A Type 3 Prophage of ‘*Candidatus Liberibacter asiaticus*’ Carrying a Restriction-Modification System

Zheng Zheng, Minli Bao, Fengnian Wu, Christopher Van Horn, Jianchi Chen,[†] and Xiaoling Deng[†]

First, second, third, and sixth author: Department of Plant Pathology, South China Agricultural University, Guangzhou, Guangdong, China; and first, fourth, and fifth authors: San Joaquín Valley Agricultural Sciences Center, Parlier, CA.
Accepted for publication 26 November 2017.

ABSTRACT

Prophages, the lysogenic form of bacterial phages, are important genetic entities of ‘*Candidatus Liberibacter asiaticus*’ (CLAs), a nonculturable α -proteobacterium associated with citrus Huanglongbing. Two CLAs prophages have been described, SC1 (NC_019549.1, Type 1) and SC2 (NC_019550.1, Type 2), which involve the lytic cycle and the lysogenic cycle, respectively. To explore the prophage repertoire, 523 CLAs DNA samples extracted from leaf petioles of CLAs-infected citrus were collected from southern China and surveyed for Type 1 and Type 2 prophages by specific PCR. Eighteen samples were found lacking both prophages. One sample, JXGC, sequenced using Illumina HiSeq, generated >100 million short sequence reads (150 bp per read). Read mapping to known prophage sequences showed a sequence coverage of 46% to SC1 and 50% to SC2. BLAST search using SC1 and SC2 as queries identified three contigs from

the JXGC de novo assembly that form a circular P-JXGC-3 (31,449 bp), designated as a new Type 3 prophage. Chromosomal integration of P-JXGC-3 was detected to occur within a helicase gene, resulting in a duplication of this gene. P-JXGC-3 had 36 open reading frames (ORFs), 10 of which were not found in Type 1 or Type 2 prophages, including four genes that encoded a restriction-modification (R-M) system (*hsdR*, *hsdS*, *hsdM1*, and *hsdM2*). Typed by prophage-specific PCR, the CLAs strains in southern China contained all combinations of the three prophage types with the exception of a Type 2–Type 3 combination, suggesting active ongoing prophage–phage interactions. Based on gene annotation, P-JXGC-3 is not capable of reproduction via the lytic cycle. The R-M system was speculated to play a role against Type 1 prophage–phage invasion.

Prophages, the lysogenic form of a bacterial phage with its DNA inserted into the chromosome, are important genetic elements of the bacterial genome and play critical roles in bacterial evolution, bacterial cell defense, and environmental adaptation including pathogenesis (Boyd and Brüssow 2002; Feiner et al. 2015). In ‘*Candidatus Liberibacter asiaticus*’ (CLAs), an unculturable α -proteobacterium associated with citrus Huanglongbing (HLB, yellow shoot disease), two types of prophages, Type 1, represented by prophage SC1, and Type 2, represented by prophage SC2, have been described (Zhang et al. 2011; Zheng et al. 2016). Research data suggested that SC1 was involved in the lytic cycle of forming phage particles and SC2 was involved in the lysogenic conversion of CLAs pathogenesis (Fleites et al. 2014; Jain et al. 2015; Zhang et al. 2011).

Among the currently published whole genome sequences of CLAs from different geographical regions around the world (Duan et al. 2009; Katoh et al. 2014; Kunta et al. 2017; Lin et al. 2013; Wu et al. 2015a, b; Zheng et al. 2014a, b, 2015). Eight CLAs genomes contain extensive prophage sequences. A CLA strain (Ishi-1, AP014595.1) from Japan was reported to show lack of prophage. Based on PCR experiments with

primer sets specific to detect Type 1 and Type 2 prophages, CLAs strains lacking both prophages were also detected in southern China (Zheng et al. 2016). It was unclear if these strains simply lacked prophages or possibly harbored unknown prophages. There have not yet been extensive studies to describe the repertoire of CLAs prophages.

The presence of more than one prophage in CLAs raised a question of possible prophage–phage interactions. An analysis on CLAs strains collected in southern China revealed the predominance of a single prophage (Type 1 or Type 2) in the region (Zheng et al. 2016). A CRISPR (clustered regularly interspaced short palindromic repeats)/*cas* system was reported in both Type 1 and Type 2 prophages and was predicted to be involved in bacterial defense by cleaving invading phage DNA based on spacer sequence(s) (Zheng et al. 2016). Another bacterial defense system is the restriction-modification system (R-M system). R-M systems are equipped with an endonuclease that cleaves any foreign DNA containing a specific recognition site and a methyltransferase that protects host DNA by modifying specific nucleotides (Luria and Human 1952; Nathans and Smith 1975; Wilson 1991). In the CLAs genome, R-M systems had only been annotated in chromosomal regions (Duan et al. 2009; Zheng et al. 2014a) but not in prophage regions.

To explore the CLAs prophage repertoire, we collected over 500 CLAs DNA samples in southern China and surveyed for the presence of new prophages. CLAs samples without Type 1 and Type 2 prophages were found and one sample (named JXGC) from Jiangxi Province with relatively high CLAs titer was selected for next-generation sequencing. Sequence analyses of strain JXGC revealed the presence of a new prophage that carried a previously unknown R-M system. A survey for prophage prevalence in southern China revealed evidence of active prophage–phage interactions in the CLAs population.

