Genomes of '*Candidatus* Liberibacter solanacearum' Haplotype A from New Zealand and the United States Suggest Significant Genome Plasticity in the Species

Sarah M. Thompson, Chris P. Johnson, Ashley Y. Lu, Rebekah A. Frampton, Kerry L. Sullivan, Mark W. E. J. Fiers, Ross N. Crowhurst, Andrew R. Pitman, Ian A. W. Scott, Aimin Wen, Neil C. Gudmestad, and Grant R. Smith

First, third, fourth, fifth, sixth, eighth, ninth, and twelfth authors: The New Zealand Institute for Plant & Food Research Limited, Lincoln 7608, New Zealand; first, third, fourth, fifth, eighth, ninth, and twelfth authors: Plant Biosecurity Cooperative Research Centre, Canberra, ACT 2617, Australia; second, tenth, and eleventh authors: Department of Plant Pathology, North Dakota State University, Fargo 58108; and seventh author: The New Zealand Institute for Plant & Food Research Limited, Mt Albert 1025, New Zealand.

Current address of A. Y. Lu and M. W. E. J. Fiers: Center for the Biology of Disease, 3000 Leuven, Belgium.

Accepted for publication 25 March 2015.

ABSTRACT

Thompson, S. M., Johnson, C. P., Lu, A. Y., Frampton, R. A., Sullivan, K. L., Fiers, M. W. E. J., Crowhurst, R. N., Pitman, A. R., Scott, I. A. W., Wen, A., Gudmestad, N. C., and Smith, G. R. 2015. Genomes of '*Candidatus* Liberibacter solanacearum' haplotype A from New Zealand and the United States suggest significant genome plasticity in the species. Phytopathology 105:863-871.

'*Candidatus* Liberibacter solanacearum' contains two solanaceous crop-infecting haplotypes, A and B. Two haplotype A draft genomes were assembled and compared with ZC1 (haplotype B), revealing inversion and relocation genomic rearrangements, numerous single-nucleotide polymorphisms, and differences in phage-related regions. Differences in prophage location and sequence were seen both within and between haplotype comparisons. OrthoMCL and BLAST analyses identified 46 putative coding sequences present in haplotype A that were not present in haplotype B. Thirty-eight of these loci were not found in sequences from other Liberibacter spp. Quantitative polymerase chain reaction (qPCR) assays designed to amplify sequences from 15 of these loci were screened against a panel of '*Ca.* L. solanacearum'-positive samples to investigate genetic diversity. Seven of the assays demonstrated within-haplotype diversity; five failed to amplify loci in at least one haplotype A sample while three assays produced amplicons from some haplotype B samples. Eight of the loci assays showed consistent A–B differentiation. Differences in genome arrangements, prophage, and qPCR results suggesting locus diversity within the haplotypes provide more evidence for genetic complexity in this emerging bacterial species.

Zebra-chip (ZC) disease is a significant disease of potato that has emerged over the last 20 years. First observed in 1994 in Mexico, it is now present in a number of countries in Central and North America as well as in New Zealand (Gudmestad and Secor 2007; Liefting et al. 2008). In 2008, the putative causal agent, '*Candidatus* Liberibacter solanacearum', was discovered in New Zealand in potato, capsicum, tomato, cape gooseberry, and tamarillo (Liefting et al. 2008, 2009). '*Ca.* L. psyllaurous' was discovered independently around the same time from tomato in the United States (Hansen et al. 2008) and is generally regarded as a synonym of '*Ca.* L. solanacearum'. The bacterium is vectored by phloem-feeding psyllids (Hemiptera: Triozidae), including *Bactericera cockerelli*, the tomato/potato psyllid (TPP) (Munyaneza et al. 2007), and can also be spread in infected plant propagules (Secor et al. 2009).

