroughly discussed with the corresponding laboratory. Additionally, a comment page, reporting the material used and the possible deviations to the protocol, was filled by all the participants.

Preliminary statistical analysis

The results were first checked to identify carry-over contamination in the no-template and negative controls. Sensitivity and specificity of the method were estimated according to Altman & Bland (1994). Sensitivity (specificity) was the proportion of true positives (negatives)

that were correctly identified by the method. Both parameters were calculated for both viruses together for each repetition in each laboratory. Additionally, extreme values of sensitivity or specificity were pointed out by a warning threshold. The threshold was set up based on a binomial distribution of the results. Indeed, in such interlaboratory evaluation, the obtained results follow a binomial distribution with two parameters: n (number of samples) and P (probability that a correct result is given by the test, corresponding to the average sensitivity or specificity). The lower limit of observed sensitivity (specificity) was defined as the threshold below which a single repetition has only 5% chance of falling if all repetitions share the same sensitivity (specificity).

Statistical analysis

The sensitivity and the specificity of the methods were calculated as described above for each virus separately. One-sided 95% confidence intervals were calculated for the global estimation of both properties for each virus, using the Agresti–Coull method (Agresti & Coull, 1998). Those confidence intervals were constructed so that, in 95% of the trials, the real value of the parameter (either specificity or sensitivity) for the corresponding virus is higher than this lower limit.

Repeatability refers to within-laboratory agreement between replicate observations of the same test performed by the same observer under similar conditions (Bland & Altman, 1986). Repeatability calculation was based on the agreement between both repetitions of each sample sent in each laboratory. Reproducibility refers to between-laboratory agreement, that is agreement between separate observations whatever the laboratory and the conditions. Repeatability and reproducibility of the RT-PCR test were estimated through the calculation of Cohen's kappa coefficients (Cohen, 1960), which measure the agreement of a classification between repetitions. In brief, a greater kappa value reflects a stronger agreement between the repetitions. The

Table 1 Primers used in this ring test

| Virus | Name | Sequence | Position in Viral RNA | Amplicon Size (in bp) |
|-------|------|----------------------------|-----------------------|-----------------------|
| ApMV | 1F | AGAAGTGACTGCCACGGTTGAAG | 237–259 | 312 |
| | 1R | CCTCAAATTCTGCTTAAAGCGGCG | 548–525 | |
| ACLSV | 5F | GCCTACAAATTAGGTGAGAGGCTC | 5564–5585 | 290 |
| | 8R | TTCCAATGGATCATGAGGTC | 5851–5832 | |
| ASPV | 4F | GAGTCTGATTATGAGGCATTTGATGC | 5926–5951 | 251 |
| | 4R | GCTTCCCTCCCATTGAGATCATAC | 6176–6153 | |
| ASGV | 5F | CCTGAATTGAAAACCTTTGCTGC | 6019–6041 | 344 |
| | 5R | CACGACTCCTAACCTCCAGTTCC | 6362–6339 | |

ACLSV, Apple chlorotic leafspot virus; ApMV, Apple mosaic virus; ASGV, Apple stem grooving virus; ASPV, Apple stem pitting virus.

kappa coefficient also represents how much better agreement is than what would result from chance only. For interpretation of the kappa value, the following guidelines were used: 0–20%, no agreement; 21–40%, weak agreement; 41–60%, moderate agreement; 61–80%, strong agreement and 81–100%, almost perfect agreement (Landis & Koch, 1977). Repeatability was evaluated by averaging the kappa coefficient between repetitions in each laboratory, and the reproducibility by averaging those coefficients between all pairs of repetitions whatever their origin (Light, 1971).

The relation between pretest and post-test probability of infection is accurately described by Baye's theorem (Fagan, 1975). This relation is based on odds [odds = probability/(1 – probability)] and likelihood ratios of the test: post-test odds = pretest odds × likelihood ratio. A likelihood ratio represents the extent to which the probability of infection for the tested sample raises or drops depending on test result. If a sample is diagnosed positive (negative), the positive (negative) likelihood ratio should be used to calculate post-test odds. The positive (negative) likelihood ratio of a test indicates how much the odds of infection is increased (decreased) if the sample is diagnosed infected (healthy) by the test. The positive likelihood ratio of a test is calculated as follow: sensitivity/(1 – specificity), while the negative likelihood ratio of a test corresponds to (1 – sensitivity)/specificity. Therefore, knowing the likelihood ratios of a particular test and the infection prevalence, the post-test probability of infection can be calculated using either the above-mentioned formulae or, alternatively, the Fagan's nomogram (Fagan, 1975).

