

# Enhancing PCR Capacity To Detect ‘*Candidatus Liberibacter asiaticus*’ Utilizing Whole Genome Sequence Information

Minli Bao,<sup>1,2</sup> Zheng Zheng,<sup>1</sup> Xiaoran Sun,<sup>3</sup> Jianchi Chen,<sup>2,†</sup> and Xiaoling Deng<sup>1,†</sup>

<sup>1</sup> South China Agricultural University, Guangzhou, China

<sup>2</sup> U.S. Department of Agriculture Agricultural Research Service, Parlier, CA, U.S.A.

<sup>3</sup> Florida Department of Agriculture and Consumer Services, Gainesville, FL, U.S.A.

## Abstract

‘*Candidatus Liberibacter asiaticus*’ (CLAs) is an unculturable  $\alpha$ -proteobacterium associated with citrus Huanglongbing (HLB; yellow shoot disease). PCR procedures that accurately confirm or exclude CLAs infection in citrus tissue/Asian citrus psyllid (ACP) samples are critical for HLB management. When CLAs was described in 1994, a 23-bp signature oligonucleotide sequence (OII) in the 16S rRNA gene (*rrs*, three genomic copies) was identified based on Sanger sequencing. OII contains single nucleotide polymorphisms (SNPs) distinguishing CLAs from non-CLAs species. The SNPs were used to design the primer HLBas, a key primer for a commonly used TaqMan PCR system (HLBas-PCR) for CLAs detection. Recent developments in next-generation sequencing technology have led to the identification of 15 CLAs whole genome sequence strains (WGSs). Analyses of CLAs WGSs have generated a significant amount of biological information that could help to improve CLAs detection. Utilizing the WGS information, this study re-evaluated

the sequence integrity of OII/HLBas and identified and/or confirmed a missing nucleotide G in the two primers. Replacement primers for OII and HLBas are proposed. At low cycle threshold (Ct) values (e.g., <30), HLBas-PCR remained reliable in CLAs determination. At high Ct values (e.g., >30), HLBas-PCR alone was unreliable in differentiating whether samples contain low CLAs titers or whether CLAs is not present. The availability of ribonucleotide reductase (RNR)-PCR derived from the five-copy *nrdB* gene helped to resolve this problem. To further enhance low CLAs titer detection, a 4CP-PCR system, based on a four-copy genomic locus, was developed. Evaluation of 107 HLB samples (94 citrus and 13 ACP) showed that 4CP-PCR was more sensitive than HLBas-PCR and shared similar sensitivity with RNR-PCR.

**Keywords:** prokaryotes, pathogen detection, techniques

‘*Candidatus Liberibacter asiaticus*’ (CLAs) is associated with citrus Huanglongbing (HLB; yellow shoot disease), a highly destructive disease in citrus production worldwide (Bové 2006). CLAs is largely detected based on genomic sequence characters because of the inability of in vitro culture and therefore limited biological information. Using a conserved primer set in conjunction with restriction enzyme analyses, a nearly complete 16S rRNA gene (*rrs*) sequence of CLAs strain Poona from India was amplified, cloned, and sequenced with the Sanger method (accession number L22532; Jagoueix et al. 1994). A 23-bp oligo sequence, OII (GCGCGTATGC AATACGAGCGGCA), was identified and served as a signature sequence for CLAs. OII, paired with a non-CLAs-specific reverse primer OI2c, was used for CLAs detection in the conventional end-point PCR format (OII-PCR) (Jagoueix et al. 1994, 1996). Referenced to the specificity of OII, primer HLBas was designed and used to develop a TaqMan PCR system (HLBas-PCR) for CLAs detection

(Li et al. 2006). HLBas-PCR has been commonly used for CLAs detection in HLB diagnostic practice, typically for validation of HLB symptomatic citrus samples.

Because CLAs has spread to previously HLB-free regions such as California (Kumagai et al. 2013), detection of low titer CLAs in early infected and asymptomatic samples becomes critical. Although a high cycle threshold (Ct) (e.g.,  $\geq 35$ ) could be attributable to a low titer of CLAs, it could also be a result of the presence of non-CLAs-specific DNA templates, such as those from related but unknown non-CLAs bacteria. Accurate detection of low titer CLAs has a significant consequence for regulatory actions in a region such as California where a current HLB management goal is to eradicate CLAs. Detection of low titer CLAs is also critical for the citrus nursery industry, since utilization of clean nursery stock is part of the standard procedure in HLB management.

Sequence integrity of OII is important for accurate PCR detection of CLAs. Based on GenBank records, a nucleotide G was found missing between positions 10 and 11 of OII or between positions 14 and 15 of HLBas (DQ302750; Tsai et al. 2006). This was followed with two sequences (GenBank records only) of a Florida strain (DQ471900) and a Brazil strain (DQ471901) from the laboratory of Dr. J. Bové, who described OII. It is understandable that early Sanger sequencing technology was less sophisticated and could be prone to sequencing errors. For example, there were nine Ns in L22532. Current next-generation sequencing (NGS) technology such as Illumina HiSeq/MiSeq significantly increases sequencing throughput and nucleotide coverages to improve sequence quality. A total of 15 CLAs strain WGSs have been acquired through NGS and are available in GenBank. Analyses on CLAs WGSs have increased current understanding of CLAs biology (Dai et al. 2019; Duan et al. 2009; Zheng et al. 2016, 2017, 2018a, b). The new knowledge provides a base to improve CLAs detection.

