International Seed Federation
Seed is Life

ISHI-Veg Guidelines for the Validation of Seed Health Methods

Version 3, November 2020

Introduction

The International Seed Health Initiative for Vegetable Crops (ISHI-Veg) supports the delivery of sufficiently healthy seed to customers by working collaboratively to develop methods for seed health testing that are internationally recognized as reference methods and accepted as industry standards. ISHI-Veg primarily focuses on developing methods for pathogenic organisms (pathogens) that are transmitted through seed. Because of their impact on seed trade, ISHI-Veg occasionally develops methods for organisms for which the pathway for introduction and establishment via seed has not been proven.

Method Validation

Validation is a requirement of method development, and it is a process that determines the fitness of a method for its intended purpose. ISHI-Veg has identified six performance criteria to make this assessment and has found them to be in line with those used by other accreditation bodies for developing seed health methods.

The scope of the method and its validation criteria should be properly defined and documented in a validation plan before the experimental studies commence. The minimum requirements set for each performance characteristic should also be established and justified. The assistance of a statistician when designing the experimental studies will ensure that the experiments proposed in the validation plan are statistically sound and appropriate for the analysis.

Data published in scientific and peer reviewed journals may be used in method validation if they assist in demonstrating that the method is fit for purpose. Data generated during the developmental phase of the method may also be used.

Factors, such as the number of target pathogen strains, availability of infected material in the case of non-culturable, new or emerging pathogens, that limit the ability to validate the method as required should be explained.

After a method has been developed and is in use, ISHI-Veg encourages labs to continue monitoring its performance and report findings that indicate it is underperforming. The corresponding group (ITG) in ISHI-Veg responsible for the method will review these reports before initiating method improvement.

Guidelines on Measuring Method Performance

Performance Characteristic	Definition
Analytical specificity of an assay	The ability of an assay to detect the target(s) pathogens (inclusivity) while excluding non-targets (exclusivity)
Analytical specificity is evaluated using pure cultures of strains/isolates/variants. When a pathogen cannot be cultured, alternatives such as infected host material or synthetic DNA/RNA representing the target of interest can be used. An assay should be evaluated with a collection of strains that represent the known phenotypic and genotypic diversity of the pathogen. In addition, the assay should be evaluated with a diverse, characterized collection of non–target strains/isolates associated with the host seed, such as those originating from the same host and location, and related species. The recommended size of a characterized collection is 20-30 isolates each of targets and non-targets. *Note: For nucleic acid-based assays in silico evaluation is recommended and should complement assay validation.*	
Analytical sensitivity	Smallest amount of the target pathogen that can be detected i.e. the limit of detection (LOD)
Analytical sensitivity is evaluated using a replicated dilution series of pure cultures of strains/ isolates/variants of the pathogen in the assay volume. Seed samples should be spiked with a dilution series, and the assay performed as described to give a positive or negative result. The number of replicates used gives an estimation of LOD at the desired detection rate of 95%. Dilution series should have a maximum dilution factor of 10x per step and consist of minimally 3 steps. The dilution with the smallest amount of target pathogen which can consistently be detected with 95% confidence is the LOD of the assay. The most diluted step should have less than 95% detection. A minimum of 3 independent dilution series with 3 replicates each should be tested. For some assays quantification of the target pathogen may be complicated due to its non-culturable state, or the inability to isolate it. In such cases sensitivity can be established by spiking infected seeds in healthy seed backgrounds. As the aim of a pre-screen is to identify seed lots that are not infected with the target pathogen(s), the LOD of the pre-screen assay should be lower than the LOD of the confirmatory detection assay [1]. *Note: For identification assays, sufficient dilutions have to be tested to ensure that the dilution used in the assay is fit for purpose.*	

[1] ISF (2018). Real-time PCR, an indirect test used for pre-screening in seed health methods. https://www.worldseed.org/our-work/phytosanitary-matters/seed-health/ishi-veg-method-development/

Performance Characteristic	Definition
Selectivity	The effect of different seed matrices on the ability of the method to detect target pathogen(s)
Selectivity is evaluated by analyzing the effect of matrix variations by spiking a known concentration of the target pathogen in a relevant collection of differing matrices. Seed matrix variation can result from differences in crop species or varieties, and variations in saprophytic levels, seed treatments and the geographical origin of the seed. Different seed matrices spiked with pure pathogen, or infected seeds (ground or whole) or leaf material should be tested. A minimum of 2 variables (species, varieties, locations, etc.) should be tested per matrix but a higher number is recommended. A suitable concentration of the target pathogen should be determined based on developmental work and according to spiking recommendations described in ISHI-Veg best practices. *Note: The tolerance level for variations in the detection of the pathogen caused by the seed matrix may vary per method and should be explained.*	
Repeatability	Degree of similarity in results of replicates of the same seed lots when the method is performed with minimal variations in a single lab
Repeatability of a method is evaluated by the same technician assaying replicates of seed samples (infected and healthy) using the same reagents and equipment. Replicates should be drawn from seed lots with a range of infection levels including one that is just above the LOD. Replicates should be uniformly infected. At least three replicates per seed lot should be evaluated. *Statistical analysis* For qualitative results, the repeatability of the method can be estimated by calculating accordance [2]. This is based on the probability of finding the same test results for identical test materials within laboratories. Accordance of at least 90% is considered to indicate good repeatability [3]. For quantitative results, ANOVA, or regression analysis can be considered. Assistance of a statistician for the analysis is highly encouraged. *Note 1: Use well-characterized seed lots* *Note 2: If infected lots are not available, (sub)samples of healthy lots can be spiked with infected seed.*	

[2] Langton, S.D., Chevennement, R., Nagelkerke, N. and Lombard, B. (2002). Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. International Journal of Food Microbiology, 79, 175-181.

