Prophage Diversity of ‘*Candidatus Liberibacter asiaticus*’ Strains in California

Z. Dai,¹ F. Wu,¹ Z. Zheng,¹ R. Yokomi,² L. Kumagai,³ W. Cai,⁴ J. Rascoe,³ M. Polek,⁵ J. Chen,^{2,†} and X. Deng^{1,†}

¹Department of Plant Pathology, South China Agricultural University, Guangzhou, Guangdong, China; ²United States Department of Agriculture (USDA)–Agricultural Research Service, San Joaquin Valley Agricultural Sciences Center, Parlier, CA, U.S.A.; ³Plant Pest Diagnostic Center, California Department of Food and Agriculture, Sacramento, U.S.A.; ⁴USDA Animal and Plant Health Inspection Service–Plant Protection and Quarantine, Beltsville, MD, U.S.A.; and ⁵National Clonal Germplasm Repository for Citrus and Dates, Riverside, CA, U.S.A.

Accepted for publication 7 October 2018.

ABSTRACT

Huanglongbing (HLB) is a highly destructive citrus disease and is associated with a nonculturable bacterium, ‘*Candidatus Liberibacter asiaticus*’. ‘*Ca. L. asiaticus*’ in the United States was first found in Florida in 2005 and is now endemic there. In California, ‘*Ca. L. asiaticus*’ was first detected in Hacienda Heights in Los Angeles County in 2012 and has now been detected in multiple urban locations in southern California. Knowledge of ‘*Ca. L. asiaticus*’ strain diversity in California is important for HLB management. In this study, genomic diversity among 10 ‘*Ca. L. asiaticus*’ strains from six California locations were analyzed using a next-generation sequencing (NGS) (Illumina MiSeq and HiSeq) approach. Draft genome sequences of ‘*Ca. L. asiaticus*’ strains were assembled. Sequences of the 16S ribosomal RNA gene and *nrdB* confirmed ‘*Ca. L. asiaticus*’ identity. Prophages were detected in all

‘*Ca. L. asiaticus*’ strains. The California ‘*Ca. L. asiaticus*’ strains formed four prophage typing groups (PTGs): PTG1, with type 1 prophage only (strains from Anaheim, San Gabriel, and Riverside); PTG2, with type 2 prophage only (strains from Hacienda Heights); PTG1-3, with both type 1 and 3 prophages (a strain from Cerritos); and PTG1-2, with both type 1 and type 2 prophages (a strain from La Habra). Analyses of the *terL* sequence showed that all California ‘*Ca. L. asiaticus*’ strains were not introduced from Florida but likely from locations in Asia. Miniature inverted-repeat transposable elements were found in all ‘*Ca. L. asiaticus*’ strains, yet, a jumping-out event was detected in the ‘*Ca. L. asiaticus*’ strain from Cerritos. Altogether, this study demonstrated that the NGS approach focusing on prophage variation was sensitive and effective in revealing diversity of ‘*Ca. L. asiaticus*’ strains in California.

Huanglongbing (HLB) (yellow shoot disease, also known as greening disease) is a highly destructive citrus disease. HLB is associated with ‘*Candidatus Liberibacter asiaticus*’, a nonculturable proteobacterium (Jagoueix et al. 1994), and is transmitted by the Asian citrus psyllid (ACP), *Diaphorina citri* (Bove 2006). In the United States, ‘*Ca. L. asiaticus*’ was first detected in Florida in 2005 (Halbert 2005) and is now endemic there. In California, ‘*Ca. L. asiaticus*’ was first detected in 2012 in a citrus tree in an urban backyard in City of Hacienda Heights in southern California (Kumagai et al. 2013). Three years later, multiple ‘*Ca. L. asiaticus*’ strains were found in city of San Gabriel, 25 km away from Hacienda Heights (Yan et al. 2016). Since then, ‘*Ca. L. asiaticus*’ has been detected in ACP or citrus trees in multiple urban locations in southern California (Fig. 1). The extent of diversity of these California ‘*Ca. L. asiaticus*’ strains remains largely unknown.

Research on ‘*Ca. L. asiaticus*’ biology is highly challenging due to the lack of in vitro cultures. However, with the advance in DNA technology, ‘*Ca. L. asiaticus*’ can be studied through genome sequence analyses. Currently, 10 whole sequenced genomes of ‘*Ca. L. asiaticus*’ strains have been deposited in GenBank (version 223). Two of them, strains HHCA (GCA_000724755.2) and SGCA5 (GCA_001430705.1), are from California (Wu et al. 2015a; Zheng et al. 2014a). The genome of ‘*Ca. L. asiaticus*’ is composed of a

generally highly conserved chromosomal region and a prophage region. Chromosomal genes such as the 16S ribosomal RNA (rRNA) gene and deoxy-ribonucleotide reductase gene (*nrdB*) have been used for ‘*Ca. L. asiaticus*’ species identification (Jagoueix et al. 1994; Li et al. 2006; Zheng et al. 2016). A genomic locus, *trn*, with short tandem repeats, has been found to be highly variable and has been used in population diversity studies (Chen et al. 2010; Katoh et al. 2011).