[†]Corresponding authors: J. Chen; E-mail: Jianchi.chen@ars.usda.gov, and X. Deng; E-mail: xldeng@scau.edu.cn

Funding: This study was funded by Chinese Modern Agricultural Technology Systems (CARS-27), the Special Fund for Three-High Agriculture in Guangdong Province, China (F15070), the National Key Research and Development Program of China, and the California Citrus Research Board.

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 is an equal opportunity provider and employer.

*The e-Xtra logo stands for “electronic extra” and indicates that six supplementary figures and four supplementary tables are published online.

MATERIALS AND METHODS

Survey of CLAs prophage type in south China. CLAs samples were collected from HLB-affected citrus trees in nine

provinces in southern China (Fig. 1) between March 2013 and November 2016. All CLAs-infected samples were collected from a total of seven different citrus varieties/species, including *Citrus sinensis*, *C. grandis*, *C. reticulata*, *C. Limon*, *C. tangerine*, *C. sunki*, and *C. aurantium*. As a negative control, citrus samples were also collected from central California where no HLB has been found yet. Each sample was collected from a different citrus tree.

DNA was extracted from each sample following the procedure described previously (Zheng et al. 2014a, b). Briefly, 0.2 g of citrus leaf midribs was used for DNA extraction by using E.Z.N.A. HP Plant DNA Kit (OMEGA Bio-Tek Co., Guangdong, China). Presence of CLAs was confirmed by PCR with primer set OI1-OI2c (Jagoueix et al. 1994) and/or TaqMan Real-time PCR developed by Li et al. (2006). A DNA sample extracted from the leaves of a single tree was considered a CLAs strain. All samples were subject to prophage-typing based on the procedure of Zheng et al. (2014a). Briefly, primer set T1-2F/T1-2R and T1-3F/T1-3R specific to Type 1 prophage, and primer set T2-2F/T2-2R and T2-3F/T2-3R specific to Type 2

prophage were used for PCR (Table 1). CLAs strains were then classified into four groups: Type 1, Type 2, both Type 1 and Type 2, and neither Type 1 nor Type 2.

Standard PCR was performed on a Bio-Rad T100 Thermal Cycler (Bio-Rad, Hercules, CA). The 25-μl reaction mixture containing 1 μl of DNA template, 0.4 μl of *Taq* DNA polymerase at 2.5 U/μl (Tiangen Biotech Co., Beijing), 2.5 μl of 2.5 mM deoxynucleotide triphosphates (dNTPs), 2.5 μl of 10× DNA polymerase buffer, 0.5 μl of each forward and reverse primer (10 μM), and 17.6 μl of ddH₂O. PCR was performed under the following procedure: initial denaturation at 96°C for 5 min, 35 cycles of amplification (94°C for 30 s, 60°C for 30 s, and 72°C for 60 s) and ended with a final extension of 72°C for 10 min. The PCR products were electrophoresed in 1% agarose gels (0.5× TBE buffer) and visualized by Goldview (Guangzhou Genesun Biotech Ltd., China) under UV illumination.

Whole genome sequencing and assembly. Strain JXGC, lacking both Type 1 and Type 2 prophages, was selected for whole

Fig. 1. Prophage typing on 523 strains of ‘*Candidatus Liberibacter asiaticus*’ (CLAs) from nine provinces in southern China. Within brackets are CLAs strain numbers. The four numbers divided by forward slashes are CLAs strains harbored Type 1/Type 2/Type 1 + Type 2/neither Type 1 nor Type 2 prophage. The location where strain JXGC was collected is marked by a dot, along with three symptomatic Huanglongbing leaves used for DNA extraction.

TABLE 1. General information of prophage type-specific primer sets and 16S rDNA-specific primer sets used in this study

Name	Sequence 5'-3'	Size (bp)	Target region	Specificity	Reference
891-1F	CTGATCCTTTACCATGCCGC	950	<i>hsdS</i>	Type 3 prophage	This study
891-1R	CAGCGAAACCGATCTTGAGG				
891-2F	ACCGGATCTACCGTAATT	884	<i>hsdR</i>	Type 3 prophage	This study
891-2R	TGTGTTTTGCGAGTGAAGGG				
T1-2F	TGGCTCGGTTTCAGGTAAAT	975	Endolysin	Type 1 prophage	Zheng et al. 2016
T1-2R	AAGGGCGACGCATGTATTTC				
T1-3F	CTCACTGCGTCTTGATTCGG	866	Hypothetical protein gene	Type 1 prophage	Zheng et al. 2016
T1-3R	CGAACGAGCGGTATGTTTGT				
T2-2F	ACCCTCGCACCATCATGTTA	813	Phage structural protein gene	Type 2 prophage	Zheng et al. 2016
T2-2R	TCGTCTTGATTGGGCAGAGT				
T2-3F	ACAGTTAAGAGCCACGGTGA	918	Hypothetical protein gene	Type 2 prophage	Zheng et al. 2016
T2-3R	AAGACGTGGGTGTTATGGGT				
OI1	GCGCGTATGCAATACGAGCGGCA	1167	16S rRNA gene	CLAs chromosomal region	Jagoueix et al. 1994
OI2c	GCCTCGCGACTTCGCAACCCAT				
HLBas	TCGAGCGGTATGCAATACG	75	16S rRNA gene	CLAs chromosomal region	Li et al. 2006
HLBar	GCGTTATCCCGTAGAAAAAGGTAG				

