



International Seed Testing Association

# **How does ISTA evaluate the performance of the laboratories in the proficiency tests ?**

ISTA GMO TF  
S GREGOIRE - C HALDEMANN



# Outline

- Organisation
- Presence/absence rating
- Quantification rating
- Further developments

# ORGANISATION



GMO TF

ISTA  
SECRETARIAT

STA-GMO TF

Documentation

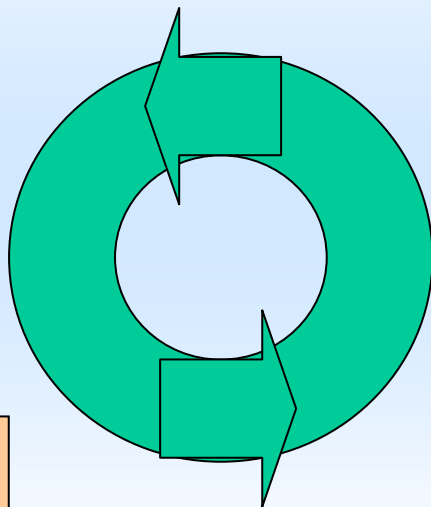
Strategic  
planning

several PTs  
evaluation

PT planning

Single PT evaluation

PT running



# Rating system

- Is in place
- Is consistent with system in place for other type of seed tests (A,B,C,BMP)
- Has already be shown at Bangkok meeting
- Has been described in Seed Testing International issues



# The rating Steps

laboratory identification is always kept confidential

- Coding, Checking result sheets, (question lab. if needed)
- Technical checks
- Statistical analysis
- Preliminary rating

• Official rating

1 GMOTF Technical expert

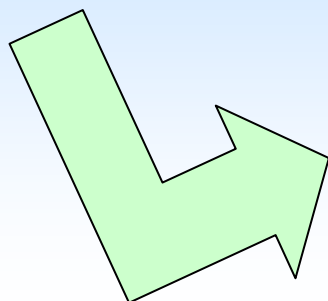
2 STA-GMOTF Statistical expert

3 ISTA Accreditation auditor

4 ISTA Test Organiser

4 experts

and ISTA Secretary General (observer)





# Decoding of rates = as for other proficiency tests

- A No problem has been detected in this test
- B There are small problems, but no specific look or action is suggested to the participant
- C Problems, ISTA indicates there might be things to consider by the laboratory to explain or correct things
- BMP Below Minimum of Performance, ISTA indicates that the results were poor and the laboratory need to find explanations and to improve/correct



# Rating as for other proficiency tests

- Rating is a general feature of ISTA proficiency tests, provided by each technical committee through the test leader and ISTA secretariat, with the assistance of the proficiency test committee and statistics committee

**The current ISTA system rates :**

*and a run of 6 tests*

<b>One test rating</b>	<b>One test Score Value</b>	<b>Overall rating on 6 tests</b>	<b>Range on 6 tests</b>
A	5 points	A	28 – 30 points
B	4 points	B	21 – 27 points
C	3 points	C	16 – 20 points
BMP	0 points	BMP	below 16 points

*Example 1:*

*5A rating and 1 BMP (Cotton)*

*5\*5points +  
1\*0point =25 points*

*Overall rating is B*

*Example 2:*

*4 B = 16 points*



# Presence/absence: 3 rating systems (RS) are available

Rate	Number of missclassified samples
<b>A</b>	<b>0 errors</b>
<b>B</b>	<b>1 or 2 errors</b>
<b>C</b>	<b>3 errors</b>
<b>BMP</b>	<b>more than 3 errors</b>

**Rating System 1: Errors are computed as absolute number of missclassifications**



# Presence/absence: 3 rating systems (RS) are available

Rate	Percent of missclassified samples
A	from 0% to 5% ]
B	>5% to 10%]
C	>10% to 20%]
BMP	more than 20%

**Rating System 2: Errors are computed as percentage of missclassifications**



# Presence/absence: 3 rating systems (RS) are available

Rate	Percent of missclassified samples
A	from 0% to 6% ]
B	>6% to 20%]
C	>20% to 30%]
BMP	more than 30%

