



**Performance Data Evaluation for the presence of seed with  
specified trait(s) in seed lots**

Version 2.1

**Note: Any electronic or hard copies of this document are uncontrolled copies / Confirm validity  
before use**

# Performance Data Evaluation for the presence of seed with specified trait(s) in seed lots

## 1 SUMMARY

This document describes how to present and evaluate performance data in the context of laboratory accreditation for determination of specified trait(s) under the Performance Based Approach.

This is related to Chapter 8 of the ISTA Rules, for the detection of seeds with specified trait(s) in seed lots, as no standardised method is included in the Rules for this type of test.

The laboratory shall specify the trait(s), and give information on the implemented method that it will use to test seeds.

The samples used must be seeds.

The laboratory shall show evidence of the ability to detect and/or to quantify the specified trait(s), when present, in a range from 0.1% to 3%.

## 2 GENERAL INFORMATION ON TRAIT(S) AND METHOD

The specified trait(s) shall be defined by the laboratory (e.g., GTS-40-3-2, Bt11, 35S promoter, NOS terminator, Cry3Bb1 protein, etc).

Specified trait(s) as described by the laboratory	Comments

The method used shall be described by the laboratory in sufficient detail, to allow auditors to understand how the results are obtained.

The method is chosen by the laboratory (e.g., lateral flow strip, grow out assay of seedlings, sub-sampling strategy, quantitative PCR method on ground flour, etc).

Description of the type of method used by the laboratory	Comments

Statement made by the lab about the way the method was validated and installed	Comments

Reference on documentation available, or other parties to contact, to be able to get information on validation of the method and its fitness for the purpose	Comments

### 3 SEEDS TO PREPARE THE SAMPLES

The samples used must be seeds.

In case the seeds are ground, the ability to obtain homogeneous and fine flour from seeds is one of the key elements to obtain a result which is representative from the sample.

In case the seeds are germinated, the percentage of germination shall be taken into account to obtain enough plants.

Samples are prepared from 2 sources of seeds which shall be in principle 100% seeds with absence of specified trait(s) on one hand, and 100% seeds with the specified trait(s) on the other hand.

ISTA recommends a 400 seeds check for the source of seed with specified trait(s) and the check of 30 000 seeds for the source of seed with absence of the specified trait(s).

ISTA recommends an individual seed test check for the source of seed with specified trait(s) with no absence of trait as a result. In that case the lower bound of purity of the seeds with the trait(s) is 99.25% with 95% confidence.

ISTA recommends pools size of no more than 1000 seeds for the source of seed with absence of the specified trait(s), with no presence as a result. In that case the lower bound of purity of the seeds with no trait(s) is 99.99%.

The fact that the 2 sources are pure enough shall be assessed by the laboratory. This can be done by a check as recommended above, or by other means at the initiative of the laboratory.

Statement of the laboratory on the assessment of purity of seed source with presence of specified trait(s)	Comments

Statement of the laboratory on the assessment of purity of seed source with absence of specified trait(s)	Comments

Approximate minimum number of seeds to gather to obtain enough seeds to prepare the different samples:

	<b>Seeds with trait(s)</b>	<b>Seeds without trait(s)</b>	<i>step</i>	<i>comments</i>
<i>Ability to check for presence/absence of specified trait(s)</i>	50	11 950	<i>ability to detect presence/absence of the specified trait(s)</i>	30 samples of 400 seeds
<i>Ability to check for quantification of presence of specified trait(s)</i>	636	55 584	<i>ability to quantify the specified trait(s)</i>	28 samples of 2000 seeds
<i>Approx. number of seeds</i>	700	67600		

**4 ABILITY TO DETECT PRESENCE AND ABSENCE OF THE SPECIFIED TRAIT(S)**

(30 samples of 400 seeds)

30 samples shall be prepared to check the ability to detect presence/absence of the specified trait(s). They shall be randomly coded by a person from 1 to 30, and given blind to the staff who will perform the tests for presence/absence.

Statement of the laboratory about blindness of the 30 samples

The person who prepared the blind samples shall report the results.

- 10 samples of 400 seeds are to be prepared in which 3 seeds contain the specified trait(s) and 397 do not contain this trait(s) (spiking level 0.75%). The seeds with absence of trait(s) can be counted (397 seeds), or if counting is not possible obtained by weight of seeds based on the 1000 seed weight.

The laboratory shall indicate the result for each sample, whether the method showed presence of the trait(s) (P), or absence (A)

sample	1	2	3	4	5	6	7	8	9	10
result										

10 P are expected. If this is not the case, the laboratory shall comment on its results.

Comments

- 10 samples of 400 seeds are prepared in which 2 seeds contain the specified trait(s) and 398 do not contain this trait(s) (spiking level 0.5%). The seeds with absence of trait(s) can be counted (398 seeds), or if counting is not possible obtained by weight of seeds based on the 1000 seed weight.