genome sequencing. The citrus tree (*C. reticulata* Blanco ‘Ponkan’) infected by strain JXGC showed typical symptoms of HLB, and was located in an orchard in Guangchang City of Jiangxi Province, China. JXGC DNA was extracted from the midrib of symptomatic leaves (Fig. 1). CLAs was quantified by the procedure of Li et al. (2006) with primer set HLBas/HLBr yielding a cycle threshold (Ct) value of 19.25. The draft whole genome sequence of strain JXGC was obtained according to a previously developed procedure (Zheng et al. 2014a). Briefly, bacterial DNA was enriched using the NEB-next microbiome DNA enrichment kit (New England BioLabs, Inc., Ipswich, MA) and enlarged through multiple displacement amplification using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences, Chicago, IL) according to the manufacturer’s instruction. Genome sequencing was carried out on a HiSeq platform (Illumina, Inc., San Diego, CA) through a commercial source. De novo assembling of prophage sequence was performed with Velvet (v 1.2.10) (Zerbino and Birney 2008) and CLAs chromosomal region sequence was assembled using the CLAs A4 genome as a reference through CLC Genomic Workbench 7.5. Genome annotation was performed on the RAST server (<http://rast.nmpdr.org/>) (Aziz et al. 2008).

Prophage identification. To evaluate the prophage status of strain JXGC, HiSeq reads were mapped to prophage SC1 (Type 1) and SC2 (Type 2) (Zhang et al. 2011) downloaded from the GenBank database. To search and obtain the possible presence of prophage distantly related to Type 1 and Type 2 prophages, de novo assembly of JXGC HiSeq data were performed using CLC genomic workbench 7.5. All de novo contigs were evaluated through BLASTn using SC1 and SC2 as queries (word_size = 28, e-value = 1e-5). Contigs with high score (>5,000) and >10,000 kb were selected. The targeted contigs were manually collected.

A major prophage contig was used as a query to search for connected contigs from de novo assembly (word_size = 16, e-value = 10). Contigs with overlaps of >20 bp and >90% identity at either ends were selected for sequence extensions. When a contig connecting both 5′ and 3′ end of the extending prophage sequence was found, the circular form, also called plasmid, was determined. When contigs connected the extending prophage sequence to the bacterial chromosome at both 5′ and 3′ ends, the lysogenic form was determined. Overlap sequences were further confirmed by the PCR amplifications and Sanger’s sequencing. Primer 3 software (Untergasser et al. 2012) was used for primer designs. PCR amplicons were sequenced after cloning in *pEASY-T1* plasmid (TransGen Biotech, Beijing) or directly. Contigs were assembled by using SeqMan software (<http://www.dnastar.com>). Whole genome sequences of CLAs strain A4 (Zheng et al. 2014a) containing a Type 2 prophage and strain Psy62 (CP001677.5) containing a Type 1 prophage were used to assist prophage contig mapping (by BLASTn). To further evaluate the circularity of the JXGC phage, the first 5,000 bp at the 5′ end of JXGC prophage was cut and attached to the 3′ end to generate a new prophage sequence, namely P-JXGC-3-B_A. For comparison, a 5,000-bp sequence randomly selected from the chromosomal region of CLAs JXGC genome was cut and attached to the end of the JXGC prophage (without the first 5,000 bp) that generated another sequence to serve as the reference, namely P-JXGC-3-B_C. Both P-JXGC-3-B_A and P-JXGC-3-B_C sequences were mapped with HiSeq reads.

The integration sites in the chromosome were identified by aligning the circular prophage sequence to the whole genome sequence of strain JXGC. Centered by the integration sites, genes/ORFs up- and down-stream were compared according to annotations and BLAST comparisons to estimate similarity at both nucleotide and amino acid sequence levels. Genes located in the integration region of the CLAs strain UF506 (HQ377374.1), Ishi-1 (AP014595.1), and psy62 (CP001677.5) were obtained based on the genome annotation. The CRISPR repeats array in the prophage region of different CLAs strain were identified following the protocol in the previous study (Zheng et al. 2016). Sequence alignment was performed in the Clustal Omega software (Sievers et al. 2011).

Prophage gene characterization. All annotated genes from the JXGC prophage were used as a query to search against Type 1 and Type 2 prophage sequences by using BLAST software (Camacho et al. 2009). Genes showing low similarity (<50% coverage with <50% similarity) with others prophage sequences were defined as the unique genes of the JXGC prophage. The genes associated with the R-M system in the prophage sequence were further analyzed by BLASTx searching against the NCBI conserved domain database (CDD, v3.15) and the Restriction Enzyme Database (<http://rebase.neb.com/rebase/rebase.html>) (Roberts et al. 2015). To compare the R-M system between the prophage region and chromosomal region of CLAs JXGC, the R-M genes from the chromosomal region were identified based on the annotation and BLASTx against the CDD database. The R-M genes were named based on conserved domain name or based on the sequence region (chromosomal or prophage region) in which they were discovered. For example, HsdR meant a restriction subunit gene that harbored a conserved domain of HsdR and HsdM I_C, where C denoted chromosome, meant the modification subunit gene that was found in the chromosomal region of CLAs.

The similarity of nucleotide and protein sequence of all R-M genes were assessed through BLAST analysis. Domain and interaction sites were obtained from the Restriction Enzyme Database (<http://rebase.neb.com/rebase/rebase.html>) (Roberts et al. 2015). The protein structure analyses were initially predicted in the Phyre server (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) using a profile-profile alignment algorithm (Kelley et al. 2015). Final three-dimensional (3D) structures were optimized by using the Pymol Molecular Graphics System (v1.7.6, Schrödinger LLC).