**Rating System 3: Errors are computed as percentage of missclassifications (less stringent than RS2)**



# Presence/absence: 3 rating systems are available

**Example: Impact of the three rating systems showed on a Proficiency Test including 12 samples.**

	0 out of 12 <sup>1)</sup>	1 out of 12	2 out of 12	3 out of 12	4 out of 12
<b>RS 1</b>	<b>A</b>	<b>B</b>	<b>B</b>	<b>C</b>	<b>BMP</b>
<b>RS 2</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>BMP</b>	<b>BMP</b>
<b>RS 3</b>	<b>A</b>	<b>B</b>	<b>B</b>	<b>C</b>	<b>BMP</b>

<sup>1)</sup> 0 missclassification out of 12 samples

**This table shows the rating results according to chosen rating system and the number of missclassification(s).**

Rating systems allows to account for 10 to 30 samples, and the level of difficulty in a given Proficiency Test



# Quantification: 3 Rating Systems are available

- System 1: true value is number of GM seeds/total number of seeds in each sample
- System 2: true value is weight of GM seeds/total weight of seeds in each sample
- System 3: true value is the median of values reported by participating laboratories (after Cochran test)

*Computations are performed from all sample results returned by the laboratories whatever the method (including sub-sampling and quantitative PCR )*



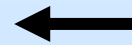
# Quantification: Rating system depends on unit of results in report sheet

System 1: true value is number of GM seeds/total number of seeds in each sample



Result reported in % number

System 2: true value is weight of GM seeds/total weight of seeds in each sample



Result reported in % mass, or %DNA copies



System 3: true value is the median of values reported by participating laboratories



Result reported in %DNA copies



# Rating quantification

- *True value is known (purity check + event + number + weight), and/or can be estimated from all results received*

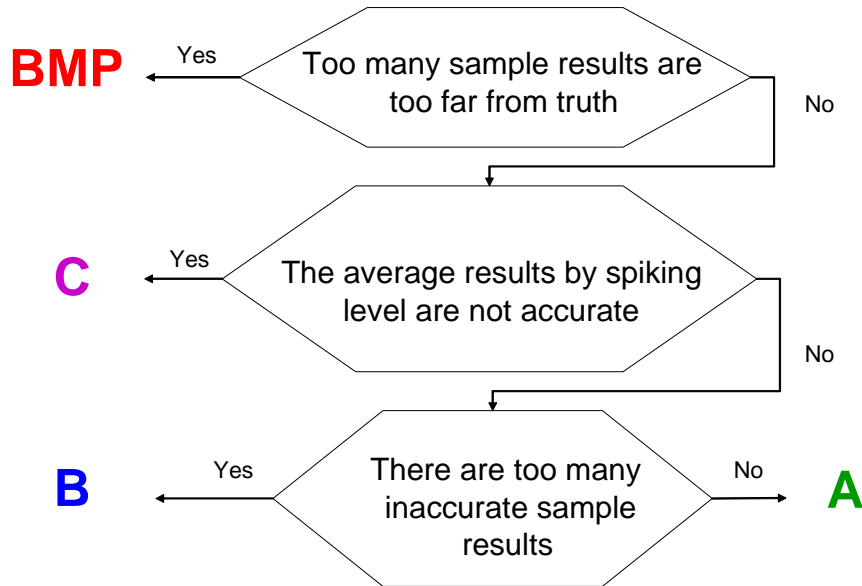
- Too many *sample results* are too far from truth =>BMP
- *Average results* by spiking level are not accurate =>C
- Too many inaccurate *sample results* =>B
- Otherwise =>A