Five haplotypes (A to E) of '*Ca.* L. solanacearum' have been described based on single-nucleotide polymorphisms (SNPs) and insertion-deletions (indels) in the ribosomal RNA (rRNA) region (Nelson et al. 2011, 2013; Teresani et al. 2014). Haplotypes A and B are associated with solanaceous crops and the TPP in Central and North America (A and B) and New Zealand (A). Haplotype C is associated with carrot and *B. apicalis* in Scandinavia, while haplotype D, also found in carrot, is associated with *B. trigonica* in Europe. The recently described haplotype E is associated with carrot and celery in Europe and has no confirmed psyllid vector (Teresani et al. 2014). '*Ca.* L. solanacearum', ZC, and the genus Liberibacter were recently reviewed by Haapalainen (2014).

ZC disease acquired its name from the stripes of brown discoloration seen in infected potato tubers that become more distinct after frying for chip production. In the aboveground potato plant, symptoms include curling leaves, chlorosis, aerial tubers, swollen nodes, and pink or purple discoloration.

In New Zealand, the first report of ZC and '*Ca.* L. solanacearum' (Liefting et al. 2008) occurred not long after the discovery of TPP in 2006 (Gill 2006). The symptoms and pathology of ZC disease in New Zealand are different and often less severe than the symptoms in the United States (Berry et al. 2011; Pitman et al. 2011, 2012). Variability in symptoms has also been noted in the United States where both haplotypes A and B are present. Field observations in Texas suggested that '*Ca.* L. solanacearum' haplotype B is more destructive than haplotype A (Wen et al. 2013). Many factors, individually or in combination, may contribute to the differences in symptomology and pathology between ZC disease in New Zealand and the United States, including potato cultivar, environmental conditions, vector behavior, and genetic differences.

Corresponding author: G. R. Smith; E-mail address: grant.smith@plantandfood.co.nz

Genome deposition: '*Ca.* L. solanacearum' NZ1, GenBank accession numbers JMTK00000000 and JMTK00000000.1 and 'Ca. L. solanacearum' HenneA, GenBank accession numbers JQIG00000000 and JQIG00000000.1.

S. M. Thompson and C. P. Johnson contributed equally to this research.

N. C. Gudmestad and G. R. Smith are joint senior authors.

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

http://dx.doi.org/10.1094/PHYTO-12-14-0363-FI

At present, the genetic variation within '*Ca.* L. solanacearum' is poorly understood. Currently the haplotypes can be identified by Sanger sequencing of the rRNA operon and the solanaceaous-infecting haplotypes A and B can be distinguished by a polymerase chain reaction (PCR) assay based on a simple sequence repeat (SSR) marker (Wen et al. 2013). Previous studies into solanaceous '*Ca.* L. solanacearum' variability have used multilocus sequence typing (MLST) (Glynn et al. 2012) and SSR markers (Lin et al. 2012). Although two dominant "clades" or "lineages" were identified in both studies (corresponding to the haplotypes A and B), there was evidence for additional genetic variability, with some SSR markers having up to eight different alleles (Lin et al. 2012).

Genetic variability has been investigated more extensively in '*Ca.* L. asiaticus'. '*Ca.* L. asiaticus' is one of three '*Candidatus* Liberibacter' spp. ('*Ca.* L. asiaticus', '*Ca.* L. americanus', and '*Ca.* L. africanus') associated with huanglongbing or citrus-greening disease. Serotypic analysis of '*Ca.* L. asiaticus' found seven different serotypes (Garnier et al. 1991). Analysis of SSR markers in '*Ca.* L. asiaticus' samples (Katoh et al. 2011) and PCR restriction fragment length polymorphism (RFLP) analysis of the OMP gene (Bastianel et al. 2005; Hu et al. 2011) also indicated substantial diversity within '*Ca.* L. asiaticus'. A number of '*Ca.* L. asiaticus' and '*Ca.* L. americanus' genomes have now been fully or partially assembled (Duan et al. 2009; Katoh et al. 2014; Lin et al. 2013a, b; Wulff et al. 2014.; Zheng et al. 2014). *Liberibacter crescens*, an endophyte of papaya and the only culturable species of this genus, has also been sequenced (Fagen et al. 2014; Leonard et al. 2012).

