Method Validation Experiments for DNA Barcoding

A Proposal

Susan J. Murch, University of British Columbia

How do we figure this out?

TECHNICAL COMMUNICATIONS

Probability of Identification: A Statistical Model for the Validation of Qualitative Botanical Identification Methods

Robert A. LaBudde
Least Cost Formulations, Ltd, 824 Timberlake Dr, Virginia Beach, VA 23464
Old Dominion University, Department of Mathematics and Statistics, Norfolk, VA 23529
James M. Harnly[1]
U.S. Department of Agriculture, Agricultural Research Center, Beltsville Human Nutrition Research Center, Food Composition and Methods Development Laboratory, Bldg 161 BARC-East, Beltsville, MD 20705

Botanical Identification Method

- Any qualitative method that reliability identifies a botanical material and returns a binary result
  - 0 = “not identified”
  - 1 = “identified”
- Based on:
  - number of replicates
  - Probability of Identification (POI)
- Needs to show:
  - inclusivity, exclusivity, probability of identification, robustness, reproducibility, repeatability, “other”

(LaBudde & Harnley, 2012)

Proof of Identity

- Statistical model to determine confidence in the qualitative method
- Depends on the concentration of target plant
- Question: how does the method make the transition from the positive to the negative response?
- Method Performance Requirements (MPR)
  - Specify the botanical material (inclusivity)
  - Specify the non-target materials (exclusivity)
  - Physical form of the materials
  - Minimum concentration
  - Define categories of “specific superior” and “specific inferior” test materials

Ideal Goal

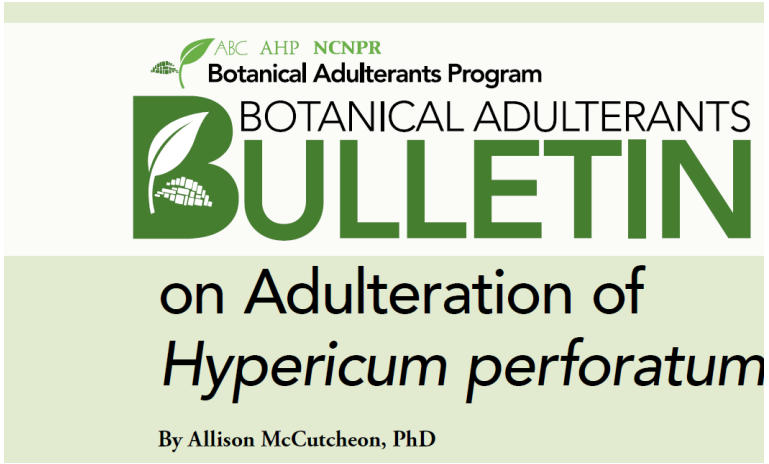
- Discriminate with a specified degree of confidence between the included and excluded

Inclusivity

- Ability of a BIM to correctly identify variants of the target material that meet the identity specification.
- Inclusivity Sampling Frame (ISF) – a list of practically obtainable botanical materials that are expected to give a positive result
  - Species
  - Subspecies
  - Cultivars
  - Growing location
  - Growing conditions
  - Post-harvest processing

What does the experiment look like?

- For each BIM, the MPR must provide a list of all necessary botanical variants that should provide a positive identification
- Case Study – St. John's wort



Inclusion:

- *Hypericum perforatum* L.

Exclusion:

- Chinese SJW (*Hypericum perforatum* subsp. chinense)
- Hybrids
- *Hypericum barbatum*
- *Hypericum hirsutum*
- *Hypericum montanum*
- *Hypericum elegans*
- *Hypericum patulatum*
- *Hypericum tetrapterum*
- 64 *Hypericum* species found in China

Study Design

= 160 samples +++

Number	Sample Identification
Inclusion Panel	
60	*Hypericum perforatum* - Wildtype
10	*Hypericum perforatum* – CV Anthos
10	*Hypericum perforatum* – CV Helos
10	*Hypericum perforatum* – CV Topas
Exclusion Panel	
10	*Hypericum perforatum* subsp. chinense
10	*Hypericum barbatum*
10	*Hypericum hirsutum*
10	*Hypericum montanum*
10	*Hypericum elegans*
10	*Hypericum patulatum*
10	*Hypericum tetrapterum*
???	64 *Hypericum* species found in China
Include or Exclude??	
???	Hybrids

Study Design (cont'd)

Number	Sample Identification
	Specific Superior Test Material (SSTM)
420	*Hypericum perforatum* (98%) + adulterant (2%) (60 per test x 7 likely adulterants)
	Specific Inferior Test Material (SITM)
60	*Hypericum perforatum* (90%) + *Hypericum perforatum* subsp. chinense (10%)
60	*Hypericum perforatum* (90%) + *Hypericum barbatum* (10%)
60	*Hypericum perforatum* (90%) + *Hypericum hirsutum* (10%)
60	*Hypericum perforatum* (90%) + *Hypericum montanum* (10%)
60	*Hypericum perforatum* (90%) + *Hypericum elegans* (10%)
60	*Hypericum perforatum* (90%) + *Hypericum patulatum* (10%)
60	*Hypericum perforatum* (90%) + *Hypericum tetrapterum* (10%)
???	64 *Hypericum* species found in China (?????)
	Include or Exclude??
???	Hybrids (?????)

= 840 samples +++

Interpreting the data

Table 2. Example performance requirements

Requirement	SSTM, %	Measure	Limit	No. of replicates to be tested	No. of failures allowed[a]
POI	100	95% 1-sided LCL	0.90 (FNF<0.10)	60	2
POI	0	95% 1-sided UCL	0.10 (FPF<0.10)	60	2

[a] In each case, no more than two failures are allowed.

Table 4. Observed SLV results for example BIM

SSTM, %	No. of test portions	No. identified	No. not identified	POI
0.0	60	1	59	0.0167
33.3	60	7	53	0.1167
66.7	60	27	33	0.4500
100.0	60	60	0	1.0000

Still need… SLV, Co-Lab Trial, chemical tests

Table 7. Collaborative study results for 0% SSTM concentration

AOAC Binary Data Interlaboratory Study Workbook Study Reported Values					Version: 2.2
	Sample ID 0% SSTM				
Sequence	Item	Symbol	Value	Approximately 95% LCL[a]	Approximately 95% UCL[b]
1	Total number of laboratories	p	10		
2	Total number of replicates	Sum(n(L))	120		
3	Overall mean of all data (grand mean)	LPOI or LPOD	0.0083	0.0015	0.0457
4	Repeatability SD	s(r)	0.0913	0.0807	0.1713
5	Among-laboratories SD	s(L)	0.0000	0.0000	0.0402
6	Homogeneity test of laboratory PODs	P-value	0.4303		
7	Reproducibility SD	s(R)	0.0913	0.0814	0.1064
8	Intraclass correlation coefficient for repeatability	I(r)	1.0000	0.8335	1.0000

[a] LCL = Lower confidence level.

[b] UCL = Upper confidence level.

Also the following admixture of dyes have been reported as co-occurring SJW adulterants: E123 Amaranth (FD&C Red #2), E133 Brilliant Blue (FD&C Blue #1), E110 Sunset Yellow (FD&C Yellow #6), and E102 Tartrazine (FD&C Yellow #5).[25]