The laboratory shall indicate the result for each sample, whether the method showed presence of the trait(s) (P), or absence (A)

sample	1	2	3	4	5	6	7	8	9	10
result										

10 P are expected. If this is not the case, the laboratory shall comment on its results.

Comments

- 10 samples of 400 seeds where the trait(s) is absent in each seed are needed. The seeds with absence of trait(s) can be counted (400 seeds), or if counting is not possible obtained by weight of seeds based on the 1000 seed weight.

The laboratory shall indicate the result for each sample, whether the method showed presence of the trait(s) (P), or absence (A)

sample	1	2	3	4	5	6	7	8	9	10
result										

10 A are expected. If this is not the case, the laboratory shall comment on its results.

Comments

If all 30 samples are true to type, the upper limit of a one-sided 95% confidence interval of the false rate (including both false positive and false negative) is 9.5%.

About 300 samples would be necessary to obtain a 95% confidence interval with 1% as the upper boundary of false rate. Given the size and required workload of many laboratories, it is impractical to require this ten-fold increase in the number of samples to achieve a 1% upper boundary of false rate. Therefore 30 samples are considered adequate.

## 5 QUANTIFICATION OF SPECIFIED TRAIT(S)

The laboratory shall state the quantification result in %, in the unit in which it will report its results. If more than one unit is to be used, the laboratory shall either show evidence for each unit chosen, or indicate how a result with a given unit can be derived from a result from a unit for which evidence has been shown. In any case the unit chosen for reporting the quantification result must tally with the unit for the spiking levels in order to allow a comparison.

Data may be provided in % seeds by number of seeds, in % seeds by mass of seeds, or in another unit, such as % DNA copies.

Statement of the laboratory on the unit used for the performance data, and reporting of results, if % of seeds by number of seeds	Comments

Statement of the laboratory on the unit used for the performance data, and reporting of results, if % of seeds by weight of seeds	Comments

Statement of the laboratory on the unit used for the performance data, and reporting of results, if % of seeds by the unit described below (% DNA copies for example)	Comments

The performance data evaluation aims at checking % results in the range from 0.1% to 3%.

Samples are all made up of 2000 seeds in total (by count, or by weight based on 1000 seed weight).

When a number of seeds with the specified trait(s) are spiked in the samples, the weight of seeds with the specified trait(s), and the weight of the seeds where the trait(s) is not present, shall be weighed for each sample.

For each of 7 levels, 4 samples of 2000 seeds shall be prepared, making 28 samples in total.

3 known levels of presence shall be checked: 0.1% (2 seeds spiked), 0.5% (10 seeds spiked), 1% (20 seeds spiked) by each laboratory.

4 levels of presence shall be selected by the person who prepares the sample, to be given blind to the testers.

- Blind level 1 shall be in the interval [0.1%, 0.5%[ in number of seeds (2 to 9 seeds spiked)
- Blind level 2 shall be in the interval [0.5%, 1%[ in number of seeds (10 to 19 seeds spiked)
- Blind level 3 shall be in the interval [1%, 2 %[ in number of seeds (20 to 39 seeds spiked)
- Blind level 4 shall be in the interval [2 %, 3%] in number of seeds (40 to 60 seeds spiked)

For each of the blind levels, the person who prepares the samples shall select at random a number of seeds within the range, and prepare 4 samples with this same number of seeds having the specified trait(s).

*The number of seeds required for the 28 samples is about:*

	Seeds with specified trait(s)	Seeds without specified trait(s)
0.1% level	$2 \times 4 = 8$	$1998 \times 4 = 7992$
0.5% level	$10 \times 4 = 40$	$1990 \times 4 = 7960$
1% level	$20 \times 4 = 80$	$1980 \times 4 = 7920$
Blind level 1	$9 \times 4 = 36 \text{ max}$	$1998 \times 4 = 7992 \text{ max}$
Blind level 2	$19 \times 4 = 76 \text{ max}$	$1990 \times 4 = 7960 \text{ max}$
Blind level 3	$39 \times 4 = 156 \text{ max}$	$1980 \times 4 = 7920 \text{ max}$
Blind level 4	$60 \times 4 = 240 \text{ max}$	$1960 \times 4 = 7840 \text{ max}$
TOTAL	636	55 584

The 28 samples for the ability to quantify the specified trait(s) shall be randomly coded from 1 to 28 and given blind of level to the staff that will perform the tests.

