Evaluation of Quantitative Real-Time PCR Assays for Detection of Citrus Greening

Turechek, W.W.[1], Irey, M.[2], Sieburth, P.[3], Brlansky, R.[4], DaGraça, J.[5], Graham, J.[4], Gottwald, T.[1], Hartung, J.[6], Hilf, M.[1], Kunta, M.[5], Manjunath, K.[7], Ling, H.[8], Ramdugu, C.[7], Roberts, P.[9], Rogers, M.[4], Shatters, R.[1], Sun, X.[10], and Wang, N.[4]

[1]United States Department of Agriculture (USDA) – Agricultural Research Service (ARS), U.S. Horticultural Research Laboratory, Fort Pierce, FL; [2]United States Sugar Corporation/Southern Gardens Citrus, Clewiston, FL; [3]Florida Department of Agriculture and Consumer Services (FDACS), Bureau of Budwood Registration, Winter Haven, FL; [4]University of Florida, Citrus Research and Education Center, Lake Alfred, FL; [5]Texas A&M University, Weslaco, TX; [6]USDA-ARS, Beltsville, MD; [7]USDA–ARS, Riverside, CA; [8]USDA–ARS, Parlier, CA, [9]University of Florida, Southwest Florida Research and Education Center, Immokalee, FL; [10]FDACS–DPI, Gainesville, FL.

Citrus huanglongbing (HLB), or citrus greening, is a serious and industry-limiting disease. It is caused by a fastidious prokaryote that survives in the sieve tube elements of infected plants. Three distinct species of the pathogen are known to cause disease with *Candidatus* Liberibacter asiaticus (Las) being the only species causing HLB in Florida, USA. Diagnosis of the pathogen is complicated by the fact that it cannot be grown in sterile culture, although recent progress towards culturing the organism has been made (1,8). Preliminary diagnoses can be made through visual symptoms, and greater certainty can be achieved through the use of PCR. The most widely used diagnostic test for pathogen identification is the quantitative real-time PCR (qPCR) assay developed by Li et al. (7). Other qPCR primer sets have been developed by Wang et al. (11), Shatters et al. (9), and by Irey (*unpublished*). It is not clear if any of these tests perform better than the others under any given set of conditions. Evaluating diagnostic tests typically relies on the ability to pre-classify so-called 'case' (diseased) and 'control' (healthy) samples with a 'gold standard' or reference test so that the sensitivity (the probability of detection in a truly infected plant) and specificity (the probability of non-detection in a truly healthy plant) of the new test can be accurately estimated – the notion being that the new test is cheaper, easier, faster, or all of the above relative to the gold standard. However, no gold standard test exists for HLB. The objectives of this study were to: 1) Evaluate the performance of several qPCR protocols as used in several diagnostic laboratories; and 2) Introduce methods of evaluating diagnostic tests in the absence of a gold standard.

Materials and Methods: A total of 276 DNA samples (extracts) were sent for analysis to thirteen labs in FL, CA, TX, and MD. The DNA samples consisted of presumed positive samples (i.e., samples collected from symptomatic tissue that tested positive in preliminary testing), presumed negative samples (i.e., samples collected from screened and tested budwood trees, or from orchards in states not known to have HLB), samples that tested as questionable in preliminary testing, samples from citrus-relatives, and water blanks containing no DNA (5). All samples were run blind by all labs. qPCR procedures varied among labs by the type of qPCR implemented (Taqman®, SYBR Green® and EvaGreen®), multiplex vs. single primer sets for PCR, different master mixes, different primer sets, different machines and different systems for determining cycling threshold values (5). Specific test results from ten laboratories were considered in this analysis and are shown in Table 1.

Table 1. Primer sets run by the collaborating labs.

	Lab									
Primer Set	1	2	3	4	5	6	7	8	9	10
Li et al. (7)	x	x	x	x	x	x		x	x	
Wang et al. (11)							x			
Irey (*unpublished*)								x		
Shatters et al. (9)										x

In a standard test evaluation, one might be tempted to consider symptomatic samples as disease positive (cases) and the asymptomatic samples as disease negative (controls) and perform a standard receiver operating characteristic (ROC) curve analysis. However, the expression or lack of expression of symptoms of HLB does not guarantee the presence or absence of the pathogen in that sample, and with the lack of a definitive gold standard test, alternatives to the ROC methodology must be employed for estimating test sensitivities and specificities.