Because of its universal presence, the *rrs* sequence has been used in bacterial classification and identification, mostly above species/genus levels (Woese and Fox 1977). CLAs has three copies of *rrs* (Duan et al. 2009; Kim and Wang 2009). This is advantageous for sensitive detection of CLAs with OII-PCR and HLBas-PCR

<sup>†</sup>Corresponding authors: X. Deng and J. Chen; [xldeng@scau.edu.cn](mailto:xldeng@scau.edu.cn) and [jianchi.chen@ars.usda.gov](mailto:jianchi.chen@ars.usda.gov)

**Funding:** This work was supported in part by the California Citrus Research Board, the National Key Research and Development Program of China (grant 2018YFD0201500), and Chinese Modern Agricultural Technology Systems (grant CARS-26). This research used resources provided by the SCINet project of the U.S. Department of Agriculture Agricultural Research Service (project number 0500-00093-001-00-D). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture (USDA). USDA is an equal opportunity employer.

\*The e-Xtra logo stands for “electronic extra” and indicates that supplementary tables and one supplementary figure are published online.

The author(s) declare no conflict of interest.

Accepted for publication 3 August 2019.

compared with the PCR systems using primers based on single-copy genes. A recent WGS analysis identified a five-copy ribonucleotide reductase (RNR) gene (*nrdB*), which was also conserved among all bacterial species (Zheng et al. 2016). Primer set RNRf/RNRr was designed based on *nrdB*. A PCR system (RNR-PCR) using both TaqMan probe and CYBR green formats was developed. RNR-PCR showed higher detection sensitivity than HLBas-PCR (Zheng et al. 2016). In the same study, a four-copy genomic locus was also described but has not yet been evaluated for CLAs detection.

In this study, WGS information was used to improve CLAs detection accuracy and sensitivity. The issues of missing nucleotide G in OI1/HLBas and interpretation of high Ct values were addressed. This study had three objectives: (i) to re-evaluate the sequence integrity of OI1/HLBas based on WGS data and evaluate the impact of the missing G for CLAs detection; (ii) to evaluate the use of RNR-PCR to assist differentiation between low CLAs titer and no CLAs in samples with high Ct values from HLBas-PCR; and (iii) to develop a new PCR system based on a four-copy genomic locus and evaluation of its potential to detect CLAs.

## Materials and Methods

**CLAs DNA sequences and samples.** All complete and partial CLAs *rrs* sequences in the GenBank database (version 231.0) were identified by keyword search and/or by BLASTn search using an *rrs* copy (CD16\_01840) from CLAs strain A4 as a query and downloaded. All CLAs WGSs in the GenBank database were also downloaded. Sequence variations in OI1/HLBas regions of all of the *rrs* sequences were examined by the standalone BLASTn program (Camacho et al. 2009) queried by OI1/HLBas with an e-value of 1,000 and a word size of 11. The original MiSeq read data (a total of  $3.28 \times 10^7$  reads with 250 bp/read) in strain A4 (Zheng et al. 2014b) were used for BLASTn searches with OI1 as a query to search for intact reads with the missing G. HLB-affected citrus tissue and Asian citrus psyllid (ACP) samples were collected from China between 2006 and 2017 and used to validate and compare different PCR systems.

**Evaluation of the missing G in primer HLBas for CLAs detection.** A previously described TaqMan PCR procedure (Li et al. 2006) was followed. The reverse primer HLBr and TaqMan probe HLBp were used. Experiments were focused on variations of the forward primer HLBas as follows: (i) a nucleotide G was added after position 14 of HLBas and formed new primer HLBas-G, which was used for PCR (HLBas-G-PCR); and (ii) adjustments were made in HLBas-G by adding a nucleotide to the 5' end or deleting a nucleotide at the 3' end to obtain a series of primers with a melting temperature ( $T_m$ ) in the range of 61 to 63°C. Experiments with different PCR systems were performed using the same CLAs-infected tissue sample and the same no-CLAs sample from China. Comparisons were made based on Ct values.

**Interpretation of high Ct values from HLBas-PCR.** As mentioned earlier, HLBas-PCR is currently the most commonly used protocol for CLAs detection. However, false-positive results can occur with high Ct values. To resolve this problem, RNR-PCR was introduced with the following hypothesis: for a sample with CLAs, if HLBas-PCR (based on three-copy *rrs*) detected the bacterium, RNR-PCR (based on five-copy *nrdB*) should also detect the bacterium with a lower Ct value. The following ratio was introduced:  $R_{nrdB/rrs} = \text{RNR-PCR Ct}/\text{HLBas-PCR Ct}$ , where  $R_{nrdB/rrs} < 1$  was considered CLAs positive and  $R_{nrdB/rrs} > 1$  was considered CLAs negative. The hypothetical scenario of  $R_{nrdB/rrs} = 1$  is further explained in the *Discussion*.

The  $R_{nrdB/rrs}$  experiments were performed using “true low titer CLAs” and “true no-CLAs” samples. The word “true” was emphasized to indicate the preknown CLAs status in these experiments. True low titer CLAs samples were obtained by a series of 10-fold dilutions of CLAs samples with HLBas-PCR Ct values mostly between 19 and 26. The highest dilution with a Ct value  $< 40$  was considered as a true low titer CLAs sample. Note that the thermal cyclers used in this study gave no Ct values if they were  $\geq 40$ . For the  $R_{nrdB/rrs}$  calculation, the no-CLAs Ct values were recorded as 40 in this study.

The no-CLAs samples were obtained from citrus samples collected from Chongqing, China and Parlier, California. Both locations are free of HLB. An additional no-CLAs sample was DNA extracted from an olive tree in Parlier.