Results

Preliminary analysis of the documented results

Results of the test (see example in Fig. 1) and comment pages were carefully scrutinised and a preliminary statistical analysis was carried on. First of all, a carry-over contamination in the no-template control and the negative control was observed with the second repetition of the method in two laboratories for ASGV/ASPV detection. The preliminary statistical analysis pointed some repetitions with sensitivity or specificity values lower than the fixed warning thresholds: laboratory B (second repetition ApMV/ACLSV), laboratory E (first repetition, ASPV/ASGV, both repetitions ApMV/ACLSV), laboratory F (second repetition, ApMV/ACLSV), laboratory G (first repetition, ASPV/ASGV) and laboratory H (first repetition ASGV/ASPV). This analysis was followed by thorough discussions with each laboratory. Some extreme values were correlated with major deviations from the

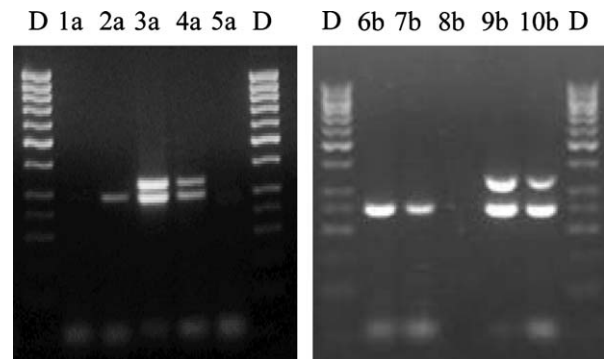


Figure 1 Visualisation of Apple chlorotic leafspot virus/Apple mosaic virus (a) or Apple stem grooving virus/Apple stem pitting virus (b) infection by gel electrophoresis of PCR products. D, O'GeneRuler™ 50 bp DNA Ladder (Fermentas, St Leon-Rot, Germany); Lanes $n = 1$ to 10, trees analysed during the ring test assay.

protocol: (a) the laboratory E did not proceed the samples on ice before RT-PCR as recommended by the protocol and (b) the second repetition of ACLSV/ApMV duplex protocol made by laboratory B was performed with foreign expired reagents.

Therefore, the results discarded for statistical analysis presented (a) an obvious carry-over contamination or (b) an extreme specificity or sensitivity value correlated with a major deviation from the protocol. So, the second repetition of ASPV/ASGV duplex protocol from laboratories A and B (carry-over contamination) were discarded as well as laboratory E results and the second repetition of ACLSV/ApMV for laboratory B. However, for the ApMV/ACLSV duplex protocol, some laboratories reported the difficulty in determining the exact size of the observed band and correctly discriminating ApMV (312 bp) from ACLSV (290 bp) band.

Specificity and sensitivity

The specificity and the sensitivity values and their 95% confidence intervals were calculated as described in *Materials and methods*. The results are shown in Fig. 2. The specificity values for ApMV, ACLSV and ASGV detection were 93.8%, 96.0% and 95.5%, respectively. The specificity of the protocol for ASPV detection was much lower (67.9%). The calculation of one-sided confidence intervals (95% confidence intervals) showed that their lower limits were higher than 90% for ApMV, ACLSV and ASGV. Whatever the virus, the sensitivity was higher than 95%, ranging from 96.4% (ACLSV) to 100.0% (ASPV). The lower limits of the one-sided confidence intervals for sensitivity ranged between 91.7% (ASGV) and 97.7% (ASPV).