*NB: Computations are made on GM spiked samples*

GMO TF Session, ISTA Annual Meeting, Zurich (Switzerland) 26.-29.06.2006

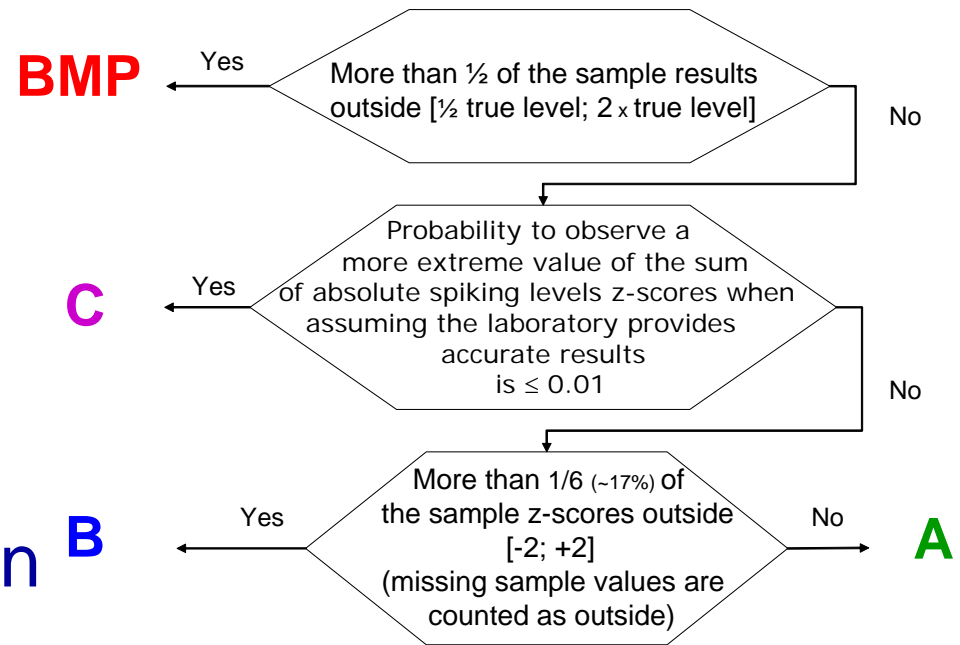


# Rating quantification



Principles

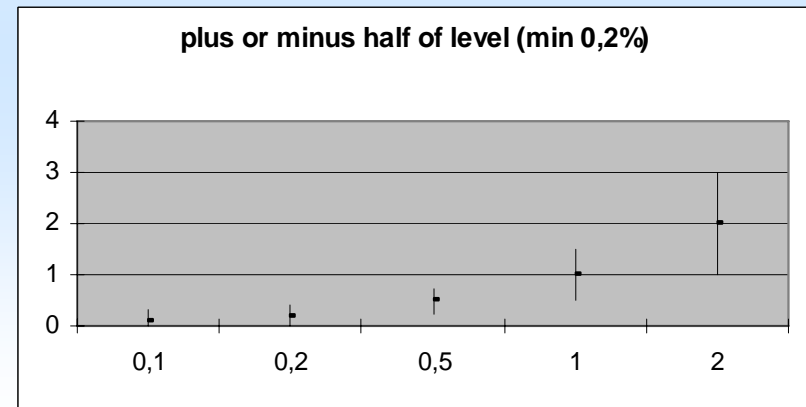
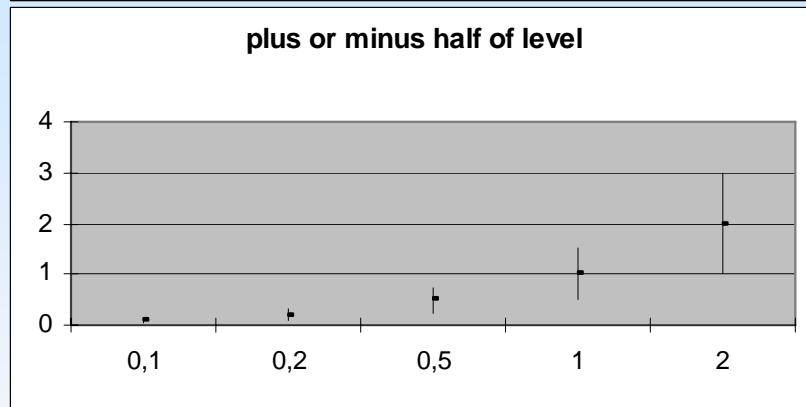
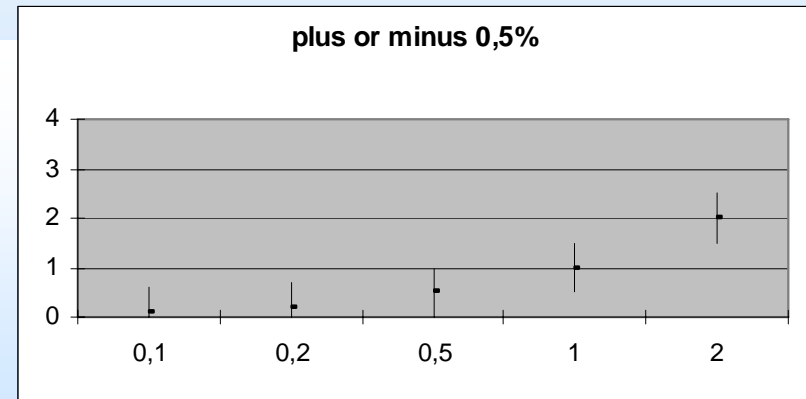
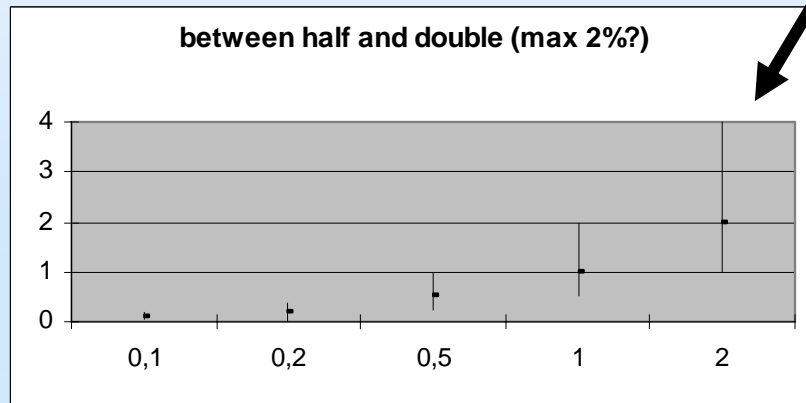
Implementation