**Development of 4CP-PCR and evaluation.** The four-copy locus or *4cp* sequences in the CLAs strain A4 genome sequence (Zheng et al. 2016) were extracted and designated as *4cp1*, *4cp2*, *4cp3*, and *4cp4*. The sequences were aligned using Clustal Omega (Sievers et al. 2011). Primer sets were designed from regions shared by all four *4cp* sequences using Primer 3 software (version 4.1.0) (Untergasser et al. 2012). Open reading frames (ORFs) in each *4cp* were identified and amino acid sequences were determined. The genetic nature of each of the four ORFs was searched using the NCBI BLASTp program and referenced to the GenBank database (version 231.0). Relationships among the four ORFs were determined using the neighbor-joining method with MEGA 7.0 (Kumar et al. 2016). A set of 107 CLAs samples (94 from citrus and 13 from ACP) collected from China was used to evaluate 4CP-PCR, along with HLBas-PCR, CLAs-4G-PCR, and RNR-PCR. PCR results were evaluated by Duncan's multiple range test using SPSS software (version 18.0; IBM Corporation).

**TaqMan PCR procedure.** PCR assays were performed in two laboratories (Guangzhou, China and Parlier, California). In the China laboratory, the CFX Connect Real-Time System (Bio-Rad, Hercules, CA) was used. PCR was performed in 20- $\mu$ l volume reactions consisting of the following reagents: 10  $\mu$ l of Bestar qPCR Master Mix (DBI Bioscience), 1  $\mu$ l of DNA template (25 ng), 0.2  $\mu$ l of TaqMan probe (5  $\mu$ M), and 0.4  $\mu$ l of each forward and reverse primer (10  $\mu$ M). The standard amplification procedure was 95°C for 2 min, followed by 40 cycles at 95°C for 10 s and 60°C for 30 s. The data were analyzed using Bio-Rad CFX Manager 2.1 software with automated baseline settings and a manually set threshold at 0.1.

In the U.S. laboratory, the Applied Biosystems StepOnePlus Real-Time PCR System was used. PCR was performed in 20- $\mu$ l volume reactions consisting of the following reagents: 10  $\mu$ l of TaqMan Fast Universal PCR Master Mix (2 $\times$ ) (Applied Biosystems), 1  $\mu$ l of DNA template (25 ng), 0.2  $\mu$ l of TaqMan probe (5  $\mu$ M), and 0.4  $\mu$ l of each forward and reverse primer (10  $\mu$ M). The standard amplification procedure was 95°C for 20 s, followed by 40 cycles at 95°C for 10 s and 60°C for 20 s. The fluorescence signal was captured at the end of each 60°C step. The data were analyzed using StepOnePlus software version 2.3 (Applied Biosystems) with automated baseline settings and a manually set threshold at 0.1.

## Results

**The missing G in OI1/HLBas.** A total of 628 CLAs *rrs* records from nine countries/regions were identified in the GenBank database (version 231.0) (Table 1; Supplementary Tables). Among the 628 sequences, 486 records did not include the OI1 primer sequence. Within the other sequences harboring OI1, 123 sequences missed the G, likely related to suboptimal sequencing and editing errors. However, 19 sequences had the G and the first record was DQ302750.1 from Taiwan in 2005 (Tsai et al. 2006).

A total of 15 CLAs WGSs have been published, all from NGS (Table 1). As expected, 10 WGSs showed three copies of *rrs*, four copies of *4cp*, and five copies of *nrdB*. The other five WGSs showed copy number variations at the three genomic loci, possibly owing to their draft sequence status. In draft genome sequences, repeat regions might fail to assemble. All *rrs* sequences in CLAs WGSs showed the presence of the G nucleotide (Table 1). BLASTn search in the strain A4 MiSeq data did not find any reads covering the missing G of OI1. Therefore, this study concluded that nucleotide G was missed in OI1 and HLBas.

**Effects of the missing G in PCR detection.** As shown in Table 2, primer HLBas-G had a higher  $T_m$  (63.48°C) than that of HLBas (62.40°C). HLBas-G-PCR slightly reduced the Ct value compared with HLBas-PCR (22.28 versus 22.35, respectively). However, there was statistically no significant difference between these two PCR systems. Three additional experiments were conducted to compare HLBas-PCR and HLBas-G-PCR as shown in Supplementary Table

S2. Experiments 1 and 3 showed lower Ct values in HLBas-G-PCR and experiment 2 showed almost identical Ct values between the two PCR systems. Regardless, both HLBas-PCR and HLBas-G-PCR successfully detected CLAs in this relatively high CLAs titer situation (e.g., Ct values in the 20s).

PCR results using forward primers different in the G positions are shown in Table 2. Compared with HLBas and HLBas-G (significance level d), three PCRs (CLAs-2G, CLAs-3G, and CLAs-5G) were statistically less sensitive (significance levels a, b, and c) and four (CLAs-1G, CLAs-4G, CLAs-6G, and CLAs-7G) were more sensitive (significance level e). From the four sensitive PCR primers, CLAs-4G was arbitrarily chosen as a replacement of HLBas.

**True low titer CLAs experiments.** A total of 15 dilution experiments were performed to compare HLBas-PCR with RNR-PCR (Supplementary Tables; experiments 1 through 15). CLAs was detected at Ct values in 39s with HLBas-PCR (Supplementary Tables; experiment 11, Ct = 39.75 at 10<sup>-4</sup> dilution; experiment 15, Ct = 39.25

at 10<sup>-5</sup> dilution; and experiment 16, Ct = 39.25 at 10<sup>-4</sup> dilution). In both experiments 11 and 15, RNR-PCR showed lower Ct values at the same dilution level. R<sub>nrdB/trs</sub> values were all <1 for all dilution tests. RNR-PCR was not performed on experiments 16 and 17. Results of experiment 15 were used as a representative and are shown in Table 3.