Traditional approaches cannot be used to investigate pathology or the potential contribution of pathogen variability to symptom expression for the nonculturable species; therefore, whole-genome comparisons can provide valuable insight into differences between the species. Genome sequences have been used to develop tools for investigating genetic variability and improved detection of '*Ca.* L. asiaticus' (Katoh et al. 2011, Kogenaru et al. 2014). Such tools are important for characterizing the strains used in biological experiments.

The '*Ca.* L. solanacearum' ZC1 genome is from a haplotype B-positive sample sourced from northern Texas and is the only completely assembled '*Ca.* L. solanacearum' genome to date. It is not known whether the SNPs and indels in the rRNA region that define the haplotypes are representative of major differences throughout the genome. Failure to understand underlying patterns of genetic diversity compromises both the interpretation of research results and the subsequent development of appropriate management strategies. Here, we describe and compare two '*Ca.* L. solanacearum' haplotype A genomes obtained by high-throughput sequencing of TPP samples from New Zealand and the United States. Comparison of these genomes with ZC1 resulted in the identification of loci that distinguish the '*Ca.* L. solanacearum' haplotype A genomes from haplotype B. Fifteen of these loci were used to investigate the genetic variability among samples of '*Ca.* L. solanacearum' from around North and Central America using quantitative (q)PCR assays.

MATERIALS AND METHODS

'*Ca.* L. solanacearum' sample selection and sequencing. *'Ca. L. solanacearum' NZ1*. TPP, originally collected from a colony infesting tamarillo (*Solanum betaceum*) in Northland, New Zealand, were established in a colony at Plant & Food Research, Lincoln in June 2010. Individual TPP were collected from the colony in September 2010 and DNA was extracted using cetyltrimethylammonium bromide, as described previously (Beard et al. 2013). The '*Ca.* L. solanacearum' titer of individuals was determined using the '*Ca.* L. solanacearum' 16S and psyllid internal transcribed spacer 2 SYBR green qPCR assays (Beard et al. 2013). A sample (NZ1) with a high '*Ca.* L. solanacearum' titer and DNA quality was whole genome amplified using a RepliG kit (Qiagen) and further purified using components of the Geneaid Genomic DNA Mini Kit for Plant Material. A 300-bp insert library was prepared and 100-bp paired-end reads were generated on one lane of an Illumina HiSeq2000 by Macrogen Inc. (Seoul, Korea). A second DNA sample was sequenced on one PacBio RS SMRT cell (Pacific Biosciences) at Washington State University.

'Ca. L. solanacearum' HenneA. TPP collected from potato fields in Edinburg, TX in March 2012 were established in a colony at North Dakota State University, Fargo. In June 2012, insects were collected and DNA from individuals was tested using qPCR to determine '*Ca.* L. solanacearum' 16S titer relative to the insect mitochondrial COI gene in each sample. The two DNA samples with the highest '*Ca.* L. solanacearum' titers were combined, giving one sample (HenneA). DNA library preparation was undertaken using an Ion Plus Fragment Library Kit (LifeTechnologies) and size selected using the Pippen Prep instrument (Sage Science, Inc.); then, the sample was bead loaded using an Ion OneTouch 2 instrument (Life Technologies). The bead-loaded library was then applied to an Ion 318 chip and sequenced on the Ion Torrent PGM as per the manufacturer's instructions.

Genome assembly. *'Ca. L. solanacearum' NZ1*. Sequences were filtered to a minimum threshold quality score of 20 and low-quality nucleotides were trimmed from the termini of sequences. Sequences less than 50 bp in length were discarded. The filtered reads were assembled using SOAPdenovo (Li et al. 2010). The ZC1 genome (GenBank accession number NC_014774.1) was used as an initial reference. A combination of iterative mapping with Bowtie (Langmead et al. 2009) of trimmed reads and SOAPdenovo contigs, using an in-house bash script in which consensus sequences were generated (Li et al. 2009a), was used for contig extension and gap closure. The PacBio sequences were used for scaffolding and ordering of the Illumina de novo contigs. Final mapping of the paired-end reads to the draft genome was undertaken to confirm contiguous mapping with the expected coverage across each contig.