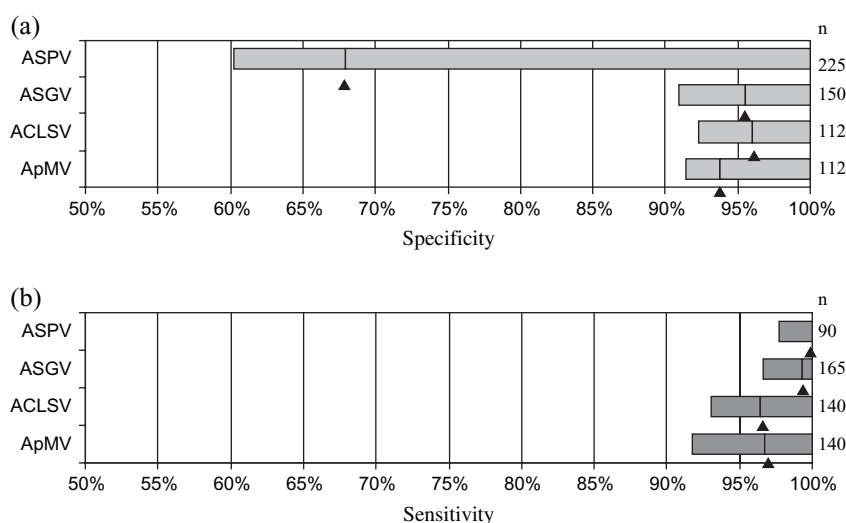


Figure 2 Graphical estimation of the one-sided confidence interval ($\alpha < 5\%$) of specificity (a) and sensitivity (b) for the duplex RT-PCR methods for each virus. □, confidence interval of sensitivity; ■, confidence interval of specificity; ▲, average value; n, number of individual results for each parameter.

Repeatability and reproducibility

The repeatability and reproducibility of the duplex RT-PCR protocol were calculated for each virus as described in *Materials and methods*. The kappa coefficients of Cohen are given in Table 2. Whatever the parameter, a clear difference was observed between ASPV and the three other viruses. For ApMV, ACLSV and ASGV, an almost perfect agreement was observed between all the data (reproducibility) as well as the data obtained within laboratories (average repeatability). For ApMV, an almost perfect agreement was observed within laboratories and a strong agreement between all the data. For ASPV, the reproducibility and the average repeatability indicated only a moderate agreement between data. Differences were observed between the laboratories. A perfect agreement (100%) was observed for laboratories A (ApMV), C (ASGV), D (ApMV, ASGV and ASPV), G (ApMV and ACLSV), H (ApMV and ACLSV) and I (ASGV). Independently of the virus, a weak agreement was obtained by laboratory F, a strong agreement by laboratories C and I

and an almost perfect agreement by laboratories A, D, G and H.

Likelihood ratios and post-test probability

The likelihood ratios and post-test probabilities were calculated as described in *Materials and methods*; results are shown in Table 3 and Fig. 3. The positive likelihood ratios for ACLSV, ApMV and ASGV were higher than 10, corresponding to the best level in the general guidelines and indicating that the test led to large change in pretest to post-test probability when a virus is detected. The positive likelihood ratio for ASPV detection was only 3.1, for example it only led to small change in pretest to post-test probability. The negative likelihood ratios for all the viruses were lower than 0.1, corresponding to the best level in the general guidelines and indicating that the test led to large change in pretest to post-test probability in case of non-detection of the viruses. The likelihood ratio obtained with ASPV (zero) reflected the 100% sensitivity obtained during this test.

Table 2 Kappa coefficient of repeatability and reproducibility (in %) of the duplex RT-PCR detection method

| Virus | Laboratories | | | | | | | | Average Repeatability | Reproducibility |
|-------|--------------|------|-------|-------|------|-------|-------|-------|-----------------------|-----------------|
| | A | B | C | D | F | G | H | I | | |
| ApMV | 100.0 | n.a. | 74.1 | 100.0 | 47.1 | 100.0 | 100.0 | 89.7 | 87.3 | 79.0 |
| ACLSV | 90.4 | n.a. | 71.2 | 90.4 | 53.3 | 90.5 | 100.0 | 90.5 | 83.8 | 85.5 |
| ASGV | n.a. | n.a. | 100.0 | 100.0 | 50.0 | 100.0 | 77.5 | 100.0 | 88.0 | 90.4 |
| ASPV | n.a. | n.a. | 64.9 | 100.0 | 0.0 | 75.0 | 87.0 | 16.9 | 57.3 | 50.6 |

ACLSV, Apple chlorotic leafspot virus; ApMV, Apple mosaic virus; ASGV, Apple stem grooving virus; ASPV, Apple stem pitting virus; n.a., not applicable, one repetition being discarded.

Table 3 Positive and negative likelihood ratios for each virus

| Virus | Positive Likelihood Ratio | Negative Likelihood Ratio |
|-------|---------------------------|---------------------------|
| ACLSV | 24.1 | 0.04 |
| ApMV | 15.6 | 0.04 |
| ASGV | 22.1 | 0.01 |
| ASPV | 3.1 | 0.00 |

ACLSV, Apple chlorotic leafspot virus; ApMV, Apple mosaic virus; ASGV, Apple stem grooving virus; ASPV, Apple stem pitting virus.