Survey of prophages prevalence. A subset of 240 CLAs samples was selected from the total 523 samples, with a consideration of maintaining the same geographical distribution, to test for the presence of prophages simultaneously in each sample using prophage type-specific primer sets (Table 1). For PCR identification of Type 1 and Type 2 prophages, each of two stable primer sets were selected as the type-specific primers based on our previous study (Zheng et al. 2016). For PCR identification of Type 3 prophage, all Type 3-specific primer sets listed in Supplementary Table S1 showed stable amplification in JXGC DNA samples without any nonspecific amplification. Therefore, two primer sets, 891-1F/891-1R and 891-2F/891-2R, targeting the *hsdS* gene and *hsdR* genes, respectively, were selected as the Type 3-specific primers for PCR identification of Type 3 prophage. Presence of all combinations of prophage types were recorded and mapped to the geographical origins.

RESULTS

Prophage survey. A total of 523 CLAs-infected citrus DNA samples, with Ct values ranging from 17 to 31 based on the method of Li et al. (2006), were collected from southern China. From these samples, 107 (20.46%) samples harbored only the Type 1 prophage, 362 (69.22%) samples harbored only the Type 2 prophage, 36 (6.88%) samples harbored both Type 1 and Type 2 prophages, and 18 (3.44%) samples had neither Type 1 nor Type 2 prophages (Fig. 1). A subset of 86 samples (22 Type 1, 60 Type 2, two both Type 1 and Type 2, and two non-Type 1 and Type 2) were further tested and confirmed by PCR experiments with additional primer sets (T1-1F/T1-1R, Type 1-specific, and T2-1F/T2-1R Type 2-specific). After partitioning into the geographical origins, a pattern similar to our previous publication (Zheng et al. 2016) was observed: a higher ratio of Type 1 prophage was present in strains collected from west China (e.g., Yunnan), a higher ratio of Type 2 prophage was present in strains collected from eastern China (e.g., Guangdong), and a lower ratio of both prophages overall. Among the 18 non-Type 1-Type 2 samples, no clear geographical skew was observed (Fig. 1). To further evaluate the status of prophage occurrences in these 18 samples, one sample (named JXGC) from Jiangxi Province

with relatively high CLas titer was selected for next-generation sequencing.

New prophage P-JXGC-3. As with all samples in this study, the citrus tree (*C. reticulata*) from which sample JXGC was collected showed typical HLB symptoms of mottling and yellowing (Fig. 1). HiSeq sequencing of JXGC DNA generated a total of 103,149,836 reads with average length of 150 bp. Mapping of the HiSeq reads to SC1 and SC2 sequences showed partial coverage to SC1 (46%) and to SC2 (50%) (Fig. 2, blue shadowed region). These mapping results suggested the possible presence of an undescribed prophage in sample JXGC. The de novo assembly of JXGC HiSeq

reads resulted in 37,098 contigs ≥ 1000 bp in length. Using SC1 and SC2 as references, Contig_891 (27,860 bp) showed a BLAST bit score result of $>10,000$, which was significantly greater than that of the next BLAST hit (Contig_12853) with a bit score of 2,470. Contig_891 partially covered SC1 (52.9%) and SC2 (53.5%), in agreement with the HiSeq read mapping results.

Sequence extension of Contig_891 based on a BLAST search against the de novo assembly identified two contigs, Contig_30437 (1,464 bp) connected to the 5' end of Contig_891, and Contig_12853 (1,506 bp) connected to the 5' end of Contig_30437. Continuing contig extension led to two results: (i) the 5' end of Contig_12853

Fig. 2. HiSeq read mapping of ‘*Candidatus Liberibacter asiaticus*’ strain JXGC to the sequences of prophages SC1, SC2, and P-JXGC-3. Annotated genes are indicated by the arrow boxes. Numbers above are nucleotide positions. On the left and under each prophage names, Consensus = assembled contigs; Coverage = read coverage from 0 (bottom) to 127 (top). Identical genes ($>85\%$ identity and 90% length coverage) are represented by the same color. The early gene regions of the three prophages are highlighted by a blue shadow. The integration site of P-JXGC-3 into strain JXGC chromosome is marked by a vertical red line. Locations of type-specific primers are marked by black lines along with primer names and orientations.

Fig. 3. Schematic representation of prophage P-JXGC-3. **A**, The circular form (plasmid) assembled by three contigs with primer names and sites for PCR confirmation. **B**, The linear form (chromosomal integration) bordered by two dash lines with assembling contigs, primer names, and sites indicated. The tandem repeat region is indicated by a red box. **C**, Gene arrangements around the two junctions (dash lines) of P-JXGC-3 and chromosome. Blue shadow boxes identify the six gene repeats. DNA sequence repeats are underlined by red lines. HP = hypothetical protein. The gene sequences located in the prophage region are labeled by green arrows and gene sequences located in the chromosomal region are labeled by blue.

connected to the 3' end of Contig_891, forming a circular plasmid (Fig. 3A); and (ii) the HiSeq read mapping test further confirmed the circularity (Supplementary Fig. S1) and Contig_12853 connected to the 3' end of Contig_12375 (13,109 bp) of the CLas chromosome. The 3' end of Contig_891 connected to another copy of Contig_12853, followed by Contig_19293 (1,869 bp) and Contig_20254 (1,480 bp) in the CLas chromosome, i.e., the linear configuration (chromosomal integration) of the prophage (Fig. 3B). The fact that coverage of Contig_12853 (151.09 X) was nearly twice that of Contig_891 (66.94 X) also supported the plasmid/chromosome integration configuration.