**BMP = If too many sample results (« many » is x% of data points) are too far from truth (4 studied criteria are shown below)**

**Selected criterium: More than half of the results are outside half and double of truth**



*Within vertical bar = not too far from truth*



C = If the average results (from 2-6 samples) by spiking level are not accurate

- Mean of sample results per spiking level is computed
- A « *spiking level z\_score* » is computed for each spiking level.
- A sum of absolute « *spiking level z\_scores* » for the different spiking levels is compared to a statistical threshold



## B = There are too many inaccurate sample results

- Individual « *sample z\_scores* » are computed
- The number of « big » absolute « *sample z\_scores* » is counted
- If number of big « *sample z\_scores* » exceeds a pre-defined limit then B rating, otherwise A rating

Number of <i>non zero level</i> samples in the proficiency test	Maximum tolerated number of <i>sample z_scores</i> out of the interval [-2,+2]
1 to 5	0
6 to 11	1
12 to 17	2
18 to 23	3
24 to 29	4

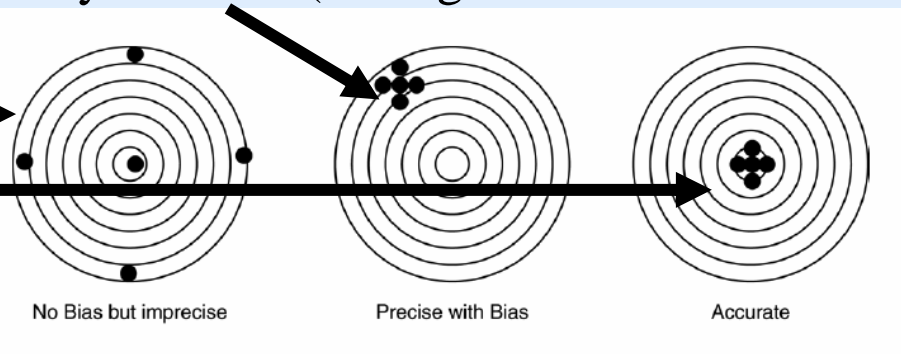
More than 1/6th are inaccurate => rate B

Illustration on a spiking level : spotted by C check (average is far from truth=>C)

B  
check

Not  
OK=> B  
  
OK=> A

Same average





# Z-scores are a deviation from a reference value

## Z-scores

Z-scores are useful to establish rules from distributions with different means and/or different standard-deviations. They are already used in ISTA PT rating systems, on Germination and Purity for example.