The likelihood ratios can further be combined with the pretest probability, for example the incidence of the infection in the analysed population, to obtain the post-test probability of an individual. This can be performed through the Fagan's nomogram (Fig. 4) or a mathematical calculation (*Materials and methods* and Fig. 3). The determination of post-test probability of a negative result with ASGV (from a population with 20% incidence of the infection) using Fagan's nomogram is given in Fig. 4. In Fig. 3, we calculated the post-test probability of infection with a positive (a) or negative (b) result depending on the incidence of each of the four viruses in the studied population. Interestingly, the post-test probability of infection is lower than 1% (5%) if a negative result is obtained with an individual sampled in a population presenting up to 20% (55%) prevalence of the infection, whatever the virus. These suggest that these methods could minimise the risk of releasing infected material when the test result is negative. Additionally, the post-test probability of infection is higher than 90% with a positive result obtained within populations with at least 30–40% prevalence, whatever the virus. With lower prevalence populations, the post-test probabilities of infection with a positive result could be quite low, suggesting the need to reconfirm a positive result by an independent methodology.

Discussion

The routine application of new PCR-based diagnostic techniques requires thorough identification and understanding of the factors that may affect their performance in different laboratories. Through the processing of a common set of samples by nine laboratories using a standardised protocol for RT-PCR amplification, we have addressed some of the issues required to translate two duplex RT-PCR protocols for diagnosis of four fruit tree viruses into practical routine application.

As recommended (Malorny *et al.*, 2003; Banoo *et al.*, 2006), the test evaluation should be performed under the range of conditions in which they are likely to be used in practice. Moreover, samples must be sent in double blind to each laboratory to allow an objective

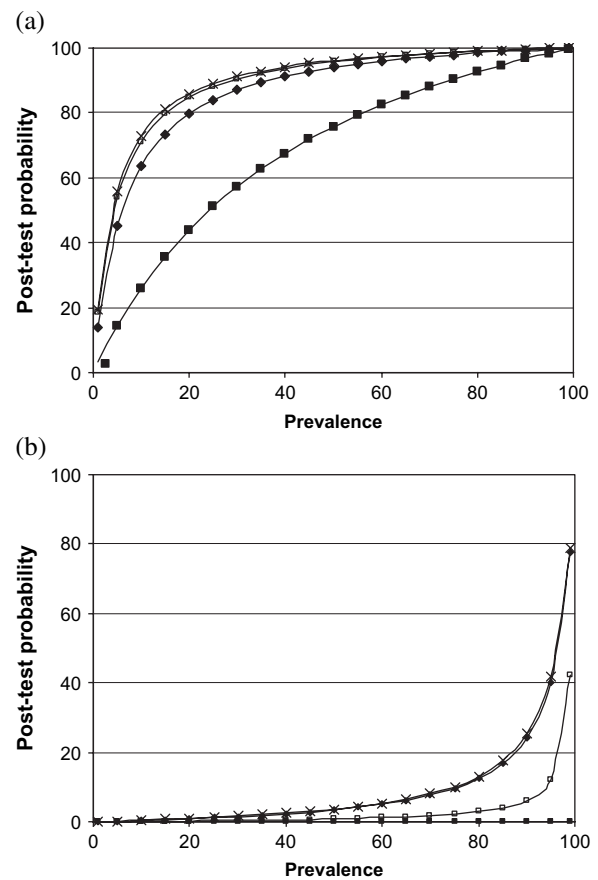


Figure 3 Post-test probability of an individual to be infected after a positive (a) or negative (b) result with the duplex RT-PCR protocol, depending on the infection incidence in the population for \times , Apple chlorotic leafspot virus; \blacklozenge , Apple mosaic virus; \square , Apple stem grooving virus and \blacksquare , Apple stem pitting virus.

interpretation of the results. So, this ring test was designed to mimic practical analysis conditions in a routine certification laboratory. The samples were selected among trees with various origins and degrees of infection. Additionally, they were processed by a simple, quick and reliable preparation method based on crude extracts preparation in an optimised buffer.