In comparison, many more variations were found in the prophage region of ‘*Ca. L. asiaticus*’. Prophages are chromosomal DNA sequences which originated from the integration of bacteriophages into bacterial chromosomes. Single-nucleotide polymorphisms (SNPs) in a phage DNA terminase large subunit gene (*ter* or *terL*) were first used to separate ‘*Ca. L. asiaticus*’ strains between China and Florida (Liu et al. 2011). A double-locus analysis using the combination of a tandem repeat number (TRN) at the *trn* locus (Chen et al. 2010) and SNPs at the *terL* locus differentiated ‘*Ca. L. asiaticus*’ strains from China, Brazil, Florida, and California (Deng et al. 2014). Three prophage types (types 1, 2, and 3) along with their circular plasmid forms were described (Zhang et al. 2011; Zheng et al. 2018). In California ‘*Ca. L. asiaticus*’ strain HHCA harbored a type 2 prophage and three strains from San Gabriel harbored a type 1 prophage (Zheng et al. 2017). Both strains from Hacienda Heights and San Gabriel were of Asiatic rather than Floridian origin (Deng et al. 2014; Zheng et al. 2017).

Miniature inverted-repeat transposable elements (MITEs) are an important driving force for bacterial genotypic and phenotypic variations (Delihas 2011; Feschotte et al. 2002). MITEs were discovered within prophages of ‘*Ca. L. asiaticus*’ (Wang et al. 2013). Two types of MITEs, MCLas-A and MCLas-B, were found in ‘*Ca. L. asiaticus*’ strains from China and Florida (Wang et al. 2013). MITEs move from place to place via a cut-and-paste mechanism. In addition, the cutting-off or jumping-out of MCLas-A

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

Funding: This work was supported, in part, by the California Citrus Nursery Board, California Citrus Research Board, and Chinese Modern Agricultural Technology Systems (CARS-26).

The author(s) declare no conflict of interest.

left behind various short sequence repeats as remnants. These remnants serve as evidence in detection of MITE movement. Based on the patterns of remnant repeats, ‘*Ca. L. asiaticus*’ strains were grouped into remnant types (RT), which showed a tendency toward regional specificity (Wang et al. 2013, 2015). A remnant type of MCLas-B was not observed but two subtypes, MCLas-B1 and MCLas-B2, were detected based on sequence variations (Wang et al. 2013).

In this study, ‘*Ca. L. asiaticus*’ strains collected from six urban locations in southern California were studied for genomic diversity, largely focusing on their prophages. Next-generation sequencing (NGS) was used to generate a large volume of short sequence reads from DNA collected from ‘*Ca. L. asiaticus*’-infected citrus trees or ACP. Draft ‘*Ca. L. asiaticus*’ genome sequences were assembled or reassembled if reported previously, such as strain HHCA (now HHCA1) and SGCA5. Prophages in different ‘*Ca. L. asiaticus*’ strains were detected using a read-mapping technique (Zheng et al. 2017) and compared. Variations of ‘*Ca. L. asiaticus*’ strains in prophage types were evaluated and discussed with regard to geographical origins.

MATERIALS AND METHODS

‘*Ca. L. asiaticus*’ DNA and sequences. Fourteen samples (citrus leaves or ACP) infected with ‘*Ca. L. asiaticus*’ were collected (Table 1). ‘*Ca. L. asiaticus*’ DNA from each sample was designated as a strain. Ten ‘*Ca. L. asiaticus*’ strains were from California (Fig. 1; Table 1). One strain from Florida and three strains from China were included for comparative purposes. California ‘*Ca. L. asiaticus*’ strains were collected by the California Department of Food and Agriculture. Of the California strains, HHCA1 (previously named as HHCA), SGCA5, SGCA1, and

SGpsy were reported previously (Zheng et al. 2017). Reference prophage sequences were downloaded from the GenBank database: type 1, SC1 (NC_019549.1); type 2, SC2 (NC_019550.1); and type 3, P-JXGC-3 (KY661963.1). SC1 and SC2 were from strain UF506 in Florida (Zhang et al. 2011). P-JXGC-3 was reported from China (Zheng et al. 2018). For all ‘*Ca. L. asiaticus*’ samples, total DNA was extracted along with host (ACP or plant), as described previously (Zheng et al. 2017). Sequences of SC1, SC2, and P-JXGC-3 were further used as queries to BLAST against representative whole-genome sequences of *Citrus* spp. and ACP in GenBank to confirm their nonhost origins. ‘*Ca. L. asiaticus*’ status was confirmed by polymerase chain reaction (PCR) using specific primer set HLBas/HLBr (Li et al. 2006) targeting the 16S rRNA gene, and primer set RNR1f/r targeting *nrdB* (Zheng et al. 2016).

DNA sequencing and assembling. Sample DNA (‘*Ca. L. asiaticus*’ and host) were sequenced by the Illumina MiSeq or HiSeq platform (Illumina, Inc.) through a commercial source after bacterial DNA enrichment and enlargement (Zheng et al. 2014a). Total DNA of the La Habra strain, LHCA, was sequenced by Illumina MiSeq platform from the Animal and Plant Health Inspection Service–Plant Protection and Quarantine, Beltsville, MD laboratory without enrichment and enlargement steps. For ACP DNA, the enrichment step was not needed and was omitted (Wu et al. 2015b). Nucleotide quality of MiSeq or HiSeq read data were checked with FastQC software (Andrews 2010) and confirmed Q scores (>30) to ensure high nucleotide reliability. Draft whole-genome sequences of ‘*Ca. L. asiaticus*’ strains were acquired through mapping HiSeq or MiSeq reads to the whole-genome sequence of ‘*Ca. L. asiaticus*’ strain A4 (GCA_000590865.2) using Bowtie2 (Langmead and Salzberg 2012). Bowtie2 scripts are available upon request. Alternatively, CLC Genomic Workbench (v10.0; Qiagen, Inc.) was used based on default setting. General