**True no-CLAs experiments.** A total of nine true no-CLAs samples were tested (Supplementary Tables; experiments 18 to 26). To note, experiment 25 used a CLAs-free ACP sample. The PCR internal control for ACP DNA (Wu et al. 2018) was positive. Experiment 26 used an olive tree sample. The PCR internal control for plant DNA (Yan et al. 2012) was positive. As expected, most PCR results showed no Ct or Ct ≥ 40. Yet three samples showed Ct values of 37 to 39 using HLBas-PCR and CLAs-4G PCR. In all cases, RNR-PCR showed higher Ct values (Ct > 40; i.e., R<sub>nrdB/trs</sub> > 1). Results of experiment 24 were used as a representative of all experiments and are shown in Table 4.

**Table 1.** Sequence integrity of the OII region (status of OII+G) in singly submitted 16S rRNA gene sequences and whole genome sequences of ‘*Candidatus Liberibacter asiaticus*’ (CLAs) in GenBank (version 231.0) and related information

Strain	OII+G <sup>w</sup>	GenBank accession	Country/ region	Host <sup>x</sup>	Sequencing format	Read length (bp)	Assembly size (Mbp)	Average nucleotide coverage (X) <sup>y</sup>	Genome completion status	Copy detected <sup>z</sup>			Sequences in GenBank (n)
										<i>rrs</i>	<i>4cp</i>	<i>nrdB</i>	
Single <i>rrs</i> record													
No OII	NA	AY919311.1	Brazil	Citrus	Sanger	1,122	1,122	1–2	NA	NA	NA	NA	486
Poona	No	L22532.1	India	Citrus	Sanger	1,448	1,448	1–2	NA	NA	NA	NA	123
Kumquat 1	Yes	DQ302750.1	Taiwan	Citrus	Sanger	1,389	1,389	1–2	NA	NA	NA	NA	19
Total records (n)													628
Whole genome sequence													
Psy62	Yes	CP001677.5	U.S.A.	ACP	454 pyrosequencing	216	1.23	16	Complete	3	4	5	1
gxpsy	Yes	CP004005.1	China	ACP	HiSeq 2000	300	1.27	50–80	Complete	3	4	5	1
A4	Yes	CP010804.1	China	Citrus	MiSeq 250	250	1.23	138	Complete	3	4	5	1
Ishi-1	Yes	AP014595.1	Japan	ACP	HiSeq 2000	75	1.19	300	Complete	3	4	5	1
JXGC	Yes	CP019958.1	China	Citrus	HiSeq 3000	150	1.23	95	Complete	3	4	5	1
AHCA1	Yes	CP029348.1	U.S.A.	ACP	HiSeq 3000	150	1.21	69	Draft	3	4	5	1
HHCA	Yes	JMIL000000000.2	U.S.A.	Citrus	MiSeq 250	250	1.15	6	Draft	3	3	5	1
FL17	Yes	JWHA000000000.1	U.S.A.	Citrus	MiSeq 300	300	1.23	50	Draft	3	4	5	1
YCpsy	Yes	LHIM000000000.1	China	ACP	MiSeq 250	250	1.23	1,120	Draft	3	4	5	1
SGCA5	Yes	LMT000000000.1	U.S.A.	Citrus	MiSeq 250	250	1.20	12	Draft	1	2	3	1
TX2351	Yes	MTIM000000000.1	U.S.A.	ACP	MiSeq 250	250	1.25	7	Draft	3	4	5	1
YNIS7C	Yes	QXD001000000.1	China	Citrus	HiSeq X Ten	150	1.19	15	Draft	3	4	5	1
SGCA1	Yes	QFZT01000587.1	U.S.A.	Citrus	MiSeq 250	250	0.24	1	Draft	1	2	1	1
TX1712	Yes	QEWL01000001.1	U.S.A.	Citrus	MiSeq 300	300	1.20	10	Draft	1	1	5	1
SGpsy	Yes	QFZJ000000000.1	U.S.A.	ACP	MiSeq 250	250	0.78	2	Draft	1	3	2	1
Total records (n)													15

<sup>w</sup> OII+G refers to the presence of nucleotide G between positions 10 and 11 of oligomer OII (GCGCGTATGCgAATACGAGCGGCA) as identified by the lowercase letter g. NA = not applicable.

<sup>x</sup> ACP = Asian citrus psyllid.

<sup>y</sup> For whole genome sequences, X = total CLAs sequence data (bp)/1.2M (rough estimate of CLAs genome size).

<sup>z</sup> Not all whole genome sequences had three copies of *rrs*, four copies of *4cp*, and five copies of *nrdB* because of their draft status.

**Table 2.** Evaluation of the absence or presence of nucleotide G in the forward primers using the U.S. Department of Agriculture Animal and Plant Health Inspection Service TaqMan PCR system (Li et al. 2006)

Primer name <sup>v</sup>	Aligned primer sequence <sup>w</sup>	T <sub>m</sub> (°C) <sup>x</sup>	CLAs citrus <sup>y</sup>	Non-CLAs citrus <sup>z</sup>
HLBas	TCGAGCGCGTATGC-AATACG	62.40	22.35 ± 0.00 d	40.00
HLBas-G	TCGAGCGCGTATGCgAATACG	63.48	22.28 ± 0.04 d	40.00
<b>CLAs-1G</b>	CAAGTCGAGCGCGTATGCg	62.18	21.92 ± 0.03 e	40.00
CLAs-2G	CAAGTCGAGCGCGTATGCgA	63.50	23.00 ± 0.02 a	40.00
CLAs-3G	AAGTCGAGCGCGTATGCgAA	62.86	22.80 ± 0.03 b	40.00
<b>CLAs-4G</b>	<b>AGTCGAGCGCGTATGCgAAT</b>	62.66	22.08 ± 0.02 e	40.00
CLAs-5G	GTCGAGCGCGTATGCgAATA	61.00	22.54 ± 0.12 c	40.00
<b>CLAs-6G</b>	TCGAGCGCGTATGCgAATAC	61.00	22.08 ± 0.01 e	40.00
<b>CLAs-7G</b>	CGAGCGCGTATGCgAATACG	62.24	22.06 ± 0.02 e	40.00

<sup>v</sup> Bold indicates the recommended primer to replace HLBas. HLB = Huanglongbing and CLAs = ‘*Candidatus Liberibacter asiaticus*’.