'Ca. L. solanacearum' HenneA. Sequence files in the fastq format were generated from the Ion Torrent sequencing data. Genome de novo assembly was performed using SeqMan NGen (DNA Star) assembly software. A second assembly was performed using the Newbler assembly software (454 Life Science) supported by the iPlant Collaborative (http://www.iplantcollaborative.org/). The two assemblies were combined and the resulting contigs were mapped to '*Ca.* L. solanacearum' ZC1 (NC_014774.1) and the '*Ca.* L. solanacearum' NZ1 draft genome. The final assembly was confirmed by mapping read data back to the draft genome using SeqMan NGen.

Confirmation of genome assembly by PCR and Sanger sequencing. Primers to confirm genome rearrangements and the location of prophage were selected using Primer3 (Untergasser et al. 2012) in Geneious using default settings. Products were amplified using KAPA HiFi HotStart ReadyMix (KAPABiosystems) using the cycling conditions listed in Supplementary Tables S3 and S4 from New Zealand '*Ca.* L. solanacearum' haplotype A samples. The products were gel purified using the Gel/PCR DNA Fragments Extraction Kit (Geneaid), as per the manufacturer's instructions. Purified products were Sanger sequenced using the primers listed in Supplementary Tables S3 and S4 at the Lincoln University Sequencing Facility (Lincoln, New Zealand).

Genome annotation, ortholog determination, and comparison. To undertake whole-genome analysis, the two assembled genomes were concatenated in numerical order of the contigs and circularized into a pseudomolecule. The three '*Ca.* L. solanacearum' genome sequences ('*Ca.* L. solanacearum' ZC1, NZ1, and HenneA) were then aligned in Geneious v6.1.7 (Biomatters; http://www.geneious.com) using the progressive Mauve v2.3.1 whole-genome aligner (Darling et al. 2010). In order to illustrate the level of sequence similarity or difference between the three '*Ca.* L. solanacearum' genomes, a 100-kbp fragment from '*Ca.* L. solanacearum' ZC1,

between positions 300,000 and 400,000, was aligned with the corresponding sequence from positions 902,948 to 1,002,948 of NZ1 using progressive Mauve aligner in Geneious. The same sequence region from NZ1 was also compared with the region 827,624 to 927,625 from HenneA.

The coding regions of the three genomes ('*Ca.* L. solanacearum' ZC1, NZ1, and HenneA) were annotated using RAST (Overbeek et al. 2014) to provide consistent annotation. Prophage regions were defined based on similarity to the prophage sequences in '*Ca.* L. solanacearum' ZC1 (Lin et al. 2011). Bacteriophage (phage)-related regions were identified through the presence of phage-related genes.

In the first analysis, the coding regions of '*Ca.* L. solanacearum' haplotype A and haplotype B were compared using OrthoMCL (Li et al. 2003) to reveal the ortholog relationships. The E-value cut off for the OrthoMCL reciprocal BLAST was $1E^{-5}$, with minimum coverage of 70% over both query and subject sequence. Coding regions from '*Ca.* L. solanacearum' NZ1 for the ortholog clusters present in both haplotype A sequences but not identified in '*Ca.* L. solanacearum' ZC1 were further screened using the BLASTn function against ZC1 as a custom database in Geneious. Loci without significant similarity to ZC1 (E-value greater than $1E^{-11}$) were considered unique to '*Ca.* L. solanacearum' haplotype A. The same process was applied to coding regions identified only in ZC1.