Several methods have been introduced to evaluate diagnostic tests in the absence of a gold standard (10). For tests producing binary results, the sensitivity and specificity must be estimated for each of the R tests being evaluated and the prevalence must be estimated in each of the S populations where the tests were applied. In general, there are $(2^R\text{-}1)S$ degrees of freedom to estimate $(2R\text{-}1)S$ parameters for an arbitrary number of tests (R) and populations (S). Let $n_s(\mathbf{x})$ be the number of individuals receiving a classification $\mathbf{x}$ (e.g., there are $2^R = 8$ possible sets of test results [classifications] when $R = 3$ tests are applied to single population: 1-1-1, 1-1-0, 1-0-1, 0-1-1, 1-0-0, 0-1-0, 0-0-1, 0-0-0), π_s be the probability of disease, $x(r)$ be the classification (test result) of a sample for test r, and α_{rs} and β_{rs} be the sensitivity and specificity of test r (r = 1,2,…R) in a sample from population s (s=1,2,…,S). Assume that the π_s's are distinct between populations and that test results are conditionally independent. Under these

assumptions the test outcomes are distributed as S independent multinomials with log likelihood (10):

$$L = \sum_{1}^{S} \sum_{x} n_s(\boldsymbol{x}) \ln \left[\pi_s \prod_{r=1}^{R} \alpha_{rs}^{x(r)} (1-\alpha_{rs})^{1-x(r)} + (1-\pi_s) \prod_{r=1}^{R} \beta_{rs}^{x(r)} (1-\beta_{rs})^{1-x(r)} \right]$$

The log likelihood is over parameterized for R less than three with no functional constraints, but there are ways of handing this situation (4). There is no simple closed-form solution for the maximum likelihood estimates (MLE), but the function can be evaluated numerically. One such method is to use the iterative Newton-Raphson (NR) algorithm to obtain the MLEs. Another approach takes advantage of the unknown disease status of individuals by treating disease status as a latent (unobserved) variable. Latent class analysis is a procedure used to identify the k distinct classes (subgroups or categories) of a nominal latent variable. Latent class analysis was introduced by Lazarsfeld (6), and was significantly extended and formalized by Goodman (2,3) who developed the expectation-maximization (EM) algorithm that now forms the basis of most programs that fit latent class models. Here, we can define π as a latent variable with two classes, diseased and disease-free or healthy.

For this study, we compared test results of four diagnostic tests (R =4) from a single population (S=1) as run in 10 of the 13 labs (Table 1). For the primers developed by Li et al. (7), the mean Ct value was used and was calculated from the results of labs 1 thru 6, 8 and 9 from each sample. The Ct values recorded from labs 7, 8 and 10 were used for analysis, and were the result of the primer sets of Wang et al. (11), Irey, and Shatters et al. (9), respectively. The qPCR data were partitioned into two groups based on their test results. A test negative sample is a sample that has exceeded a predetermined threshold Ct value, and a test positive sample is a sample that has not exceeded the threshold Ct value. Since it is unknown which threshold Ct value best classifies diseased and healthy samples, analyses were conducted individually for Ct threshold values from 25 to 40.

PROC NLIN was used to obtain the MLEs via the Newton-Raphson algorithm for the model above (SAS Institute, ver. 9.2, Cary, NC). We used PROC LCA to fit the latent class model by treating disease status as a 2-class latent variable. The MLEs were obtained via the EM algorithm. PROC LCA can be obtained free-of-charge through Pennsylvania State University's Methodology Center's website (http://methodology.psu.edu/index.php/downloads/proclcalta). Youden's index, J (=α+β-1), was calculated from the MLE's and was used to identify Ct values located farthest from the diagonal line that bisects a plot of the true positive proportion (α) versus the false positive proportion (1-β), and serves as one means of identifying practical threshold values.