For sequence validation, over 20,000 bp of the prophage sequence, including contig junctions, were confirmed by PCR with 25 primer sets and Sanger sequencing of PCR amplicons. The junction between Contig_891 and Contig_30437 contained tandem repeats (Supplementary Fig. S2), explaining the disconnection of the two contigs during de novo assembling. The disconnection between Contig_30437 and Contig_12853 was caused by the repeat region in Contig_12853 that also connected to Contig_19293.

The new prophage was designated as P-JXGC-3 following the nomenclature rule in the previous publication (Zheng et al. 2017). The circular form of plasmid P-JXGC-3 had a size of 31,449 bp and

had been deposited in GenBank under the accession number KY661963. P-JXGC-3 had a G+C content of 39.9% and 36 open reading frames (ORFs). The prophage belongs to a new Type 3 prophage because of the significant difference (~50%) to Type 1 and Type 2 prophages. Of the 36 predicted genes in P-JXGC-3, 10 genes were unique to P-JXGC-3, while 26 genes were highly similar or identical to SC1 or SC2 (Supplementary Table S2). In addition, a CRISPR/cas system found in the early gene region of P-JXGC-3 prophage was similar to a CRISPR locus found in the Type 1 prophage (Supplementary Fig. S3).

Whole genome sequence of strain JXGC. Using the CLas A4 genome as a reference and the described procedure (Zheng et al. 2014b), the whole genome sequence of CLas strain JXGC was determined to have a total of 1,225,162 bp with GC content of 36.4% and 1,118 ORFs. The JXGC whole genome sequence had an average coverage of 72.78 X. The JXGC genome sequence has been deposited in GenBank with the accession number CP019958.

Integration of P-JXGC-3. According to the annotated JXGC genome sequence, P-JXGC-3 integration occurred at position 1,194,198 within ORF B2I23_05410 and ended in position 150, crossing the designated start point of the submitted sequence, within

Fig. 4. Restriction and modification (R-M) systems in ‘*Candidatus Liberibacter asiaticus*’ strain JXGC. **A**, Genes are identified by arrow boxes with loci names above. **B**, Three-dimensional structure of R-M system subunit proteins. In subunit R, purple = P-loop NTPase domain; yellow = N-terminus domain. In subunit M, cyan = N-terminus domain; green = AdoMet-MTase domain. In subunit S, pink = target recognition domain. Protein name and accession ID are labeled under the three-dimensional structure.

ORF B2I23_00005. Both ORFs were 1,386 bp in size and annotated as DNA/RNA helicases (Fig. 3C). The same helicase gene was also found in the P-JXGC-3 prophage (ORF PJXGC_gp33). Alignments of the amino acid sequence of the three helicase genes, as well as the nucleotide sequences of the genes, showed a high level of similarity (Supplementary Fig. S4). Further analyses showed each of the two helicase genes in the JXGC genome were part of a repeat region containing six genes/ORFs (shadowed regions in Fig. 3C). However, at the nucleotide level, only a partial helicase gene and the downstream ligase gene (1,433 bp) were highly similar (>99%) (Supplementary Table S3, red underlines in Fig. 3C). An integrase gene (B2I23_05415) was located 2,124 bp downstream of the prophage in the bacterial chromosomal region (Fig. 3C).

R-M systems. In P-JXGC-3, four genes were annotated as members of a R-M system, *hsdR* encoding a restriction endonuclease, *hsdS* encoding a restriction endonuclease with two target recognition domains and *hsdM1* and *hsdM2* each encoding a DNA-methyltransferase (Fig. 4). All four genes were in the same orientation with the order as *hsdR*-*hsdS*-*hsdM1*-*hsdM2* (Fig. 4). This belongs to the Type 1 R-M system (Murray 2000).

In the JXGC chromosomal region, five R-M system subunit genes in three locations were found (Fig. 4). The five chromosomal genes

had no significant similarity with any of the four P-JXGC-3 R-M system genes at the nucleotide sequence level (Table 2). At the amino acid sequence level, low levels of similarities were detected (Table 2). However, at the 3D structure level, significant structure similarities among some genes were found (Fig. 4). For example, *hsdS_C* (B2I23_03640) and *hsdS* (PJXGC_gp08) from chromosomal and prophage regions, respectively, shared 178 amino acids with only 25% similarity. Yet, the predicted 3D structures of the proteins encoded by the two genes were very similar (Fig. 4B). Among the prophage genes, the 3D structure of the two prophage methyltransferases, *hsdM1* (PJXGC_gp09) and *hsdM2* (PJXGC_gp10), were very similar, although they only had 46% similarity in amino acid sequence and insignificant similarity at the nucleotide level.