A second OrthoMCL analysis was carried out to identify loci that distinguished '*Ca.* L. solanacearum' haplotype A from other Liberibacter spp. The Liberibacter spp. genomes listed in Table 1 were reannotated using RAST (Overbeek et al. 2014). OrthoMCL analysis was then preformed as above. The loci from '*Ca.* L. solanacearum' NZ1 identified by OrthoMCL as specific to haplotype A were screened using the BLASTn function against the National Center for Biotechnology Information (NCBI) nr database in Geneious. Loci with similarity to Liberibacter spp. sequences with an E-value of less than $1E^{-11}$ were discarded. The remaining sequences were considered to differentiate the haplotype A genomes from previously published genomes and were further analyzed using qPCR.

Primer design and assay conditions. Because a limited number of genomes were available for the whole-genome comparison, qPCR assays were developed to investigate the presence or absence of loci in a broader range of samples. Primers were designed within the putative distinguishing loci using Primer3 (Untergasser et al. 2012) in Geneious using default settings, with product size specified as 70 to 250 bp. qPCR reactions for initial screening contained 1× SensiFAST SYBR Hi-ROX mastermix (Bioline), 100 nM concentration of each primer, and 1 µl of DNA in a total reaction volume of 11 µl. Assays were run on a StepOne Plus qPCR machine (Applied Biosystems) using a touchdown protocol (Larsen et al. 2002) consisting of 15 cycles as follows: 5 s at 95°C, then 30 s at temperatures decreasing from 70 to 62°C, with a decrease of 2°C every 3 cycles. An additional 25 cycles with an annealing temperature of 60°C were then performed. Following amplification, products were subjected to melt curve analysis (60 to 95°C at 0.3°C s^{-1}). All primers were tested against three '*Ca.* L. solanacearum' haplotype A samples (from New Zealand TPP, New Zealand potato, and U.S. TPP), a haplotype B sample (from U.S. TPP), healthy potato, and healthy TPP (as determined by the '*Ca.* L. solanacearum' 16S SYBR assay) (Beard et al. 2013).

Investigating locus diversity of '*Ca.* L. solanacearum' samples. Fifteen primer sets that amplified only haplotype A samples were selected for screening against 43 additional samples (Table 2) from geographically diverse sources. These samples included haplotype A and haplotype B samples as well as samples with haplotype A and B dual infections. Haplotype status was determined using the SSR assay (Wen et al. 2013). qPCR reactions for further testing contained 1× SSoAdvanced Universal SYBR Green Supermix (Bio-Rad), 250 nM forward and reverse primers, and 2 µl of DNA in a total reaction volume of 20 µl. Assays were performed on a Bio-Rad CFX96 qPCR machine using a protocol consisting of 40 cycles of 10 s at 98°C and 30 s of 60°C. Following amplification, products were subjected to melt curve analysis (60 to 95°C at 0.3°C s^{-1}). The qPCR '*Ca.* L. solanacearum' species-specific primers LsoF and HLBr (Li et al. 2009b) were used to give a baseline cycle threshold representative of '*Ca.* L. solanacearum' titer.

RESULTS

Genome assembly. The New Zealand haplotype A sample, '*Ca.* L. solanacearum' NZ1, was assembled into five contigs, ranging in size from 18.4 to 870 kbp. These were assembled from 86,759,926 Illumina reads. The 45,111 reads from PacBio sequencing, with an average read length of 3,183 bp, were used to confirm the de novo contig joining, structural rearrangements in comparison with ZC1, and locations of prophage. Mapping of the Illumina reads to the final draft genome confirmed continuous mapping of each contig, with approximately 1,000× coverage. This draft genome of '*Ca.* L. solanacearum' NZ1 has a length of 1,312,416 bp and a GC content of 35.3%, and contains 1,236 predicted coding sequences (CDS), as annotated by RAST (Table 1). This whole-genome shotgun assembly has been deposited at GenBank under accession number JMTK00000000: the version described in this article is version number JMTK01000000. '*Ca.* L. solanacearum' HenneA, the U.S. haplotype A sample from Texas, was assembled into