Fig. 1. Map of six locations of the greater Los Angeles area of California area where samples of ‘*Candidatus Liberibacter asiaticus*’ strains were collected. Each location label starts with name of the city followed by name of the ‘*Ca. L. asiaticus*’ strain, its prophage typing group (PTG), and type of miniature inverted-repeat transposable elements (MITEs) in each line. A = MITE MCLasA, B1 = MITE MCLasB1, and B2 = MITE MCLasB2.

information of assembly and consensus extraction were obtained using CLC Genomic Workbench.

For prophage detection, MiSeq or HiSeq reads were mapped to reference prophages SC1, SC2, and P-JXGC-3 using Bowtie2. Mapping profiles were visualized by IGV software v 2.3.97 (Robinson et al. 2011) or CLC Genomic Workbench. Structurally, each reference prophage was composed of an early-gene region and a late-gene region (Zhang et al. 2011). Because all prophages share similar early-gene regions, a '*Ca. L. asiaticus*' strain with reads covering >30% of the late-gene region was considered to be harboring the reference prophage type. The threshold of 30% was empirically estimated based on mapping profile observations. Prophage sequences of each '*Ca. L. asiaticus*' strain were acquired through consensus of read mapping.

Characterization of '*Ca. L. asiaticus*' strains. Each '*Ca. L. asiaticus*' strain was characterized by four parameters, as follows. (i) Prophage typing group (PTG), was identified according to the prophage types a strain harbored. (ii) SNP signatures in *terL* sequences were identified. Briefly, the *terL* gene sequence in each prophage of a '*Ca. L. asiaticus*' strain was extracted in reference to the *terL* sequence in '*Ca. L. asiaticus*' strain A4 (CD16_RS05440) with the assistance of BLAST search (Camacho et al. 2009). All *terL* sequences were aligned by MUSCLE or ClustalW algorithms implemented in Mega 7 software (Edgar 2004; Kumar et al. 2016; Thompson et al. 1994). SNPs were identified manually in reference to SC1, SC2, and P-JXGC-3. (iii) MITE characteristics were defined. The two MITEs (MCLasA and MCLasB) described previously (Wang et al. 2013), and all possible RTs of MCLas-A, which varied from zero to three repeats in the terminal inverted repeat (TIR) regions generated in silico, were used as queries for BLAST search against MiSeq or HiSeq read data of each '*Ca. L. asiaticus*' strain. A MITE was determined by detection of TIR and direct repeat and an additional 30 bp outward. (iv) TRN characteristics at the *trn* locus were defined (Chen et al. 2010; Deng et al. 2014). Two methods were used: (a) TRN was counted from assembled '*Ca. L. asiaticus*' draft genomes; the *trn* locus was located by primer sequences of LapGP1f/r (Chen et al. 2010) or through BLAST; and (b) TRN was counted directly from MiSeq or HiSeq reads because the quality of nucleotide was high, as indicated by the >30 Q scores. Short reads covering the *trn* locus were identified by LapGP1f/r through BLAST search and extracted for manual TRN count. It should be noted that locus *trn* was in the '*Ca. L. asiaticus*' chromosomal region, rather than in the prophage (Chen et al. 2010).

Phylogeny of prophages. Contigs in each draft prophage were ordered according to reference prophages (SC1, SC2, or P-JXGC-3) and concatenated to a single nucleotide sequence required for data input in Mega 7 (Kumar et al. 2016). Prophage sequences were aligned through the CLUSTAL package (Thompson et al. 1994) embedded in Mega 7. The general time-reversible substitution model was chosen and a phylogenetic tree was constructed using the maximum-likelihood method (Nei and Kumar 2000) with 500 bootstrap tests in Mega 7.

RESULTS

Selected metrics of the genome sequence assemblies of the 14 '*Ca. L. asiaticus*' strains (10 from California) are listed in Table 1. Primer sequences for '*Ca. L. asiaticus*' detection through the 16S rRNA gene and *nrdB* gene were also detected with 100% match in each assembly. This study generated eight new draft whole-genome sequences of California '*Ca. L. asiaticus*' strains. Two of them, strains AHCA1 and SGCA1, were deposited into GenBank. The other draft genome sequences were not submitted because a significant portion (>40%) of these draft genome sequences consisted of contigs <200 bp. Current GenBank policy does not accept contig <200 bp. All draft genome sequences are available upon request. Due to '*Ca. L. asiaticus*' titer variation in the DNA samples, the sequences that covered the approximately 1.2 M '*Ca. L. asiaticus*' whole genome varied, as shown in assembly size, total contig number, and coverage (X) (Table 1). Note that small numbers for coverage (X) in Table 1 meant partial covering of the whole '*Ca. L. asiaticus*' genome. Quality of sequence nucleotide was high (Q score > 30). The draft or partial genome sequences were satisfactory for genome analyses in this study.