<sup>w</sup> The missing G in HLBas was identified by a lowercase letter g.

<sup>x</sup> T<sub>m</sub> = melting temperature.

<sup>y</sup> PCR cycle threshold (Ct) values are presented. The same CLAs-infected citrus sample from Guangdong Province of China was used for all PCR experiments with duplicates. Different letters indicate a significant difference at α = 0.01 from Duncan’s multiple range test.

<sup>z</sup> The no-CLAs citrus sample was from Parlier, California; 40.00 indicates Ct values ≥ 40.00.

**Primer design and characterization of 4cp loci.** According to A4 WGS (CP010804.1), each 4cp locus was 1,227 bp including a single ORF (Figs. 1 and 2). Locations of primer and probe sequences of the 4CP-PCR system are shown in Figure 1 and Supplementary Tables. The sequences of the primer and probe regions at four

**Table 3.** A representative true low titer ‘*Candidatus Liberibacter asiaticus*’ (CLAs) PCR experiment (experiment 15) on dilutions of a citrus DNA sample collected from Guangdong, China and infected with CLAs

Dilution	PCR systems <sup>y</sup>			$R_{nrdb/rrs}^z$
	HLBas	CLas-4G	RNR	
10 <sup>0</sup>	21.57	21.18	20.77	0.96
10 <sup>-1</sup>	24.88	24.43	23.82	0.96
10 <sup>-2</sup>	28.30	27.97	27.14	0.96
10 <sup>-3</sup>	31.86	31.56	30.44	0.96
10 <sup>-4</sup>	35.29	35.11	33.77	0.96
10 <sup>-5</sup>	<b>39.25</b>	<b>38.07</b>	<b>36.98</b>	0.94
10 <sup>-6</sup>	40.00	40.00	<b>38.72</b>	0.97
10 <sup>-7</sup>	40.00	40.00	40.00	NC

<sup>y</sup> HLB = Huanglongbing, HLBas = HLBas-PCR (HLBas/HLBp/HLBr), CLas-4G = CLas-4G-PCR (CLas-4G/HLBp/HLBr), and RNR = ribonucleotide reductase-PCR (RNRf/RNRp/RNRr). Values are PCR cycle threshold (Ct) values. Ct values at the highest dilutions are in bold.

<sup>z</sup>  $R_{nrdb/rrs}$  = Ct value ratio of RNR-PCR/HLBas-PCR. When both PCR Ct values were 40.00,  $R_{nrdb/rrs}$  was not calculated (NC).

**Table 4.** A representative true no-‘*Candidatus Liberibacter asiaticus*’ (CLAs) experiment (experiment 24) on dilutions of a citrus DNA sample free of CLAs collected from Chongqing, China

Dilution	PCR systems <sup>y</sup>				$R_{nrdb/rrs}^z$
	HLBas	CLas-4G	RNR	18S	
10 <sup>0</sup>	<b>38.19</b>	<b>39.78</b>	40.00	13.25	1.05
10 <sup>-1</sup>	<b>39.76</b>	40.00	40.00	16.24	1.01
10 <sup>-2</sup>	40.00	40.00	40.00	19.66	NC
10 <sup>-3</sup>	40.00	40.00	40.00	23.37	NC
10 <sup>-4</sup>	40.00	40.00	40.00	28.97	NC
10 <sup>-5</sup>	40.00	40.00	40.00	33.56	NC
10 <sup>-6</sup>	40.00	40.00	40.00	34.64	NC
10 <sup>-7</sup>	40.00	40.00	40.00	35.56	NC

<sup>y</sup> HLB = Huanglongbing, HLBas = HLBas-PCR (HLBas/HLBp/HLBr), CLas-4G = CLas-4G-PCR (CLas-4G/HLBp/HLBr), RNR = ribonucleotide reductase-PCR (RNRf/RNRp/RNRr), and 18S = 18S-PCR (18Sf/18Sr) based on Yan et al. (2012) as the plant DNA internal control. Numbers are PCR cycle threshold (Ct) values. Ct values at the highest dilutions are in bold.

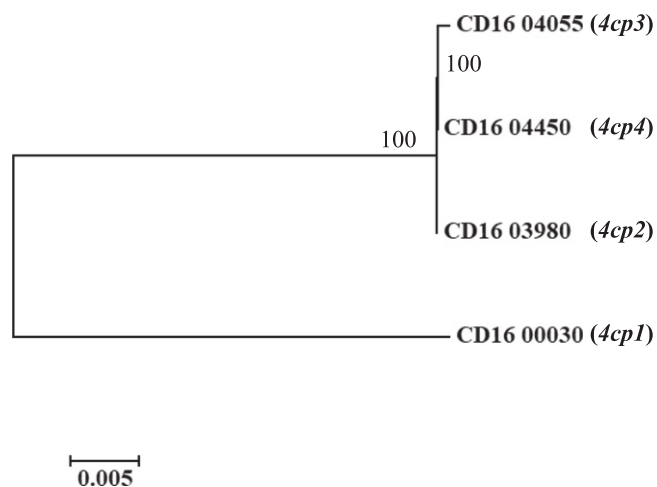
<sup>z</sup>  $R_{nrdb/rrs}$  = Ct value ratio of RNR-PCR/HLBas-PCR. When both PCR Ct values were 40.00,  $R_{nrdb/rrs}$  was not calculated (NC).