TABLE 1. Genome sequences used in this study

Isolate ID	Source[a]	Country	Year	GenBank accession	Contigs	Size (Mbp)	GC (%)	RAST-predicted CDS[b]	Reference
'*Candidatus* Liberibacter americanus'									
PW_SP	Periwinkle	Brazil	2014	AOFG00000000.1	22	1.18	31.1	982	Lin et al. 2013a
Sao Paulo	Periwinkle	Brazil	2014	NC_022793.1	1	1.20	31.1	1,003	Wulff et al. 2014.
Ca. L. asiaticus									
A4	Periwinkle	China	2014	JFGQ00000000.1	10	1.21	36.4	1,173	Zheng et al. 2014
Gxpsy	*Diaphorina citri*	China	2013	NC_020549.1	1	1.27	36.6	1,225	Lin et al. 2013b
psy62	*D. citri*	United States	2009	NC_012985.3	1	1.23	36.5	1,170	Duan et al. 2009
Liberibacter crescens BT-1	Culture	Puerto Rico	2012	NC_019907.1	1	1.50	35.4	1,419	Leonard et al. 2012
'*Ca.* L. solanacearum'									
ZC1	*Bactericera cockerelli* (TPP)	United States	2011	NC_014774.1	1	1.26	35.2	1,207	Lin et al. 2011
NZ1	*B. cockerelli* (TPP)	New Zealand	2014	JMTK00000000.2	5	1.31	35.3	1,236	This study
HenneA	*B. cockerelli* (TPP)	United States	2014	JQIG00000000.1	7	1.21	34.9	1,225	This study

[a] TPP = tomato/potato psyllid.
[b] CDS = coding sequences.

seven contigs, ranging in size from 3.8 to 690 kbp from 4,152,520 Ion Torrent reads. Final mapping of the reads against the draft genome confirmed continuous mapping with approximately 45× coverage. The draft genome of '*Ca.* L. solanacearum' HenneA is 1,211,361 bp in length, contains 1,225 CDS predicted by RAST, and has a GC content of 34.9%. This draft genome does not include assembly of repetitive elements—primarily prophage regions. This whole-genome shotgun assembly has been deposited at GenBank under accession numbers JMTK00000000, JMTK00000000.1, JQIG00000000, and JQIG00000000.1: the version described in this article is version JQIG00000000.1.

Genome comparisons. After comparison of the haplotype A genomes and ZC1 using a Mauve alignment, a high degree of similarity (block synteny and direction) between NZ1 and HenneA was visualized (Fig. 1). The most notable difference between the haplotype A genomes is the location of their prophage domains. In '*Ca.* L. solanacearum' NZ1, there are three prophage domains (Fig. 2, green bars in track 1). The presence of three prophage domains in '*Ca.* L. solanacearum' NZ1 is supported by PacBio reads mapping from the core genome into the phage domains and by the level of coverage seen in these regions, consistent with that of the rest of the genome. Sanger sequencing of the region from the integrase into the adjacent core genome confirmed the presence of phage sequence at all three locations. The locations of prophage 2 and 3 correspond to contig ends in the incomplete '*Ca.* L. solanacearum' HenneA assembly, suggesting that they are located in the same positions as NZ1. Prophage 1, however, is absent from '*Ca.* L. solanacearum' HenneA. The Mauve alignment revealed two major structural rearrangements between the haplotype A genomes and the haplotype B genome. There is an approximately 57-kbp fragment (magenta block) starting at position 698,883 in '*Ca.* L. solanacearum' ZC1 that is inverted in both NZ1 and HenneA. Second, the teal-colored gene block at position 981,006 to 1,095,345 in '*Ca.* L. solanacearum' ZC1 is relocated in both NZ1 and HenneA. These major rearrangements were confirmed by Sanger sequencing of PCR products at the start and end of each rearranged block. '*Ca.* L. solanacearum' NZ1 and HenneA both differ from ZC1 in the location of the prophage domains. '*Ca.* L. solanacearum' ZC1 has two prophage regions, one at position 176,396 and the other at position 1,214,970 (Lin et al. 2011) (Fig. 1, white gaps in the ZC1 sequence red and blue blocks, respectively). There are no prophage sequences in the corresponding locations in either NZ1 or HenneA. The NZ1 prophage 1 has a high degree of similarity to both prophages in ZC1 (87.7% to prophage 1 and 84.3% to prophage 2) whereas '*Ca.* L. solanacearum' NZ1 prophages 2 and 3 are less similar to the ZC1 prophage sequences. The presence of this third prophage in NZ1 results in an increased genome size of NZ1 (1.31 Mbp) compared with the ZC1 genome (1.26 Mbp). Preliminary SNP analysis showed greater differences between haplotypes than within haplotype A (Fig. 3).