PTGs of '*Ca. L. asiaticus*' strains. The noncitrus or ACP origin tests found no significant sequence similarity between the three prophages (SC1, SC2, and P-JXGC-3) and the sequences of *Citrus sinensis* (National Center for Biotechnology Information [NCBI]: taxid:2711), *C. clementina* (NCBI: taxid:85681), and ACP (NCBI: taxid:121845). Specifically, no BLAST hit with a length of >42 bp, significantly smaller than the read length of 100 bp or more (Table 1). The read-mapping technique effectively detected the presence of prophage in all '*Ca. L. asiaticus*' strains (Fig. 2A, B, and C; Table 2). Coverage status in the late-gene region was a good indicator for prophage type identification. For the 10 California '*Ca. L. asiaticus*' strains, type 1 prophage was found in 8 strains, type 2 prophage was in 3 strains, and type 3 prophage was in 1 strain

TABLE 1. General information for whole-genome sequences of '*Candidatus Liberibacter asiaticus*' strains in this study

Strain	Geographic origin	Host	Sequencing format	Read length (bp)	Total reads	Sample collection date	Total ' <i>Ca. L. asiaticus</i> ' reads/bps	Assembly size (bp)	Contigs (n)	Coverage (X) ^a	GenBank accession ^b	Reference
1, AHCA1	California, United States	ACP	HiSeq 3000	150	611,118,478	08/02/2017	563,855/84,578,250	1,209,331	22	69	GCA_003143875.1	This study
2, AHCA2	Anaheim	Citrus	HiSeq 3000	100	593,33,332	10/18/2017	34,811/3,481,100	925,768	3,696	3	–	This study
3, CTCA1	Cerritos	ACP	HiSeq 3000	100	610,349,210	03/15/2017	305,845/30,584,500	1,147,287	769	25	–	This study
4, HHCA1	Hacienda Heights	Citrus	MiSeq 250	250	34,481,138	05/16/2014	9,317/2,329,250	1,150,620	239	2	GCA_000724755.2	Zheng et al. 2014a
5, HHCA2	Hacienda Heights	Citrus	MiSeq 250	250	29,149,334	02/03/2017	18,071/4,517,750	1,129,846	675	4	–	This study
6, LHCA1	La Habra	Citrus	MiSeq 300	300	49,149,732	03/29/2017	22,249/6,674,700	1,165,516	493	5	–	This study
7, RSCA1	Riverside	ACP	HiSeq 3000	100	471,308,572	09/28/2017	96,997/9,699,700	1,075,978	1,564	8	–	This study
8, SGCA5	San Gabriel	Citrus	MiSeq 250	250	35,614,768	09/04/2015	59,947/14,986,750	1,201,385	56	12	GCA_001430705.1	Wu et al. 2015a
9, SGCA1	San Gabriel	Citrus	MiSeq 250	250	38,192,184	10/16/2015	1,843/460,750	235,690	610	1	GCA_003149415.1	This study
10, SGpsy	San Gabriel	ACP	MiSeq 250	250	40,868,296	10/16/2015	8,293/2,073,250	778,008	1,394	2	–	This study
11, FL17	Polk County, Florida, United States	Citrus	MiSeq 300	300	51,684,100	09/19/2014	146,817/44,045,100	1,227,253	3	11	GCA_000820625.1	Zheng et al. 2015
12, A4	Guangdong, China	Citrus	MiSeq 250	250	32,767,554	01/14/2014	536,881/134,220,250	1,233,514	1	138	GCA_000590865.2	Zheng et al. 2014b
13, JXGC	Jiangxi	Citrus	HiSeq 3000	150	103,149,836	07/12/2017	604,058/90,608,700	1,225,162	1	76	GCA_002216815.1	Zheng et al. 2017
14, YCpsy	Guangdong	ACP	MiSeq 250	250	39,773,438	07/31/2015	5,595,634/1,398,908,500	1,233,647	9	1,120	GCA_001296945.1	Wu et al. 2015b

^a Coverage (X) = Total '*Ca. L. asiaticus*' sequence data (bp) per 1.2 M (rough estimate of '*Ca. L. asiaticus*' genome size).

^b Symbol – indicates not submitted due to the fact that a significant portion (>50%) of the draft genome consisted of contigs <200 bp. Current GenBank policy does not accept contigs <200 bp. Yet, all un-deposited whole-genome sequences are available upon request.

(CTCA1). We noted the presence of two prophage strains, CTCA1 (type 1 and 3) and LHCA1 (type 1 and 2). The 10 California ‘*Ca. L. asiaticus*’ strains formed four PTGs: PTG1 with type 1 prophage only, PTG2 with type 2 prophage only, PTG1-2 with both type 1 and type 2 prophages, and PTG1-3 with both type 1 and type 3

prophages (Table 2; Fig. 1). For the non-California strains, strain FL17 was in PTG1, strain A4 was in PTG2, and strain YCpsy with two prophages was in PTG1-3 (Table 2).