[3] ISTA (2019). Procedure 'validation methods and organizing and analyzing results of interlaboratory comparative tests (CT)'. Version 1.0. https://www.seedtest.org/upload/cms/user/TCOM-P-10-ValidatingmethodsandresultsofCTsV1.141.pdf. Last checked on 29-07-2020.

Performance Characteristic	Definition
Reproducibility	Degree of similarity in results when the method is performed across labs with replicate seed subsamples

Reproducibility of a method is evaluated in a comparative test (CT). Data generated from the CT is used to establish reproducibility of the method and to provide quantification of variation in test performance under different laboratory conditions.

For validation, methods should be performed across experienced labs on replicate seed subsamples. These subsamples should be drawn from seed lots with a range of infection levels including one that is just above LOD. Replicates should be infected as uniformly as possible which can be assessed by homogeneity testing prior to the CT.

Statistical analysis

For qualitative results, the reproducibility of the method can be estimated by calculating concordance [2]. This is based on the probability of finding the same test results for identical test materials between laboratories. Concordance of at least 90% is considered to indicate good reproducibility [3].

The concordance odds ratio (COR) is an indicator of the relationship between accordance and concordance, or in other words, the presence or absence of differences between laboratories. Ideally, the estimate for the COR should be 1.0; higher values indicate that the method does not provide consistent results between laboratories. For each method an acceptable COR value should be determined above which the method will not be adopted as an ISHI-method. For quantitative results, ANOVA, or regression analysis can be considered.

Assistance of a statistician for the analysis is highly encouraged.

Note 1: For a CT it is important to check stability of the pathogen in the seed since various factors such as time and conditions during shipment may have an effect on the pathogen and therefore on the CT test results.

Note 2: If infected lots are not available, healthy lots can be spiked with infected seed. Samples for testing can then be drawn from this seed.

Diagnostic performance	An evaluation of the ability of the method to discriminate between positive and negative seed lots

To assess diagnostic performance the results of the method should be compared to the true disease status of the seed lot. Ideally this should be done with an independent assessment of the disease status using a reference method or an existing validated test.

Diagnostic sensitivity and diagnostic specificity are measures of the test's ability to discriminate between the presence and absence of the target pathogen. Diagnostic sensitivity is the test's capacity to give a positive result when the pathogen is present, while diagnostic specificity is a measure of how certain a negative result is a true negative. Evaluating these

Performance Characteristic	Definition

metrics as accurately as possible is essential, as they are measures of confidence in the test results [4].

The two measures of diagnostic performance are obtained by a calculation based on true and false results (see below).

		Validated method result / independent assessment	
		Positive	**Negative**
Test outcome	**Positive**	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)
		DIAGNOSTIC SENSITIVITY = (TP / (TP + FN)) x 100%	**DIAGNOSTIC SPECIFICITY = (TN / (FP + TN)) x 100%**

To obtain a good estimate of diagnostic performance, a statistician should be consulted when designing the experimental studies on validation on the required number of samples to be tested. They should be drawn from relevant origins, production years and infection levels among others that have been listed, and selected based on the most important sources of variation.

Note 1: If infected lots are not available, (sub)samples of healthy lots can be spiked with infected seed.

Note 2: When a reference standard to which the performance of the method can be compared is not available, the method developer must establish a reference that provides an independent assessment of the true health status of lots.

Note 3: An alternative way to assess diagnostic performance is to evaluate samples in a comparative test. The data generated by a large number of labs includes estimates of laboratory competence and a measure of robustness of method performance. Robustness demonstrates how precise the method is when subjected to variable factors, a normal occurrence when the method is used by different personnel of different labs.

[4] Cardwell, K., Dennis, G., Flannery, A.R., Fletcher, J., Luster, D., Nakhla, M., Rice, A., Shiel, P., Stack, J., Walsh, C. and L. Levy. 2018. Principles of Diagnostic Assay Validation for Plant Pathogens: A Basic Review of Concepts. doi:10.1094/PHP-06-18-0036-RV.

Version History

Version	Date	Changes (minor editorial changes not indicated)
1	May 2018	Performance characteristics identified
2	May 2020	*Analytical sensitivity:* experimental set-up, definition and requirements of LOD more clearly defined. *Repeatability:* statistical analysis added (quantitative and qualitative). *Reproducibility:* statistical analysis added (quantitative and qualitative) and homogeneity testing. *Diagnostic performance:* editorial change of note 1, note 2 adjusted from "either a CT or single lab testing" to "CT data can be added". *Post-implementation surveillance:* removed as a performance characteristic and added general text to the introduction encouraging labs to report less than optimal performance of a method. Introductory text modified to include • The focus of ISHI-Veg methods. • The scope of the method and performance requirements for each criterium to be part of the validation plan. • An encouragement to consult the statistician.
3	November 2020	*Analytical specificity:* repetition from introduction about the inclusion of a justification of a deviation deleted. *Analytical sensitivity:* reference added to explain the term 'pre-screen assay'. *Diagnostic performance:* definition adjusted, minimum seed sample size to be tested removed, diagnostic sensitivity and specificity formulas added, a new note added (note 2) and one note re-written (note 3). Version history included.