The first step in the result analysis is to accurately check if the guidelines have been followed and if special troubles have been encountered. This was carried out through adocumentation of the results including a Microsoft Excel sheet, the gel pictures and a comment page. Additionally, the results were also commented through phone discussions and a meeting between all the participants. All the obtained information was crucial for the adequate result analysis. The preliminary statistical analysis was carried on to identify extreme values. The majority of extreme values were correlated with

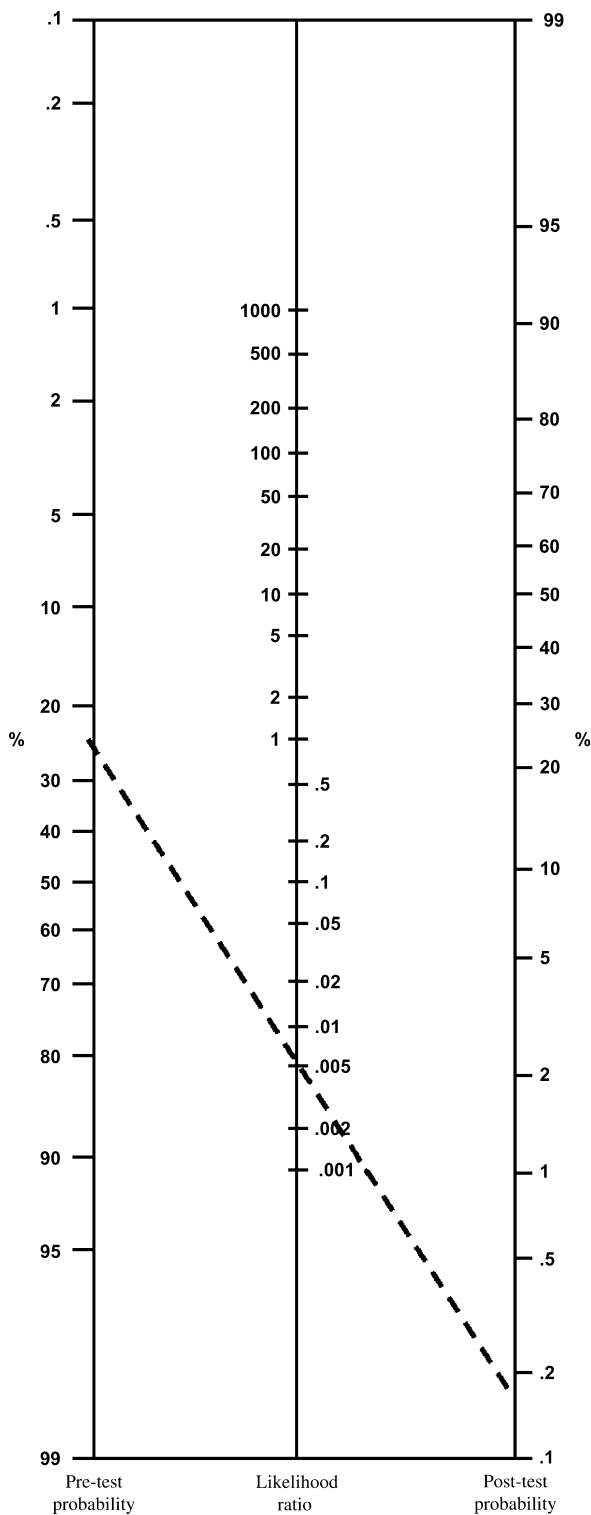


Figure 4 Fagan's nomogram to calculate post-test probability of an individual to be infected depending on the likelihood ratio and the pretest probability, for example the incidence of the infection in the analysed population. Discontinuous line allows calculation of the post-test probability of a negative result for ASGV with a pretest probability of 20%.

a major deviation of the protocol. In conclusion, a careful documentation of the results combined with a preliminary statistical analysis selected the interpretable results for statistical analysis.

Specificity and sensitivity values for ASPV, ASGV and ACLSV were similar or higher than those obtained in the literature for specificity (82–100%) and sensitivity (38–96%) (Malorny *et al.*, 2003; Josefsen *et al.*, 2004; Jeffries & James, 2005; Taha *et al.*, 2005; Lòpez *et al.*, 2006; Truyen *et al.*, 2006). Nevertheless, the specificity value of ASPV detection was low (67.9%), reflecting the high percentage of false positives. Several hypotheses (contamination during sample preparation, reproducibility of primer specificity, etc.) may explain this result. Clearly, this point needs to be addressed during a second ring test to better understand the origin of these results for ASPV. Additionally, it should be recommended to detect separately ApMV and ACLSV infection as many laboratories encountered difficulties to discriminate ApMV and ACLSV bands.