Prophage phylogeny. Phylogenetic relationships among all the prophages in this study based on whole prophage sequences

Fig. 2. Profiles of ‘*Candidatus Liberibacter asiaticus*’ MiSeq or HiSeq read mapping in reference to prophage A, SC1; B, SC2; and C, P-JXGC-3. Names of ‘*Ca. L. asiaticus*’ strains and prophages are listed on the left. Call for the presence of a prophage was based on significant read mapping in the late-gene region and highlighted by bold boxes. Reference prophage genome sequences were at the top of A, B, and C. Arrow boxes = open reading frames; *terL* = gene of phage terminase large unit; and MCLasA, MCLasB1, or MCLasB2 = miniature inverted-repeat transposable elements (MITEs). For each ‘*Ca. L. asiaticus*’ strain, gray graphs represent read coverage in log scale. Regions corresponding to *terL* in each prophage are highlighted by the shaded box. For MITE analysis, open rectangle = MCLasA, cross rectangle = jumping-out of MCLasA with remnants, open triangle = MCLasB1, line-triangle = MCLasB2, X = false read mapping confirmed by read analyses, and O = no read mapping.

are described in Figure 3. Supported by high bootstrap values, prophages clustered along the line of prophage types represented by SC1, SC2, and P-JXGC-3. In consideration of the whole prophage sequences (Fig. 2A, B, and C), the tree clusters were obviously due to the differences in the late-gene regions. Because of the incomplete and partial coverage status, subgrouping based on tree topology within a prophage type may not be reliable. However, some clustering trends deserve mentioning. Within the type 1 prophages, the three San Gabriel prophages grouped together and the two Anaheim prophages grouped together. Within the type 2 prophages, the two Hacienda Heights prophages grouped together. All these suggested that ‘*Ca. L. asiaticus*’ strains collected in the same locations harbored the same type of prophage.

The *terL* locus. Based on read mapping (Fig. 2A, B, and C), the 5’ half of *terL* showed more variation in read coverage than that of the 3’ half. Selected variable regions of *terL* at nucleotide sequence level are shown in Figure 4, with Figure 4A representing the 5’ region and Figure 4B and C representing the 3’ region. Significant SNPs among types were present but no SNP within a prophage type (Fig. 4A). SNPs within prophage type were present, mainly related to SC1 and SC2 of Florida origin (Fig. 4B and C). In fact, these are regions of primer sets FC3f/r and CT3f/r used to differentiate Asiatic and Floridian strains (Deng et al. 2014). The point to make here was that all California ‘*Ca. L. asiaticus*’ prophages were not associated with the Florida ones (SC1 and SC2), with the exception of strain FL17 as explained later (in Discussion).

MITEs. Read mapping also found MITEs in all prophages (Fig. 2A, B, and C). Based on map consensus and confirmed by read sequences, MCLasA, MCLasB1, and MCLasB2 (Fig. 5) were identified and labeled, and jumping-out of CLasA was shown (Fig. 2B). Different combinations of prophage type and MITE type were observed and summarized in Table 2. If a MITE was detected but no prophage was called (e.g., strain RSCA1) (Fig. 2B), this MITE was assigned to the called prophage with no MITE (e.g., P-RSCA1-1)

(Fig. 2A). It was assumed that a MITE should have a prophage or host (Feschotte et al. 2002).

The *trn* locus. TRNs at the *trn* locus were summarized in Table 3. A draft genome sequence had only a single TRN. In contrast, TRN counts from MiSeq or HiSeq reads detected multiple TRNs in three of the 16 ‘*Ca. L. asiaticus*’ strains (MiSeq or HiSeq datasets). However, in all cases, a dominant TRN type could be identified. The dominant TRN type might or might not agree with the TRN type counted from the draft genome sequence. An example is strain SGCA5 with TRN = 6 in the draft genome count but TRN ≥ 19 in read counts. The discrepancy was likely due to the fact that the draft SGCA5 genome sequence was assembled using strain YCpsy (TRN = 6) as reference (Wu et al. 2015a) (i.e., an error from the reference assembling method).

DISCUSSION

Although California is still officially considered to be ‘*Ca. L. asiaticus*’ free due to the vigorous statewide survey efforts and mandatory removal of infected trees, additional ‘*Ca. L. asiaticus*’ detections are expected due to the long latent period of HLB. When a new ‘*Ca. L. asiaticus*’ strain is identified, the source and its relationship to previously known strains are among the first questions asked. Answers to these questions are challenging because information or a database of ‘*Ca. L. asiaticus*’ strains in California is lacking. This study analyzed a total of 10 ‘*Ca. L. asiaticus*’ strains recently found in California at the whole-genome sequence level and established a PTG system for ‘*Ca. L. asiaticus*’ strain evaluation. Note that the *trn* locus is not in the prophage but could provide additional information to assist the evaluation. Our results suggested that the California ‘*Ca. L. asiaticus*’ strains detected thus far were likely the result of several introductions, not from Florida but from other uncertain Asiatic origins, substantiating the previous observations on the strains from Hacienda Heights and San Gabriel (Deng et al. 2014; Zheng et al. 2017).

Fig. 2. (Continued from previous page)

Using the PTG system, we were able to evaluate relationships of ‘*Ca. L. asiaticus*’ strains between plant and insect hosts and within and among locations. For the Anaheim samples, one ‘*Ca. L. asiaticus*’ strain was from ACP (AHCA1) and the other was from

citrus (AHCA2). Both were in the same PTG1 (Fig. 1). Similarly, the three San Gabriel ‘*Ca. L. asiaticus*’ strains (two from citrus and one from ACP) (Table 1) were in the same PTG1 (Fig. 1). These suggested that a PTG was not altered by passage through the vector

TABLE 2. General information for prophages about ‘*Candidatus Liberibacter asiaticus*’ strains from California and Florida in the United States and Guangdong, China and grouping of ‘*Ca. L. asiaticus*’ strains based on prophage types