Conserved domain analysis through 3D structures indicated the HsdS protein (406 amino acids) contained two target recognition domains (Fig. 4), which can impact target sequence specificity to both the restriction and modification activities of the R-M complex (Murray 2000). Both HsdM1 and HsdM2 had the AdoMet-MTase domains (NCBI CDD Accession: cl17173), which required S-adenosyl-L-methionine as a substrate for methyltransfer (Meselson and Yuan 1968), and a key N-terminal domain (HsdM-N, NCBI CDD Accession pfam12161) shown to be alpha-helical (Fig. 4). It was also revealed

TABLE 2. Pair-wise BLAST search comparisons of restriction-modification (R-M) system genes in chromosomal region (*hsdMI_C*, *hsdMII_C*, *hsdR_C*, *hsdMIII_C*, *hsdM_C*, and *hsdS_C*) and prophage region (*hsdR*, *hsdS*, *hsdM1*, and *hsdM2*) in ‘*Candidatus Liberibacter asiaticus*’ strain JXGC^a

Gene (ID)	Amino acid/ nucleotide acid (bp)	<i>hsdMI_C</i>	<i>hsdMII_C</i>	<i>hsdR_C</i>	<i>hsdMIII_C</i>	<i>hsdS_C</i>	<i>hsdR</i>	<i>hsdS</i>	<i>hsdM1</i>	<i>hsdM2</i>
<i>hsdMI_C</i> B2I23_01705	375/1,128		NA	NA	NA	NA	NA	NA	NA	NA
<i>hsdMII_C</i> B2I23_03460	83/252	NA		NA	15/45 (33%)	NA	NA	NA	20/72 (28%)	NA
<i>hsdR_C</i> B2I23_03465	110/333	NA	NA		NA	NA	19/79 (24%)	NA	NA	18/72 (25%)
<i>hsdMIII_C</i> B2I23_03635	674/2,025	NA	13/14	NA		NA	NA	NA	108/440 (25%)	97/311 (31%)
<i>hsdS_C</i> B2I23_03640	426/1,281	NA	NA	NA	NA		NA	45/178 (25%)	NA	NA
<i>hsdR</i> PJXGC_gp07	1,016/3,051	NA	NA	15/17	NA	NA		NA	NA	NA
<i>hsdS</i> PJXGC_gp08	406/1,221	NA	NA	NA	NA	NA	NA		NA	NA
<i>hsdM1</i> PJXGC_gp09	470/1,413	NA	10/10	NA	12/12	NA	NA	NA		190/417 (46%)
<i>hsdM2</i> PJXGC_gp10	496/1,491	NA	NA	24/30	20/25	NA	NA	NA	106/163	

^a Similarities of amino acid sequences are in upper-right diagonal. Top similarity (matched nucleotide number/total compared nucleotide number) (percentage) are shown; BLAST parameter for protein sequences: word_size = 3, expect threshold = 1,000, gap open existence = 11, gap open extension = 1. Similarities of nucleotide sequences are in lower-left diagonal. Top similarities (matched nucleotide number/total compared nucleotide number) corresponding to amino acid sequence hits are shown. Similarity percentages were not calculated due to short comparable sequence stretches. BLAST parameters for nucleotide sequences: word_size = 7, expect threshold = 1,000, gap open existence = 5, gap open extension = 1, match/mismatch = 2, -3; NA = no similarity value available.

No.	Combination	Samples	Frequency (%)	PCR result				
				M	Type 1	Type 2	Type 3	O11/ O12c
1	Type 1	14	5.8					
2	Type 1 + Type 3	48	20.0					
3	Type 1 + Type 2	20	8.3					
4	Type 1 + Type 2 + Type 3	12	5.0					
5	Type 2	136	56.7					
6	Type 2 + Type 3	0	0.0	Not found				
7	Type 3	10	4.2					
8	None of three types	0	0.0	Not found				
	(Total samples)	240	100.0					

Fig. 5. Frequencies of different prophage combinations in 240 ‘*Candidatus Liberibacter asiaticus*’-infected citrus DNA samples determined by type-specific primer sets. For PCR result, lane M, DNA ladder (top to bottom: 2,000, 1,550, 1,450, 1,000, 750, and 500 bp); and primers used: Type 1, left = T1-2F/T1-2R, Right = T1-3F/T1-3R; Type 2, left = T2-2F/T2-2R, Right = T2-3F/T2-3R; Type 3, left = 891-1F/891-1R, Right = 891-2F/891-2R; 16S = O11/O12c.

that the HsdR protein (1,016 amino acids) contained a N-terminus (HSDR_N, pfam04313) and a P-loop NTPase domain (cl21455) that included an ATP binding site and a putative Mg²⁺ binding site, which were essential for DNA translocation and endonuclease activity (Fig. 4).

Prevalence of different prophage combinations. As shown in Figure 5, from the possible eight combinations of three prophages, six combinations were found. Of the subset of 240 CLas samples (Supplementary Fig. S5), 70 samples (29.17%) harbored the Type 3 prophage: 10 were Type 3 alone, 48 coexisted with Type 1, and 12 coexisted with both Type 1 and Type 2. Noticeably, the combination of Type 2 and Type 3 were not detected. The CLas strain lacking all three prophages was also absent. Presence of a Type 2 prophage alone accounted for the highest proportion (136/240 or 56.67%). The combination of Type 1 and Type 3 accounted for the second highest (48/240 or 20%), higher than that of the Type 1 alone (14/240, or 5.8%). In addition, among six types of CLas strains, no citrus cultivar skew was found, e.g., three different types of CLas strains were identified from 20 HLB-affected lemon trees (*C. limon*) in the same orchard located in the middle region of Hainan Provinces and Type 2 CLas strains were confirmed from the 15 DNA samples extracted from three different citrus species (*C. sinensis*, *C. reticulata*, and *C. limon*) in the same orchard located in Guangxi Province (Supplementary Table S4).