Results and Discussion: General results from the laboratory comparison study are published as part of the PROCEEDINGS OF THE INTERNATIONAL RESEARCH CONFERENCE ON HUANGLONGBING (5), and are not the focus here. However, one result reported in that summary of particular relevance to this study was labs running the primer set of Li et al. (7) produced similar Ct values for any given sample. This result was used to justify averaging the Ct values across labs and using these mean values in subsequent analyses. Because the other tests evaluated were run in single labs, averaging was not necessary.

Estimation of test sensitivities (α) and specificities (β), as well as disease prevalence (π), were nearly identical between the NR and EM algorithms. Thus, only the MLE's from EM algorithm are shown (Table 2). This result was not unexpected since the same model was being fit and the same parameters were being estimated. In practice, however, it was more difficult to achieve convergence and/or reasonable estimates using the NR algorithm as parameter estimation was sensitive to the starting values. Unlike ROC analysis, the sensitivities (specificities) do not systematically increase (decrease) as the threshold increases. This is because disease prevalence is not treated as a constant and can change accordingly.

The sensitivities for all four tests were very high ($\alpha \geq 0.97$) for threshold Ct values above 32, with the Irey primers having the highest sensitivity over the greatest range of threshold Ct values (Table 2). The specificities for all four tests were very high as well ($\beta \geq 0.96$) for threshold Ct values below 36, with the Li primers having the highest specificity over the greatest range of Ct values. Maximum values of J were near or equal to 1 for all primer sets, and were obtained at Ct values of 31 and 32 for all primer sets except for the Li primers. J obtained its maximum at Ct = 36 for the Li primers, and was the only primer pair to achieve the maximum value of 1 for J, indicating that a cutoff of Ct = 36 provided perfect discrimination between test positive and test negative samples for this set of data. In practice, many diagnostic labs choose Ct = 30 as the cutoff, opting for greater specificity over sensitivity. Results from this study indicate that greater sensitivity can be achieved with minimal loss in specificity by choosing Ct = 31 or 32 for the Wang, Irey and Shatters primers and Ct = 36 for the Li primers.

In conclusion, all four primer sets evaluated have merit in the diagnosis of Las. Researchers and diagnosticians can use these results to help determine which test and which threshold is best suited to their application based on the need for either greater sensitivity or specificity (…the inherent tradeoff). With the exception of the Li primers, Youden's index identified what most would consider a relatively high threshold Ct value as the optimal cutoff. Although several reasons could be hypothesized as to why this is so, the underlying cause is that presumed healthy samples generate a distribution of Ct values that ultimately result in prediction errors. This is not unique to this system, as nearly all diagnostic tests will, to various degrees, give false positive and/or negative results. The challenge to quantifying these errors for detecting Las was complicated

by the lack of a gold standard. The methods demonstrated here provide the tools to estimate these errors in the face of uncertain disease status. The dataset used to demonstrate this analysis, although sufficient, was not optimal. In general, the distribution of Ct values between the presumed positive and negative samples was non-overlapping, and the number of uncertain samples tended to associate with the presumed negative samples (see poster figure). This separation could be identified in the analyses through *J*.

Literature Cited

1. Davis, M.J., Mondal, S.N., Chen, H., Rogers, M.E., and Brlansky, R.H. 2008. Co-cultivation of '*Candidatus* Liberibacter asiaticus' with Actinobacteria from citrus with huanglongbing. Plant Disease 92:1547-1550.

2. Goodman, L.A. 1974a. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 61:215-231.

3. Goodman, L.A. 1974b. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I-A: Modified latent structure approach. American Journal of Sociology 79:1179-1259.

4. Hui, S.L., and Walter, S.D. 1980. Estimating the error rates of diagnostic tests. Biometrics 36:167-171.

5. Irey, M., Sieburth, P., Brlansky, R., DaGraça, J., Graham, J., Gottwald, T., Hartung, J., Hilf, M., Kunta, M., Manjunath, K., Ling, H., Ramdugu, C., Roberts, P., Rogers, M., Shatters, R., Sun, X., and Wang, N. Lessons learned from a comparison and evaluation of multiple HLB testing laboratories employing common and different testing methodologies applied to a common set of samples. Proceedings of International Research Conference On Huanglongbing, Orlando, Florida December 1-5, 2008 (*In press*).