Strain	Prophage	Seq (bp) ^b	N ^c	Reference coverage (%) ^a			Type	Cov (X) ^d	MITE ^e	'Ca. L. asiaticus'	
				SC1	SC2	P-JXGC-3				Strain	PTG ^f
California											
1	P-AHCA1-1	36,165	11	90.3	(46.4)	(51.1)	1	45	B1	AHCA1	1
2	P-AHCA2-1	22,589	94	56.4	(27.0)	(31.9)	1	1	B1	AHCA2	1
3	<i>P-CTCA1-1</i>	35,107	43	87.7	(46.4)	78.0	1	40	<i>A(RT0-2)</i>	<i>CTCA1</i>	1–3
4	<i>P-CTCA1-3</i>	24,537	38	...	...	...	3	20	<i>B2</i>	...	...
5	P-HHCA1-2	35,419	15	(42.3)	90.8	(52.6)	2	5	A	HHCA1	2
6	P-HHCA2-2	31,971	38	(38.6)	82.0	(44.1)	2	3	A	HHCA2	2
7	<i>P-LHCA1-1</i>	37,655	7	94.0	96.6	(58.8)	1	23	<i>B2</i>	<i>LHCA1</i>	1–2
8	<i>P-LHCA1-2</i>	37,672	7	...	...	...	2	19	<i>B2</i>	...	...
9	P-RSCA1-1	35,194	73	87.9	(46.5)	(55.3)	1	14	<i>A, B2</i>	<i>RSCA1</i>	1
10	P-SGCA1-1	26,896	27	67.2	(41.6)	(43.9)	1	2	A	SGCA1	1
11	P-SGCA5-1	34,442	46	86.0	(53.1)	(56.2)	1	8	A	SGCA5	1
12	P-SGpsy-1	20,083	28	50.1	(28.0)	(31.5)	1	1	A	SGpsy	1
Non-California											
1	P-FL17-1	38,657	19	96.5	(55.2)	(61.1)	1	270	<i>A, B1</i>	<i>FL17</i>	1
2	P-A4-2	37,384	35	(48.7)	95.9	(56.5)	2	112	A (RT1-3)	A4	2
3	<i>P-YCpsy-1</i>	38,324	19	95.7	(51.4)	100.0	1	1,522	<i>B2</i>	<i>YCpsy</i>	1–3
4	<i>P-YCpsy-3</i>	31,449	1	...	...	...	3	1,776	<i>B2</i>	...	...
Reference											
1	SC1	40,048	1	na	na	na	1	na	B1	UF506	1
2	SC2	38,997	1	na	na	na	2	na	A	UF506	2
3	P-JXGC-3	31,449	1	na	na	na	3	na	B2	JXGC	3

^a SC1 (40,048 bp), SC2 (38,997 bp), and P-JXGC-3 (31,449 bp). Numbers used for prophage type assignments are in bold. Numbers in parentheses were considered to be low and not used for prophage type assignments; na = not applied.

^b Assembled sequence length.

^c Number of contigs.

^d Average nucleotide coverage.

^e Characteristic of miniature inverted-repeat transposable element (MITE) remnant type (RT) is in parentheses. ‘*Ca. L. asiaticus*’ strains with two MITEs detected are in italics.

^f PTG = prophage typing group.

Fig. 3. Phylogenetic trees of whole prophages of ‘*Candidatus Liberibacter asiaticus*’ strains, generated using the maximum-likelihood method. Numbers at the nodes are bootstrap values.

and provided evidence of ‘*Ca. L. asiaticus*’ spread by ACP in these locations (in contrast to the possibility that ‘*Ca. L. asiaticus*’ was still confined to the tree of original infection), suggesting a necessity for ACP control.

In Hacienda Heights, the same type 2 prophage was detected in the two ‘*Ca. L. asiaticus*’ strains. Strain HHCA2 was from a backyard tree located three residential properties from where the HHCA1 was collected. The tree was large, with symptomatic leaves high in the canopy that could only be seen by the surveyor from a tall ladder. It should be noted that, between 2012 and 2018, more than 2,300 citrus tree samples and more than 4,200 ACP samples have been collected from Hacienda Heights and tested repeatedly. The 4-year lag in HHCA2 detection suggested that movement of ‘*Ca. L. asiaticus*’ in Hacienda Heights area was slow. Analyses of such factors could be responsible for the slow movement of ‘*Ca. L. asiaticus*’ would provide new information for HLB management.

On the other hand, the same PTG1, as well as the TRN < 10 group, of ‘*Ca. L. asiaticus*’ strains between Anaheim and Riverside (Fig. 1; Table 3) raises a question about possible connections between ‘*Ca. L. asiaticus*’ strains originating from two distinct areas. The two cities are 60 km apart and they are connected by major transport corridors (Fig. 1). Analyses of the ‘*Ca. L. asiaticus*’ strains, if found in locations between the two cities, would be helpful in tracking the path of ‘*Ca. L. asiaticus*’ movement. In contrast, from Anaheim northwest 50 km to San Gabriel, four different PTG strains were found (Fig. 1), suggesting different ‘*Ca. L. asiaticus*’ introductions. It is also interesting to note that, although ‘*Ca. L. asiaticus*’ strains from both Anaheim and San Gabriel were in PTG1, their TRN status was different (TRN < 10 in Anaheim versus TRN > 10 in San Gabriel), suggesting that the ‘*Ca. L. asiaticus*’ strains between the two cities were not the same, and adding further evidence for the complexity of ‘*Ca. L. asiaticus*’ introduction in the Anaheim-San Gabriel area.