4cp loci were 100% identical. It should be noted that, initially, locus 4cp4 showed SNP/nucleotide insertion at the 3’ end of primer 4CP. Further read-mapping experiments indicated that the SNP/nucleotide insertion was attributable to sequence assembling errors (Supplementary Fig. S1). BLAST search against the GenBank database for the genetic nature of the four 4cp ORFs did not identify any significant match, other than hypothetical proteins in the genomes of CLAs strains. A total of 59 SNPs were present among the four 4cp ORFs. Phylogenetically, CD16\_00030 (4cp1) was distinct from the other three ORFs. Within the three ORFs, CD16\_03980 (4cp2) and CD16\_04450 (4cp4) were more closely related to each other than to CD16\_04055 (4cp3) (Fig. 2).

**Evaluation of 4CP-PCR.** A total of 107 DNA samples (94 citrus and 13 ACP) collected from China were tested by 4CP-PCR, along with comparisons with other PCR systems (Fig. 3). No significant difference was observed between 4CP-PCR and RNR-PCR. However, 4CP-PCR was significantly different (more sensitive) from the two *rrs*-based HLBas-PCR and CLas-4G-PCR.

## Discussion

The uniqueness of oligonucleotide OI1 plays a key role in the species description of CLAs (Jagoueix et al. 1994). The fact that OI1 has a missing G raises the issue of taxonomic integrity of CLAs. For this reason, OI1-G is proposed to replace OI1. This is also important for metagenomics research of HLB. In typical metagenomics research, a citrus tissue or ACP sample suspected of being CLAs infected is



**Fig. 2.** Phylogenetic relationships of sequences of open reading frames in the four 4cp loci of ‘*Candidatus Liberibacter asiaticus*’ strain A4. The scale bar is the rate of nucleotide substitutions. Numbers at nodes are bootstrap values (1,000 replications).

	Start	4CPf	
4cp1 (CD16_00030) 4764	atgTTT~~~~~	<u>GCCAAAAGTCTTGCACGTAA</u> TGTTATAACTAATTGGTTT	3648
4cp2 (CD16_03980) 885149	atgTTT~~~~~	<u>GCCAAAAGTCTTGCACGTAA</u> TGTTATAACTAATTGGTTT	886342
4cp3 (CD16_04055) 909353	atgTTT~~~~~	<u>GCCAAAAGTCTTGCACGTAA</u> TGTTATAACTAATTGGTTT	910546
4cp4 (CD16_04450) 1000031	atgTTT~~~~~	<u>GCCAAAAGTCTTGCACGTAA</u> TGTTATAACTAATTGGTTT	1001224
MetPhe		AlaLysSerLeuAlaArgAsnValIleThrAsnTrpPhe	
		4CPp	Stop 4CPr-rc
4cp1 (CD16_00030) 3649	AG	<u>TCAGTTTACAATTACTGTGGTCTGA</u> TAGC	<u>TAGAGGCGGTAAGTTTTTCGC</u> GAG 3522
4cp2 (CD16_03980) 886343	AG	<u>TCAGTTTACAATTACTGTGGTCTGA</u> TAGC	<u>TAGAGGCGGTAAGTTTTTCGC</u> GAG 886392
4cp3 (CD16_04055) 910547	AG	<u>TCAGTTTACAATTACTGTGGTCTGA</u> TAGC	<u>TAGAGGCGGTAAGTTTTTCGC</u> GAG 910596
4cp4 (CD16_04450) 1001225	AG	<u>TCAGTTTACAATTACTGTGGTCTGA</u> TAGC	<u>tagAGGCGGTAAGTTTTTCGC</u> GAG 1001274
		SerGlnPheThrIleThrValValAspSer***	

**Fig. 1.** Alignments of selected regions of the four 4cp sequences addressing open reading frames (ORFs; in parentheses) and 4CP-PCR primers and probe sequences (underlined) in the whole genome sequences (WGS) of ‘*Candidatus Liberibacter asiaticus*’ strain A4 (CP010804.1). To note, 4CPr-rc is the reverse complement of 4CPf (GCGAAACTTACGCCTCTA). Position numbers were based on A4 WGS. The start and stop positions of ORFs are indicated. The tilde symbol (~) indicates omitted nucleotides. Lowercase letters (also underlined with an asterisk) at the beginning of 4CPr-rc of 4cp4 are corrected nucleotides from the current version of A4 WGS.

collected. Total DNA is extracted and sequenced through NGS, such as Illumina MiSeq or HiSeq. One technique to detect the presence of CLAs is to BLAST search NGS reads for those matching with OI1. Because of the missing G, the OI1 sequence could not match the target CLAs reads with 100% identity, resulting in a false-negative detection.

Although primer HLBas has a missing G, HLBas-PCR, with the increase of 1 to 2 Ct value, did not seem to significantly affect the positive CLAs determination for samples with high titers, such as Ct values < 30 (Fig. 3; Table 2). Such high titers (low Ct values) are commonly associated with citrus samples showing HLB symptoms. In HLB endemic areas such as Guangdong, China, HLB is often diagnosed through symptomatology and confirmed by HLBas-PCR for the presence of CLAs. Yet, for academic integrity, the hidden error of HLBas should be corrected. CLAs-4G is proposed to replace HLBas based on the results of this study, although other primers (CLAs-1G, CLAs-6G, and CLAs-7G) could also be used (Table 2). As metagenomics research becomes more popular, PCR primer sequences are often used for CLAs identification. Like OI1-G, the use of CLAs-4G will help to avoid false-negative results.