Statement of the laboratory about blindness of the 28 samples

**6 DESCRIPTION OF SAMPLES, AND REPORT OF RESULTS**

		0.1% level	blind level 1	0.5% level	blind level 2	1% level	blind level 3	blind level 4
Spiking level (% number of seeds)		0.1		0.5		1.0		
number of spiked seeds	sample 1							
number of spiked seeds	sample 2							
number of spiked seeds	sample 3							
number of spiked seeds	sample 4							
weight of spiked seeds	sample 1							
weight of spiked seeds	sample 2							
weight of spiked seeds	sample 3							
weight of spiked seeds	sample 4							
weight of seeds without trait(s)	sample 1							
weight of seeds without trait(s)	sample 2							
weight of seeds without trait(s)	sample 3							
weight of seeds without trait(s)	sample 4							
% other unit	sample 1							
% other unit	sample 2							
% other unit	sample 3							
% other unit	sample 4							

**Results**

result in % number of seeds	sample 1							
result in % number of seeds	sample 2							
result in % number of seeds	sample 3							
result in % number of seeds	sample 4							
result in % weight of seeds	sample 1							
result in % weight of seeds	sample 2							
result in % weight of seeds	sample 3							
result in % weight of seeds	sample 4							
result in % other unit*	sample 1							
result in % other unit*	sample 2							
result in % other unit*	sample 3							
result in % other unit*	sample 4							

\* The unit must correspond to the % other unit chosen above.

It is not permitted to give a table of results with missing samples. In case of problems with a sample, another sample of the same level shall be prepared and analysed.



The values given by the staff that performs the analysis shall be reported without any change by the person who prepared the coded samples (no retest, no re-evaluation if a value is far from the true value, etc).

Statement of the laboratory on the fact that the results obtained have been reported with no changes

Statement of the laboratory about how to derive a unit from another, if applicable

## 7 STATISTICAL EVALUATION OF PERFORMANCE DATA

Accuracy and repeatability shall be determined as an objective criteria to evaluate the performance data (for details please refer to the ISTA Handbook on Statistics in Seed Testing).

Accuracy is a way to measure closeness of agreement between a test result and the true value.

Repeatability is a measure of dispersion of test results under repeatable conditions. It will be computed as the square-root of the average of the variances of the 4 samples per level and will be expressed in % of the mean of the true levels in the test.

The four samples per level have exactly the same true value expressed as % in number of seeds if the number of seeds is counted (spiked seeds with specified trait(s) and seeds). In other cases, the true value of the 4 samples is very similar but not strictly equal. The 4 samples are not 4 repeats from a unique sample, as it might be the case in other types of repeatability assessments.

*Example for computation of repeatability std-dev in % of the mean:*

True level	Sample	Result	Variance
0.1%	sample 1	0.12	0.003825
	sample 2	0.24	
	sample 3	0.11	
	sample 4	0.12	
Blind-0.3%	sample 1	0.3	0.00046667
	sample 2	0.34	
	sample 3	0.29	
	sample 4	0.31	
0.5%	sample 1	0.52	0.003
	sample 2	0.51	
	sample 3	0.49	
	sample 4	0.4	
Blind-0.6%	sample 1	0.6	0.001225
	sample 2	0.65	
	sample 3	0.68	

	sample 4	0.62	
1.0%	sample 1	1.1	0.01666667
	sample 2	1	
	sample 3	1.2	
	sample 4	1.3	
Blind-1.5%	sample 1	1.5	0.01666667
	sample 2	1.6	
	sample 3	1.7	
	sample 4	1.8	
Blind-2.2%	sample 1	2.5	0.04333333
	sample 2	2.6	
	sample 3	2.8	
	sample 4	2.3	
<b>Average variance (repeatability variance)</b>			0.01216905
<b>Mean of the true levels</b>			0.885714
<b>Repeatability std-dev in % of the mean</b>			<b>12.45%</b>

## 8 EVALUATION OF THE DATA SUBMITTED BY THE LABORATORY

The performance data are a part of the different elements of assessment for accreditation. The grades given below to evaluate the performance data are a help for the auditors.

### 8.1 Ability to detect the trait(s):

Grade 1: all 30 samples correctly identified

Grade 2: 1 or 2 samples wrongly identified

Grade 3: more than 2 samples wrongly identified

### 8.2 Accuracy of trait(s) quantification:

In some documents +/- 20% or +/- 25% deviation from true value is stated as the maximum tolerable limit of accuracy. For low % values, in our case a range of 0.1%-3%, the distribution of data points is not symmetrical, and the difference (true value) - (quantile) is smaller below the true value compared to the difference with the same quantile above the true value. For this reason, the thresholds are established with an upper interval which is twice the lower interval. For instance the lower limit is 25% of the true value, and the upper limit is 50% of the true value.

For a true value of 1%, limits are then -0.25 and +0.5, where the accuracy is

$[(\text{observed value}) - (\text{true value})] / \text{true value}$

Grade 1: all 28 samples have the accuracy within -0.25 and +0.5 (inclusive of boundaries)

Grade 2: no samples have the accuracy smaller than -0.5 or greater than 1

Grade 3: one or more samples have the accuracy smaller than -0.5 or greater than 1

### 8.3 Repeatability of trait(s) quantification:

Grade 1: Repeatability std-dev in % of the mean is below 20%

Grade 2: Repeatability std-dev in % of the mean is below 30%

Grade 3: Repeatability std-dev in % of the mean is above 30%

## **9 CONCLUSION**

Auditors will use the information provided in the performance data evaluation document submitted by the laboratory, and will perform computations with the data provided.

Performance data evaluation is a part of the accreditation scheme when there is no ISTA method available in the ISTA Rules.