To date, ‘*Ca. L. asiaticus*’ has not been detected in the Central Valley of California, where the majority of the California commercial citrus is located. A major mountain range separates southern California from the Central Valley but major transport highways exit (Fig. 1). If ‘*Ca. L. asiaticus*’ is detected in the Central Valley, it will be of utmost importance to be able to quickly analyze and compare the valley strains to those identified in southern California. The ‘*Ca. L. asiaticus*’ PTG system developed in this study could serve as a basis for ‘*Ca. L. asiaticus*’ strain analysis to

assist HLB management efforts. There is currently no other published system available for differentiation of ‘*Ca. L. asiaticus*’ strains within California.

In addition to the PTG system, this NGS-based study also extended new knowledge about the loci of *terL*, *trn*, and MITE in ‘*Ca. L. asiaticus*’. As mentioned above (Fig. 4B and C), the prophage of strain FL17 from central Florida grouped with all California ‘*Ca. L. asiaticus*’ prophages, but not SC1 and SC2, currently used to represent Florida ‘*Ca. L. asiaticus*’ prophages. By MiSeq reads counting, strain FL17 showed a TRN = 12, in contrast to the TRN = 5 in the draft genome sequence (Table 3). This is again likely a sequence assembling error because strain Psy62 (TRN = 5) was used as reference (Zheng et al. 2015). The corrected TRN = 12 matched with a previous report of a TRN > 10 ‘*Ca. L. asiaticus*’ strain in central Florida (Chen et al. 2010). Therefore, strain FL17 was in the Asiatic group. In the double-locus analysis with *terL* and *trn* on 83 ‘*Ca. L. asiaticus*’ strains from Florida (Deng et al. 2014), 5 such minority ‘*Ca. L. asiaticus*’ strains represented by strain FL17 were found, compared with the 78 majority ‘*Ca. L. asiaticus*’ strains

TABLE 3. Tandem repeat number (TRN) analyses at the *trn* locus (CD16_RS05440 in strain A4) of ‘*Candidatus Liberibacter asiaticus*’ strains

Number	Strain	TRN in draft genome	TRN in MiSeq or HiSeq reads ^a
1	AHCA1	7	7 × 27
2	AHCA2	Not found	8 × 1
3	CTCA1	6	≥11 × 9
4	HHCA1	8 ^b	Not found
5	HHCA2	4	4 × 2
6	LHCA1	6	6 × 2
7	RSCA1	6	8 × 2
8	SGCA1	Not found	Not found
9	SGCA5	6	≥19 × 5
12	SGpsy	Not found	Not found
13	FL17	5	9 × 2, 11 × 2, 12 × 19
14	A4	6	6 × 105
15	JXGC	6	12 × 1, 11 × 2, 10 × 28, 8 × 1, 5 × 1
16	YCpsy	6	6 × 1, 11 × 2, 13 × 1, 16 × 64, 17 × 71, 18 × 2

^a Data are presented in the format of TRN × number of reads.

^b Identified by polymerase chain reaction amplification and Sanger sequencing (Zheng et al. 2014a).

Fig. 4. Multiple sequence alignments of selected variable regions of *terL* in prophages in ‘*Candidatus Liberibacter asiaticus*’ strains. **A**, Represents the 5’ half region and **B** and **C**, represents the 3’ half region. **B**, Region of primer FC3f and primer CT3f and **C**, region of primer FC3r and primer CT3r (Deng et al. 2014). Numbers at both end of a sequence indicate the nucleotide positions in SC1, SC2, and P-JXGC-3, respectively; an asterisk (*) highlights the presence of a single-nucleotide polymorphism (SNP) highlighted by shadow boxes; # indicates a nucleotide not found in the sequence; and – indicates a nucleotide deletion.

Fig. 5. Multiple sequence alignments of miniature inverted-repeat transposable elements and flanking regions (30 bp). **A**, MCLasA sequences and **B**, MCLasB sequences. Note the subdivision into MCLasB1 and MCLasB2, mainly by single-nucleotide polymorphisms, indicated by an asterisk (*) at the bottom. DR = direct repeat region, TIR = terminal inverted repeat, CR = central region, - indicates a nucleotide deletion compared with SC2 or SC1, # indicates nucleotide not found in the sequence, // indicates omission of nucleotide for illustration simplicity, and - indicates a nucleotide deletion compared with the reference sequences (SC2 in MCLasA alignment and P-JXGC-3 in MCLasB2 alignment). Note the pattern of zero to two repeats in TIR regions (i.e., RT0-2 remnant type of P-CTCA1-1 and the RT3-1 remnant type of P-A4-2) (A).

represented by strain Psy62. It remains unclear how strain FL17 was related to the ‘*Ca. L. asiaticus*’ strains in California, particularly strain SGCA5, which is also a TRN > 10 Asiatic group strain.