Vigorous HLB surveys are commonly performed for non-HLB endemic areas such as California or for enforcing the use of CLAs-free nursery stock in both HLB and non-HLB endemic areas. Asymptomatic samples with high Ct values near the threshold (e.g., Ct = 35) will likely be encountered. As shown in the dilution experiments with true low titer CLAs samples, the presence of CLAs could be determined at Ct values of 39.25 by HLBas-PCR (Table 3; Supplementary Tables). The true CLAs infection status was confirmed by RNR-PCR with a slightly lower Ct of 38.72 (or  $R_{nrdb/rns} = 0.94$  or <1). It should be noted that the interpretation of a Ct of 39.25 as CLAs positive is limited to the experimental conditions in this study and the Ct threshold for a CLAs-positive call could vary among different laboratories. In contrast, true CLAs-negative samples (from Chongqing, China and Parlier, California) yielded either no Ct (recorded as 40 in this study) or high Ct values (38.19; Table 4; Supplementary Tables). The addition of RNR-PCR resulted in higher Ct values. Therefore, it is recommended that RNR-PCR be used to resolve interpretation problems from high Ct values by HLBas-PCR or CLAs-4G-PCR.

One question that could be asked is “What about a sample with  $R_{nrdb/rns} = 1$ ?” This is theoretically possible at least in the context of a coexistence of CLAs with unknown but CLAs-related bacteria in a citrus or ACP sample. In such a situation, the CLAs infection status of a sample remains undetermined. The use of 4CP-PCR could be a

good reference in resolving the problem. However, we have not yet observed any sample with  $R_{nrdb/rns} = 1$ . On the other hand, this also raises the issue of the importance of HLB metagenomic research. Metagenomic information on CLAs-infected citrus/ACP that could be used to assist CLAs detection is currently very limited or nonexistent.

One suggestion could be to use RNR-PCR to replace HLBas-PCR for CLAs detection. This replacement is further supported by the conserved nature of the *nrdb* gene (Zheng et al. 2016). It should be noted that the reason *rns* is used for bacterial taxonomy is because *rns* is involved in protein synthesis, a process shared by all prokaryotes (Woese and Fox 1977). Gene *nrdb* plays a key role in ribonucleotide synthesis during DNA replication, a function even more conserved than protein synthesis. However, historically, CLAs, as well as many other bacteria, has been defined by *rns* sequences. Databases of bacterial *rns* sequences are well established. On the other hand, the *nrdb* database remains limited. More WGSs of CLAs strains from different regions around the world will help to broaden the benefit of using RNR-PCR for HLB management.

Detection of CLAs is hampered by the limited available biological information. WGSs of CLAs could be acquired from infected ACPs using NGS (Duan et al. 2009; Wu et al. 2015b) and from HLB-affected plants (Wu et al. 2015a; Zheng et al. 2014a, b) (Table 1). WGS analyses provided a significant amount of biological information about CLAs, filling in the gap caused by the unculturable nature of CLAs. It is important to quickly use the new CLAs information to enhance current HLB management, such as the development of RNR-PCR (Zheng et al. 2016) and 4CP-PCR. As shown in Figure 3, 4CP-PCR was as sensitive as RNR-PCR, and therefore is a good candidate for cross-checking RNR-PCR when needed. This will help to ensure the reliability of using RNR-PCR for CLAs detection. A disadvantage of 4CP-PCR is that unlike *nrdb*, the genetic nature of the *4cp* locus remains unclear.

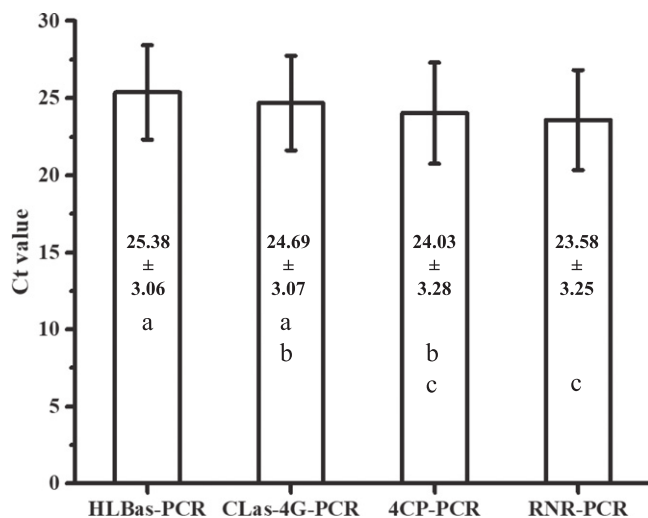
In summary, this study took advantage of the available NGS and Sanger sequencing database of CLAs to detect and/or confirm a sequence error of missing a nucleotide G in OI1, the signature oligo-sequence of CLAs, and in HLBas, the key primer of a commonly used PCR system for CLAs detection. The impact of the missing G in HLBas-PCR for CLAs detection was insignificant in samples with a high CLAs titer. Yet it could not distinguish the low CLAs titer from no CLAs in clinical citrus and ACP samples. The incorporation of RNR-PCR based on five-copy *nrdb* helped to resolve this problem. Finally, a 4CP-PCR system based on four-copy *4cp* was developed. 4CP-PCR had similar detection sensitivity to CLAs as RNR-PCR and offers an additional alternative to verifying diagnostic results by regulatory agencies. Altogether, this study demonstrated the importance and benefits of using WGS information to improve current CLAs detection and therefore HLB management.

## Acknowledgments

We thank the U.S. Department of Agriculture Forest Service International Programs for administrative assistance. We also thank Y. Andrade for technical assistance.