DISCUSSION

With limited knowledge about CLas prophage repertoire, the undetection of Type 1 or Type 2 prophages in a CLas strain was regarded as absence of a prophage (Kato et al. 2014; Zheng et al. 2016). This study initially found that 10 of the 18 CLas samples lacking Type 1 (SC1) and Type 2 (SC2) prophages harbored a Type 3 prophage (P-JXGC-3) (Fig. 5), demonstrating that CLas can harbor other prophages. Further tests on the rest of the eight CLas samples with specific PCR primer sets also detected the presence of the Type 3 prophage (data not shown), i.e., no CLas samples lacking all three prophages were detected in this study. However, we cannot exclude the possibility of more prophages to be found. The scope of this study was limited to southern China. Accurate description of CLas prophage diversity around the world, or even within southern China, requires more studies in the future.

When SC1 and SC2 were first described, each of the two prophages was structurally divided into early gene and late gene regions. The early gene region was highly conserved and associated with DNA replication, and the late gene region was unique and associated with other biological functions such as genes encoding proteins involved in the lytic cycle, e.g., lysozyme and holin in SC1 (Zhang et al. 2011). SC2 did not have lytic cycle genes and was found to be involved in lysogenic conversion (Zhang et al. 2011). Similarly, P-JXGC-3 can also be divided into early gene and late gene regions. The early gene region is highly similar to those of SC1 and SC2. In fact, P-JXGC-3 was found because of the conserved early gene region (Fig. 2). In the early gene region, 22 genes were shared among the three prophages. Two genes were unique to Type 1, and one gene was shared by both Type 1 and Type 2 (Fig. 2). Structures of the three prophages suggested a CLas prophage model consisting of a core (early genes) region, designated as C, and a flexible (late genes) region, designated as F, or prophage P-C_F. In extending this generalization, it is possible that there could be other prophages of CLas with the F possessing genes of different biological functions.

Details on how the prophage integrated into the CLas chromosome were not previously described, although it was implied that the prophage integration was near a guanylate kinase gene (Zhang et al. 2011). By aligning the plasmid and chromosomal sequences, we found that the helicase gene was used as an integration site (Fig. 3C). The exact attachment sites (*attB* and *attP*) could not be determined in this study due to the lack of experimental data. Yet, the integrase gene

downstream is speculated to be involved in the prophage integration process. Interestingly, prophage integration led to a duplication of the helicase gene, rather than a gene inactivation. The same helicase gene-based integration sequence pattern can also be found in the whole genome sequence of other CLas strains (Supplementary Fig. S6). It remains unclear why the helicase gene was used as a site of prophage integration in CLas.

SC2 encoded a virulence factor, peroxidase (SC2_gp095), which suppressed the H₂O₂-mediated defense system and host symptom development (Jain et al. 2015). The homologous gene of SC2_095 was also found in the SC1 prophage (SC1_095). This gene was not found in P-JXGC-3. However, the citrus leaves infected by CLas strain JXGC still showed the typical symptoms (mottling and yellowing) in the late stage (Fig. 1), suggesting that CLas virulence could involve at least some of the six unannotated genes in P-JXGC-3 (other than the R-M system), or chromosome-borne genes.

The most significant finding is the R-M system carried by P-JXGC-3 (Fig. 4). According to literature, most R-M systems are found in chromosomal regions (Oliveira et al. 2014), which is also the case in CLas. The presence of an R-M system in a prophage suggested that the system was recently acquired and likely benefited the CLas host, as is the case in culturable bacteria (Dempsey et al. 2005; Furuta and Kobayashi 2013; Hendrix et al. 1999; Kita et al. 1999, 2003; Kobayashi 2001; Kobayashi et al. 1999). Considering the fact that invading DNA is the target of an R-M system, we speculate that CLas could employ P-JXGC-3 to fight against SC1 phage invasion. The relatively high frequency (second highest) of Type 1–Type 3 prophage association (Fig. 5) supports this hypothesis, i.e., the presence of Type 1 prophage–phage is a positive pressure in selecting CLas cells harboring P-JXGC-3 in the bacterial population. Along the same line, P-JXGC-3 prophage carries a CRISPR/*cas* system with the spacer sequences identical or highly similar to that of Type 1 prophages, which could be a defense mechanism against the lytic Type 1 phage/prophage. However, experimental proofs are needed to confirm these hypotheses. On the other hand, Type 2 prophages were not lytic to CLas. Therefore, there was no selection pressure for the Type 2–Type 3 prophage combination (Fig. 5).

In summary, through the use of next-generation sequencing technology and analyses on a large number of CLas samples collected from the historical HLB region in southern China, a new Type 3 prophage of CLas was described. This prophage carried an R-M system that could be involved in CLas prophage–phage interactions. It should be noted that the exact function of the R-M system in Type 3 prophages has yet to be determined experimentally. In vitro culture of CLas is a key for such experiments, but has not yet been successful. Before the in vitro culture issue is resolved, we continue to explore the use of CLas genome sequences to acquire information of CLas biology to meet the need of HLB research and management.

ACKNOWLEDGMENTS

We thank S. Vargas for technical assistance.