The genome of ‘*Ca. L. asiaticus*’ was originally considered to have no transposon (Duan et al. 2009). Close examination later revealed the presence of MITEs (Wang et al. 2013). The biological role of ‘*Ca. L. asiaticus*’ MITEs remains unknown at the present. However, detection of a remnant type (RT0-2) in P-CTCA1-1 (Fig. 5) demonstrated evidence that MITE MCLasA in California was active, like those in Florida and China. In addition, we noted that MCLasB2 was associated with all three double-prophage ‘*Ca. L. asiaticus*’ strains (CTCA1, LHCA, and YCpsy) (Table 2). ‘*Ca. L. asiaticus*’ RT profiles seem to have geographical specificity. A preliminary study showed that, among the 30 MCLasA jumping-out strains, 17 were RT0-3 followed by 11 RT3-1 in Guangdong; and, among the 35 MCLas jumping-out strains, 12 were RT0-3 and 11 were RT3-0 (Wang et al. 2013). Only a single RT0-2 ‘*Ca. L. asiaticus*’ strain was found in California and more analyses are needed to reveal the California RT profile.

In summary, this study employed an NGS approach to characterize the whole genome of 10 ‘*Ca. L. asiaticus*’ strains from six locations in southern California. Prophages were found in all ‘*Ca. L. asiaticus*’ strains and used for strain characterization. At least four PTGs were identified, suggesting that California ‘*Ca. L. asiaticus*’ strains were likely introduced from different sources other than Florida. The NGS datasets were also ready for analyses on specific genomic loci such as *terL*, *trn*, and MITEs, which supplemented PTG analysis and led to new biological discoveries of ‘*Ca. L. asiaticus*’. We proposed a preliminary workflow for ‘*Ca. L. asiaticus*’ diversity research and to generate timely information to assist current HLB management efforts in California.

ACKNOWLEDGMENTS

We thank S. Vargas for technical assistance. 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).

LITERATURE CITED

- Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Online publication. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Bove, 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 Bioinf.* 10:421.
- Chen, J., Deng, X., Sun, X., Jones, D., Irey, M., and Civerolo, E. 2010. Guangdong and Florida populations of ‘*Candidatus Liberibacter asiaticus*’ distinguished by a genomic locus with short tandem repeats. *Phytopathology* 100:567-572.
- Delihias, N. 2011. Impact of small repeat sequences on bacterial genome evolution. *Genome Biol. Evol.* 3:959-973.
- Deng, X., Lopes, S., Wang, X., Sun, X., Jones, D., Irey, M., Civerolo, E., and Chen, J. 2014. Characterization of ‘*Candidatus Liberibacter asiaticus*’ populations by double-locus analyses. *Curr. Microbiol.* 69:554-560.
- 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.
- Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792-1797.
- Feschotte, C., Zhang, X., and Wessler, S. 2002. Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. Pages 1147-1158. *Mobile DNA II*. N. L. Craig, ed. ASM Press, Washington, DC.
- Halbert, S. E. 2005. The discovery of huanglongbing in Florida. In: *Proc. 2nd Int. Citrus Canker Huanglongbing Res. Workshop H-3*, Orlando, FL.
- Jagoueix, S., Bove, 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. *Int. J. Syst. Bacteriol.* 44:379-386.
- Katoh, H., Subandiyah, S., Tomimura, K., Okuda, M., Su, H. J., and Iwanami, T. 2011. Differentiation of ‘*Candidatus Liberibacter asiaticus*’ isolates by variable-number tandem-repeat analysis. *Appl. Environ. Microbiol.* 77:1910-1917.

- 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.
- Langmead, B., and Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- 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.
- Liu, R., Zhang, P., Pu, X. L., Xing, X. Q., Chen, J. C., and Deng, X. L. 2011. Analysis of a prophage gene frequency revealed population variation of '*Candidatus Liberibacter asiaticus*' from two citrus-growing provinces in China. *Plant Dis.* 95:431-435.
- Nei, M., and Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29:24-26.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Wang, X., Chen, J., Tan, J., Duan, S., Deng, X., Chen, J., and Zhou, C. 2015. High genetic variation and recombination events in the vicinity of non-autonomous transposable elements from '*Candidatus Liberibacter asiaticus*'. *J. Integr. Agric.* 14:2002-2010.
- Wang, X., Tan, J., Bai, Z., Deng, X., Li, Z., Zhou, C., and Chen, J. 2013. Detection and characterization of miniature inverted-repeat transposable elements in '*Candidatus Liberibacter asiaticus*'. *J. Bacteriol.* 195:3979-3986.
- 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 Announce.* 3:e01508-15.
- 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 Announce.* 3:e01316-15.
- Yan, Z., Rascoe, J., Kumagai, L. B., Keremane, M. L., and Nakhla, M. K. 2016. Citrus huanglongbing (HLB) discoveries in California in 2015 and 2012 are of different genotypes of *Candidatus Liberibacter asiaticus* (cLas) by double-locus Genomic variation analysis. *Plant Dis.* 100:645.
- Zhang, S. J., Flores-Cruz, Z., Zhou, L. J., Kang, B. H., Fleites, L. A., Gooch, M. D., Wulff, N. A., Davis, M. J., Duan, Y. P., 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., Van Horn, C., Chen, J., and Deng, X. 2018. A type 3 prophage of '*Candidatus Liberibacter asiaticus*' carrying a restriction-modification system. *Phytopathology* 108:454-461.
- Zheng, Z., Deng, X., and Chen, J. 2014a. Draft genome sequence of '*Candidatus Liberibacter asiaticus*' from California. *Genome Announce.* 2:e00999-14.
- Zheng, Z., Deng, X., and Chen, J. 2014b. Whole-genome sequence of '*Candidatus Liberibacter asiaticus*' from Guangdong, China. *Genome Announce.* 2:e00273-14.
- 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 Announce.* 3:e00169-15.
- 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* 107:662-668.
- 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: Article 39020.