## Literature Cited

- Bové, J. M. 2006. Huanglongbing: A destructive, newly-emerging, century-old disease of citrus. *J. Plant Pathol.* 88:7-37.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
- Dai, Z., Wu, F., Zheng, Z., Yokomi, R., Kumagai, L., Cai, W., Rascoe, J., Polek, M., Chen, J., and Deng, X. 2019. Prophage diversity of ‘*Candidatus Liberibacter asiaticus*’ strains in California. *Phytopathology* 109:551-559.
- Duan, Y., Zhou, L., Hall, D. G., Li, W., Doddapaneni, H., Lin, H., Liu, L., Vahling, C. M., Gabriel, D. W., Williams, K. P., Dickerman, A., Sun, Y., and Gottwald, T. 2009. Complete genome sequence of citrus huanglongbing bacterium, ‘*Candidatus Liberibacter asiaticus*’ obtained through metagenomics. *Mol. Plant Microbe Interact.* 22:1011-1020.
- Jagueix, S., Bové, J. M., and Garnier, M. 1994. The phloem-limited bacterium of greening disease of citrus is a member of the  $\alpha$  subdivision of the Proteobacteria. *J. Syst. Bacteriol.* 44:379-386.
- Jagueix, S., Bové, J. M., and Garnier, M. 1996. PCR detection of the two ‘*Candidatus*’ liberibacter species associated with greening disease of citrus. *Mol. Cell. Probes* 10:43-50.



**Fig. 3.** Comparison of PCR detection sensitivities on ‘*Candidatus Liberibacter asiaticus*’ (CLAs) using 107 samples from China (94 citrus, 13 Asian citrus psyllids) among different TaqMan PCR systems. HLBas-PCR = HLBas/HLBp/HLBr, CLAs-4G-PCR = CLAs-4G/HLBp/HLBr, 4CP-PCR = 4CPf/4CPp/4CPr, and RNR-PCR = RNRf/RNRp/RNRr. Numbers within each bar box are cycle threshold (Ct) means  $\pm$  standard deviations. Letters below the Ct means indicate a significant difference at  $\alpha = 0.01$  from Duncan’s multiple range test. HLB = Huanglongbing and RNR = ribonucleotide reductase.

- Kim, J. S., and Wang, N. 2009. Characterization of copy numbers of 16S rDNA and 16S rRNA of *Candidatus Liberibacter asiaticus* and the implication in detection in planta using quantitative PCR. *BMC Res. Notes* 2:37.
- Kumagai, L. B., LeVesque, C. S., Blomquist, C. L., Madishetty, K., Guo, Y., Woods, P. W., Rooney-Latham, S., Rascoe, J., Gallindo, T., Schnabel, D., and Polek, M. 2013. First report of *Candidatus Liberibacter asiaticus* associated with citrus Huanglongbing in California. *Plant Dis.* 97:283.
- Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33:1870-1874.
- 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. *J. Microbiol. Methods* 66:104-115.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., and Thompson, J. D. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539.
- Tsai, C. H., Su, H. J., Liao, Y. C., and Hung, T. H. 2006. First report of the causal agent of Huanglongbing (*Candidatus Liberibacter asiaticus*) infecting kumquat in Taiwan. *Plant Dis.* 90:1360.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115.
- Woese, C. R., and Fox, G. E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74:5088-5090.
- Wu, F., Jiang, H., Beattie, G. A., Holford, P., Chen, J., Wallis, C. M., Zheng, Z., Deng, X., and Cen, Y. 2018. Population diversity of *Diaphorina citri* (Hemiptera: Liviidae) in China based on whole mitochondrial genome sequences. *Pest Manag. Sci.* 74:2569-2577.
- Wu, F., Kumagai, L., Liang, G., Deng, X., Zheng, Z., Keremane, M., and Chen, J. 2015a. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from a citrus tree in San Gabriel, California. *Genome Announc.* 3:e01508-e01515.
- Wu, F., Zheng, Z., Deng, X., Cen, Y., Liang, G., and Chen, J. 2015b. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from *Diaphorina citri* in Guangdong, China. *Genome Announc.* 3:e01316-15.
- Yan, J., Yuan, F., Long, G., Qin, L., and Deng, Z. 2012. Selection of reference genes for quantitative real-time RT-PCR analysis in citrus. *Mol. Biol. Rep.* 39:1831-1838.
- Zheng, Z., Bao, M., Wu, F., Van Horn, C., Chen, J., and Deng, X. 2018a. A type 3 prophage of '*Candidatus Liberibacter asiaticus*' carrying a restriction-modification system. *Phytopathology* 108:454-461.
- Zheng, Z., Chen, J., and Deng, X. 2018b. Historical perspectives, practical managements and research advancements of citrus HLB in southern China, where the disease has been endemic for over a hundred years. *Phytopathology* 108:1224-1236.
- Zheng, Z., Deng, X., and Chen, J. 2014a. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from California. *Genome Announc.* 2:e00999-14.
- Zheng, Z., Deng, X., and Chen, J. 2014b. Whole-genome sequence of '*Candidatus Liberibacter asiaticus*' from Guangdong, China. *Genome Announc.* 2:e00273-14.
- Zheng, Z., Wu, F., Kumagai, L. B., Polek, M., Deng, X., and Chen, J. 2017. Two '*Candidatus Liberibacter asiaticus*' strains recently found in California harbor different prophages. *Phytopathology* 104:158-162.
- Zheng, Z., Xu, M., Bao, M., Wu, F., Chen, J., and Deng, X. 2016. Unusual five copies and dual forms of *nrdB* in '*Candidatus Liberibacter asiaticus*': Biological implications and PCR detection application. *Sci. Rep.* 6:39020.