6. Lazarsfeld, P.F. 1950. The logical and mathematical foundations of latent structure analysis. *In*: S.A. Stouffer et al. (eds.) Measurement and Prediction. Princeton University Press, Princeton, N.J.

7. Li, W., Hartung, J.S., and Levy, L. 2006. Quantitative real-time PCR for detection and identification of *Candidatus Liberibacter* species associated with citrus huanglongbing. Journal of Microbiological Methods 66:104-115.

8. Sechler, A, Schuenzel, E.L., Cooke, P., Donnua, S., Thaveechai, N., Postnikova, E., Stone, A.L., Schneider, W.L., Damsteegt, V.D., and Schaad, N.W. 2009. Cultivation of '*Candidatus* Liberibacter asiaticus', '*Ca*. L. africanus', and '*Ca*. L. americanus' associated with huanglongbing. Phythyopathology 99:480-486.

9. Shatters, Jr., R. G., Hunter, W., Hall, D., Niedz, R., Evens, T. J. Optimizing qPCR for detection of *Candidatus* Liberibacter species in plants and psyllid samples. Proceedings of the International Research Conference On Huanglongbing, Orlando, Florida December 1-5, 2008 (*In press*).

10. Walter, S.D. and Irwig, L.M. 1988. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J. Clin. Epidemiol. 41:923-937.

11. Wang, Z., Yin, Y., Hu, H., Yuan, Q., Peng, G., and Xia, Y. 2006. Development and application of molecular-based diagnosis for '*Candidatus* Liberibacter asiaticus', the causal pathogen of citrus huanglongbing. Plant Pathology 55:630-638.

Table 2. Maximum likelihood estimates of the sensitivity (α), specificity (β), and disease prevalence (π) for four qPCR primer sets for the Ct value thresholds listed. MLE's were estimated with the expectation-maximization (EM) algorithm, treating prevalence as a latent variable with two classes. The maximum value of Youden's index (*J*) for each primer set is shown in boldface; the corresponding Ct value could be considered optimal under certain conditions and assumptions.

		Primer Set											
		Wang et al. (11)			Irey (*unpublished*)			Shatters et al. (9)			Li et al. (7)		
Ct Value	π	α	β	*J*	α	β	*J*	α	β	*J*	α	β	*J*
28	0.21	0.93	1.00	0.927	0.98	0.99	0.969	0.91	1.00	0.915	0.91	1.00	0.915
29	0.22	0.95	1.00	0.947	1.00	0.99	0.991	0.95	1.00	0.952	0.92	1.00	0.920
30	0.23	0.98	0.99	0.979	1.00	0.99	0.985	0.94	1.00	0.938	0.94	1.00	0.938
31	0.24	0.98	1.00	***0.980***	1.00	0.99	***0.995***	0.98	1.00	***0.985***	0.94	1.00	0.938
32	0.24	0.98	1.00	***0.980***	1.00	0.98	0.981	0.98	1.00	***0.985***	0.97	1.00	0.969
33	0.24	0.98	0.99	0.978	1.00	0.98	0.978	0.96	1.00	0.958	0.96	1.00	0.963
34	0.26	0.94	0.99	0.932	1.00	0.99	0.986	0.97	0.97	0.943	0.93	1.00	0.928
35	0.26	0.98	0.99	0.971	1.00	0.97	0.969	0.98	0.96	0.936	0.95	1.00	0.949
36	0.26	0.99	0.98	0.962	1.00	0.94	0.941	0.99	0.96	0.947	1.00	1.00	***1.000***
37	0.29	0.94	0.97	0.914	1.00	0.95	0.945	0.94	0.98	0.921	0.97	1.00	0.972
38	0.32	0.93	0.94	0.868	1.00	0.89	0.889	0.89	0.98	0.864	0.99	1.00	0.999
39	0.38	0.86	0.86	0.723	1.00	0.88	0.875	0.77	0.99	0.768	1.00	0.92	0.922
40	0.78	1.00	0.78	0.780	1.00	0.78	0.780	1.00	0.78	0.780	1.00	0.78	0.780