In the diagnostic literature, the significance tests are often two-sided. Nevertheless, we estimated one-sided confidence interval for the different parameters because the first interest of the user of a diagnostic test is the minimum average performance he could expect. So, the confidence interval between a perfect value (100%) and a lower limit is much more informative than the confidence interval between a lower and a higher limit, often different from 100%. So, the choice of a one-sided interval allowed the reallocation of the confidence on the side of interest. The estimation of the confidence limit was performed by Agresti–Coull method, which presents much better properties than the standard Wald interval for proportion (Agresti & Coull, 1998), often proposed in diagnoses testing and known to be biased. The level of confidence of all confidence intervals was fixed at 0.95. The lower limit of the one-sided confidence intervals of sensitivity for all the viruses was higher than 91%, indicating that there is 95% of probability that the real sensitivity of the detection protocols is higher than this value. Similar lower limits, for example higher than 90%, were observed for the specificity of the protocol for ASGV, ApMV and ACLSV.

Repeatability and reproducibility of a diagnostic test are crucial characteristics but, unlike sensitivity and specificity, they are not often taken into account when evaluating the usefulness of a test (Begg, 1987). The reproducibility calculation was based on results obtained with common samples and reagents but in different laboratories with different operators, environments, PCR tubes, pipettes, thermocyclers, etc. Because it takes into account the agreement occurring by chance, kappa coefficients were preferred on simpler accordance or

concordance analysis based on percentage agreement calculation. Repeatability and reproducibility kappa coefficients obtained during this ring test for ASGV, ApMV and ACLSV were very good. Nevertheless, the low repeatability obtained by laboratory F underlined that some troubles may arise in a particular laboratory. Therefore, and as recommended, for example, in the ISO/IEC 17025:2005 guidelines (Anon., 2005), a detailed accompanying protocol, a good preimplementation training, a tight supervision and a continuous quality control of the personnel and the instruments are critical in the adequate use of any diagnostic protocol in routine analysis, even if it has been validated by multicentre trials.

The post-test probability of infection for a sample can be calculated using the positive or negative likelihood ratio and the infection incidence (pretest probability) in the studied population. This post-test probability of infection has a crucial importance in the analysis of the results and in the risk management associated with the use of a determined technique in a certification scheme. Unfortunately, while their calculation is common in human disease diagnostic, the likelihood ratios and post-test probabilities associated with a diagnostic technique have been very scarcely determined for plant virus diagnostic. It is therefore urgent to transfer this methodology in the plant virus diagnostic area to achieve a better risk management of the techniques used in certification schemes. For ACLSV, ApMV and ASGV diagnostic, the analysis of the post-test probabilities associated with a positive or a negative result showed that the evaluated protocols could be integrated in a phytosanitary certification scheme as first screening techniques whose positive results must be confirmed by an independent methodology. Complementary assays are required for ASPV diagnostic.

In conclusion, this paper underlines the usefulness of multicentre trials for new diagnostic protocols and, to our knowledge, publishes for the first time a detailed calculation of post-test probabilities for plant virus diagnostic protocols. Additionally, it also provides useful guidelines and suggestions for researchers desirous of validating their new plant virus diagnostic protocols through interlaboratory evaluation. More particularly, this interlaboratory evaluation of the RT-PCR protocols demonstrated their reliability to detect ACLSV, ASGV and ApMV infection. The interlaboratory evaluation also allowed the development of comprehensive and detailed protocols and underlined that the application of these methods must be performed by well-trained staff within quality control environment. Finally, even if a particular method has been carefully validated, it is still very important to organise periodically ring trials to ensure adequate utilisation of the technique by the laboratories in charge of phytosanitary certification.

Acknowledgement

We thank Frédéric Vansteen and Sylvie Vanmellaert for the excellent sample management and technical assistance. This work has been funded by the General Management of Agriculture (DGA) of the Walloon Region Ministry (Belgium).