LITERATURE CITED

- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. 2008. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* 9:75.
- Boyd, E. F., and Brüssow, H. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol.* 10:521–529.
- 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.
- Dempsey, R. M., Carroll, D., Kong, H., Higgins, L., Keane, C. T., and Coleman, D. C. 2005. *Sau42I*, a BcgI-like restriction-modification system

- encoded by the *Staphylococcus aureus* quadruple-converting phage Phi42. Microbiology 151:1301-1311.
- 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.
- Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. 2015. A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. Nat. Rev. Microbiol. 13:641-650.
- Fleites, L. A., Jain, M., Zhang, S., and Gabriel, D. W. 2014. '*Candidatus Liberibacter asiaticus*' prophage late genes may limit host range and culturability. Appl. Environ. Microbiol. 80:6023-6030.
- Furuta, Y., and Kobayashi, I. 2013. Restriction-modification systems as mobile epigenetic elements. In: Bacterial Integrative Mobile Genetic Elements. Landes Bioscience, Austin, TX.
- Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., and Hatfull, G. F. 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192-2197.
- Jagoueix, S., Bové, J. M., and Garnier, M. 1994. The phloem-limited bacterium of greening disease of citrus is a member of the α subdivision of the *Proteobacteria*. Int. J. Syst. Bacteriol. 44:379-386.
- Jain, M., Fleites, L., and Gabriel, D. W. 2015. Prophage encoded peroxidase in '*Candidatus Liberibacter asiaticus*' is a secreted effector that suppresses plant defenses. Mol. Plant-Microbe Interact. 105:1043-1049.
- Katoh, H., Miyata, S., Inoue, H., and Iwanami, T. 2014. Unique features of a Japanese '*Candidatus Liberibacter asiaticus*' strain revealed by whole genome sequencing. PLoS One 9:e106109.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10:845-858.
- Kita, K., Kawakami, H., and Tanaka, H. 2003. Evidence for horizontal transfer of the *EcoT38I* restriction-modification gene to chromosomal DNA by the P2 phage and diversity of defective P2 prophages in *Escherichia coli* TH38 strains. J. Bacteriol. 185:2296-2305.
- Kita, K., Tsuda, J., Kato, T., Okamoto, K., Yanase, H., and Tanaka, M. 1999. Evidence of horizontal transfer of the *EcoO109I* restriction-modification gene to *Escherichia coli* chromosomal DNA. J. Bacteriol. 181:6822-6827.
- Kobayashi, I. 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res. 29:3742-3756.
- Kobayashi, I., Nobusato, A., Kobayashi-Takahashi, N., and Uchiyama, I. 1999. Shaping the genome-restriction-modification systems as mobile genetic elements. Curr. Opin. Genet. Dev. 9:649-656.
- Kunta, M., Zheng, Z., Wu, F., da Graca, J. V., Park, J. W., Deng, X., and Chen, J. 2017. Draft whole-genome sequence of '*Candidatus Liberibacter asiaticus*' strain TX2351 isolated from Asian citrus psyllids in Texas, USA. Genome Announc. 5:e00170-17.
- Li, W. B., 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.
- Lin, H., Han, C. S., Liu, B., Lou, B., Bai, X., Deng, C., Civerolo, E. L., and Gupta, G. 2013. Complete genome sequence of a Chinese strain of '*Candidatus Liberibacter asiaticus*'. Genome Announc. 1:e00184-13.
- Luria, S. E., and Human, M. L. 1952. A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. 64:557.
- Meselson, M., and Yuan, R. 1968. DNA restriction enzyme from *E. coli*. Nature 217:1110-1114.
- Murray, N. E. 2000. Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol. Mol. Biol. Rev. 64:412-434.
- Nathans, D., and Smith, H. O. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. Annu. Rev. Biochem. 44:273-293.
- Oliveira, P. H., Touchon, M., and Rocha, E. P. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. 42:10618-10631.
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. 2015. REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. Nucleic Acids Res. 43:D298-D299.
- 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.
- 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.
- Wilson, G. G. 1991. Organization of restriction-modification systems. Nucleic Acids Res. 19:2539-2566.
- Wu, F., Kumagai, L., Liang, G., Deng, X., Zheng, Z., Keremane, M., and Chen, J. 2015b. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from a citrus tree in San Gabriel, California. Genome Announc. 3:e01508-15.
- Wu, F., Zheng, Z., Deng, X., Cen, Y., Liang, G., and Chen, J. 2015a. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from *Diaphorina citri* in Guangdong, China. Genome Announc. 3:e01316-15.
- Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18:821-829.
- Zhang, S., Flores-Cruz, Z., Zhou, L., Kang, B. H., Fleites, L., Gooch, M. D., Wulff, N. A., Davis, M. J., Duan, Y., and Gabriel, D. W. 2011. '*Ca. Liberibacter asiaticus*' carries an excision plasmid prophage and a chromosomally integrated prophage that becomes lytic in plant infections. Mol. Plant-Microbe Interact. 24:458-468.
- Zheng, Z., Bao, M., Wu, F., Chen, J., and Deng, X. 2016. Predominance of single prophage carrying a CRISPR/*cas* system in '*Candidatus Liberibacter asiaticus*' strains in southern China. PLoS One 11:e0146422.
- Zheng, Z., Deng, X., and Chen, J. 2014a. Whole-genome sequence of '*Candidatus Liberibacter asiaticus*' from Guangdong, China. Genome Announc. 2:e00273-14.
- Zheng, Z., Deng, X., and Chen, J. 2014b. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from California. Genome Announc. 2:e00999-14.
- Zheng, Z., Sun, X., Deng, X., and Chen, J. 2015. Whole-genome sequence of '*Candidatus Liberibacter asiaticus*' from a huanglongbing-affected citrus tree in central Florida. Genome Announc. 3:e00169-15.
- Zheng, Z., Wu, F., Kumagai, L., Polek, M., Deng, X., and Chen, J. 2017. Two '*Candidatus Liberibacter asiaticus*' strains recently found in California harbor different prophages. Phytopathology 107:662-668.