The definition of a z-score is as follows:

Consider a value  $x$  from a distribution with mean  $\mu$  and standard-deviation  $\sigma$ . The formula for converting  $x$  into its corresponding z-score is:

$$z = \frac{x - \mu}{\sigma}$$

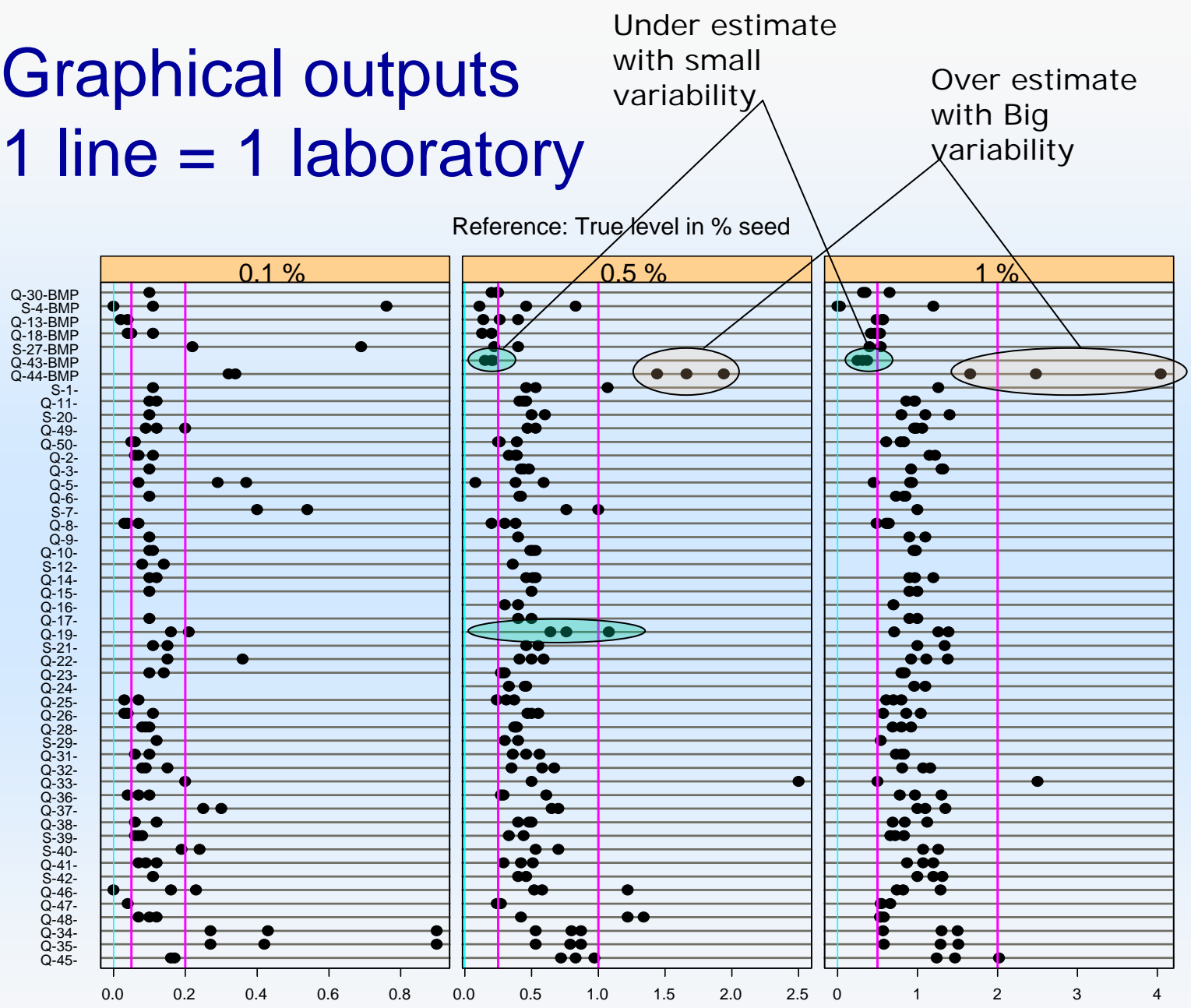
This value indicates how far and in what direction  $x$  deviates from  $\mu$ , in units of  $\sigma$ . The distribution's mean and the distribution's standard-deviation of the z-scores are equal to 0 and 1 respectively. When the distribution of reference is normal, the z-scores distribution is also normal and thus the probability to have a z-score in the interval  $[-2; +2]$  is approximately 0.95. This property is used in the PT rating.