References

- Agresti A., Coull B.A. (1998) Approximate is better than "Exact" for interval estimation of binomial proportions. *American Statistician*, **52**, 119–126.
- Altman D., Bland J.M. (1994) Statistics note, diagnostic tests I: sensitivity and specificity. *British Medical Journal*, **308**, 1552.
- Anon. (2005) *ISO/IEC 17025:2005. General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva, Switzerland: International Organization for Standardisation.
- Banoo S., Bell D., Bossuyt P., Herring A., Mabey D., Poole F., Smith P.G., Sriram N., Wongsrichanalai C., Linke R., O'Brien R., Perkins M., Cunningham J., Matsoso P., Nathanson C.M., Olliaro P., Peeling R.W., Ramsay A. (2007) Evaluation of diagnostic tests for infectious infections, general principles. *Nature Reviews Microbiology*, September, S21–S31.
- Begg C.B. (1987) Biases in the assessment of diagnostic tests. *Statistics in Medicine*, **6**, 411–423.
- Bland J.M., Altman D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **1**, 307–310.
- Bootman J., Heath A.B., Hughes P., Holmes H. (1999) An international collaborative study on the detection of an HIV-1 genotype B field isolate by nucleic acid amplification techniques. *Journal of Virological Methods*, **78**, 21–34.
- Cohen J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Fagan T. (1975) Nomogram for Bayes theorem. *The New England Journal of Medicine*, **293**, 257.
- Ferris N.P., King D.P., Reid S.M., Hutchings G.H., Shaw A.E., Paton D.J., Goris N., Haas B., Hoffmann B., Brocchi E., Bugnetti M., Dekker A., De Clercq K. (2006) Foot-and-mouth disease virus: a first inter-laboratory comparison trial to evaluate virus isolation and RT-PCR detection methods. *Veterinary Microbiology*, **117**, 130–140.
- Jeffries C., James C. (2005) Development of an EU protocol for the detection and diagnosis of *Potato spindle tuber potyviroid*. *Bulletin OEPP/EPPO*, **35**, 125–132.
- Josefsen M., Cook N., D'Agostino M., Hansen F., Wagner M., Demnerova K., Heuvelink A.E., Tassios P.T., Lindmark H., Kmet V., Barbanera M., Fach P., Loncarevic S, Hoorfar J. (2004) Validation of a PCR-based method for detection

- of food-borne thermotolerant campylobacters in a multi-center collaborative trial. *Applied and Environmental Microbiology*, **70**, 4379–4383.
- Kummert J., Vendrame M., Steyer S., Lepoivre P. (2000) Development of routine RT-PCR tests for certification of fruit tree multiplication material. *Bulletin OEPP/EPPO*, **30**, 441–448.
- Landis R., Koch G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- Light R.J. (1971) Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychological Bulletin*, **76**, 365–377.
- López M., Llop P., Gorris M., Peñalver J., Donat V., Cambra M., Keck M. (2006) European protocol for diagnosis of *Erwinia amylovora*. *Acta Horticulturae*, **704**, 99–103.
- Malorny B., Tassios P., Radström P., Cook N., Wagner M., Hoorfar J. (2003) Standardization of diagnostic PCR for the detection of foodborne pathogens. *International Journal of Food Microbiology*, **83**, 39–48.
- Massart S., Roussel S., Kummert J., Dutrecq O., Jijakli M.H. (2008) Development of routine duplex RT-PCR tests for certification of fruit tree multiplication material. *Acta Horticulturae*, **781**, 107–112.
- Olmos A., Capote N., Bertolini E., Cambra M. (2007) Molecular diagnostic methods for plant viruses. In *Biotechnology and Plant Disease Management*, pp. 227–249. Eds Z.K. Punja, S. De Boer and H.I. Sanfacon. Wallingford, UK: CABI Publishing.
- Ruelle J., Kabamba Mukadi B., Schutten M., Goubau P. (2004) Quantitative real-time PCR on Lightcycler® for the detection of human immunodeficiency virus type 2 (HIV-2). *Journal of Virological Methods*, **117**, 67–74.
- Taha M.-K., Alonso J.M., Cafferkey M., Caugant D.M., Clarke S.C., Diggle M.A., Fox A., Frosch M., Gray S.J., Guiver M., Heuberger S., Kalmusova J., Kesanopoulos K., Klem A.-M., Kriz P., Marsh S.J., Mölling P., Murphy K., Olcén P., Sanou O., Tzanakaki G., Vogel U. (2005) Interlaboratory comparison of PCR-based identification and genogrouping of *Neisseria meningitidis*. *Journal of Clinical Microbiology*, **43**, 144–149.
- Truyen U., Wilhem S., Genzow M., Schagemann G. (2006) Porcine reproductive and respiratory syndrome virus (PRRSV), a ring test performed in Germany to assess RT-PCR detection methods. *Journal of Veterinary Medicine*, **53**, 68–74.