# Graphical outputs

## 1 line = 1 laboratory

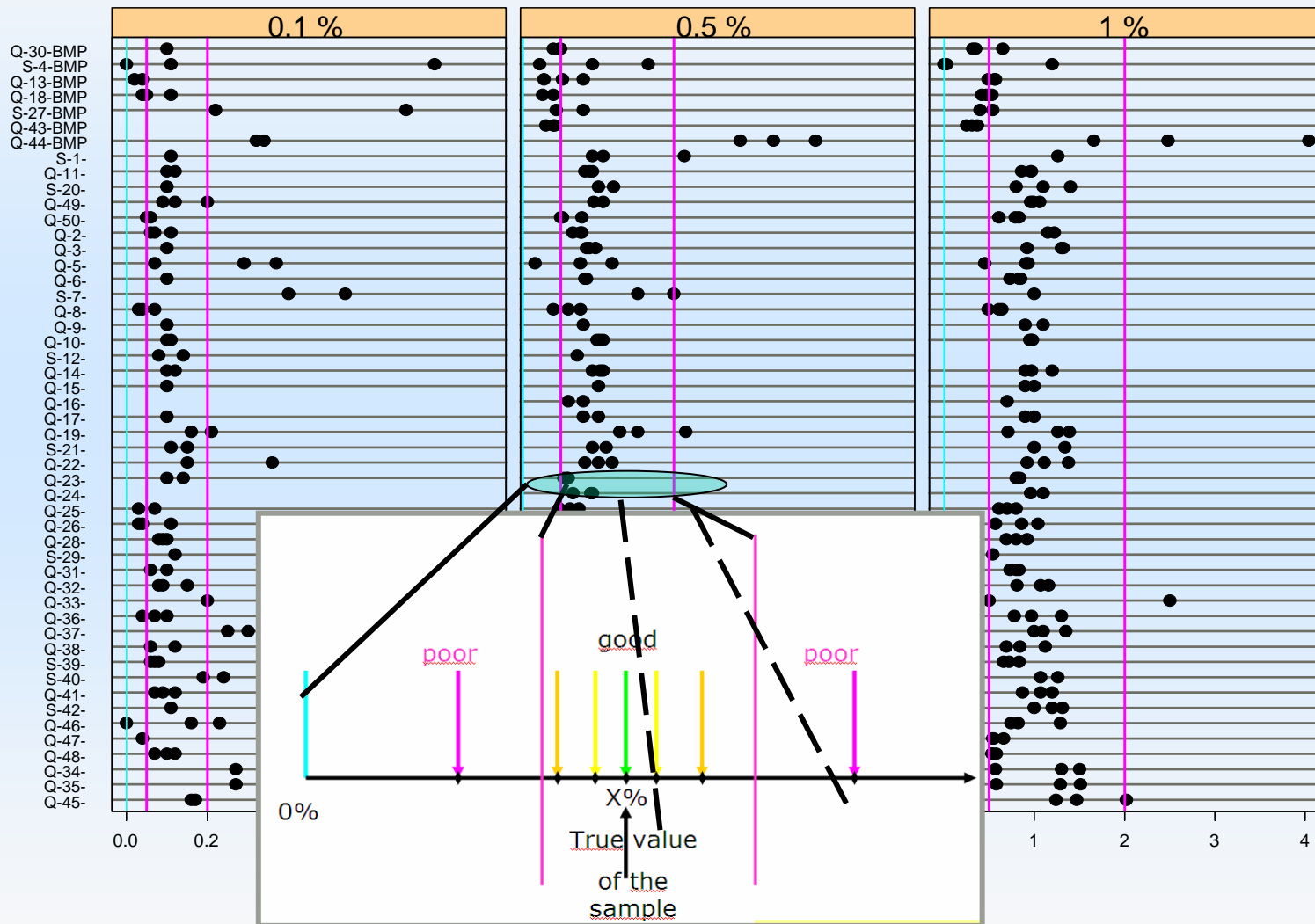
BMP  
Check  
example





# Same types of graphs are available for BMP, C, B and A rating, with their specific thresholds

Reference: True level in % seed



BMP  
Check  
example





## Further developments

Careful look by experts is important, but we move to more automatic decisions, as in other types of tests

Numerical criteria can change when necessary

Continuous feedback with participating laboratories

Each new PT may include a surprise!

Comment

Check for GM purity?

Question

Stacked genes ...?

Improve