Ortholog comparison. OrthoMCL analysis of '*Ca.* L. solanacearum' ZC1, NZ1, and HenneA identified 180 orthologs potentially unique to haplotype A (Supplementary Table S1). This set was subsequently screened using BLASTn against the '*Ca.* L. solanacearum' ZC1 genome to remove any loci that had significant similarity to noncoding regions. BLASTn confirmed 46 loci (highlighted in green) as unique to haplotype A, that are mostly concentrated in one-half of the genome (Fig. 2, track 3) and located within 12 phage-related regions (Fig. 2, track 1). Of these 46 loci, 37 were annotated by RAST as hypothetical proteins. The other nine had annotations, including type III restriction-modification system methylation subunit, type III restriction-modification system DNA endonuclease, adenine or cytosine DNA methyltransferase, ISSpo3 transposase, alcohol dehydrogenase, modification methylase *Eco*RI, and terminase. OrthoMCL followed by BLAST analysis identified 24 loci in '*Ca.* L. solanacearum' ZC1 with no significant sequence homology to either NZ1 or HenneA (data not shown). All 24 of these had RAST annotations of hypothetical protein. OrthoMCL analysis also identified 839 orthologous groups common to both haplotypes.

To identify genes present in '*Ca.* L. solanacearum' haplotype A but not in other Liberibacter genomes, a second OrthoMCL analysis was undertaken using NZ1, HenneA, and genomes of other Liberibacter spp. listed in Table 1. In total, 623 ortholog clusters were common to all Liberibacter sequences used in this analysis, while 145 ortholog clusters were only found in haplotype A. These 145

TABLE 2. Fifteen primer sets selected for further analysis

Locus tag (NZ1)	Annotation	Primer name	Primer sequence	Product size (bp)
DJ66_0348	Hypothetical protein	D348F	AAAGGCGAAAAAGGTCAGTTC	119
		D348R	AAGAATCCCCCACTACCACA	
DJ66_0385	Modification methylase *Eco*RI	D385F	TGAAGCCATTAGAAAGTGATGGTG	105
		D385R	CACGGAAAAGAGAAAATGGAGGA	
DJ66_0388	Hypothetical protein	D388F	GTTTTCTCCGCCCCAATTGC	106
		D388R	TGAGCAGCCCTGATAACGAC	
DJ66_0405	Hypothetical protein	D405F	ACCGTTGTGCCATTGATCCT	224
		D405R	GCATGTCCGCACATCCTTTT	
DJ66_0406	Adenine/cytosine DNA methlytransferase	D406F	CAACATCGCGGGGCTATAGT	200
		D406R	CCAGCAGGACAAGGACCATT	
DJ66_0411	Hypothetical protein	D411F	AGGGTTTTTCTGTGGCTTTAGC	83
		D411R	ATCGGTTTGGGTTCTGAGCC	
DJ66_0426	Hypothetical protein	D426F	TGGTACTTGCATCAATCGGGAA	144
		D426R	TCTCAGGAACACCTACGGGA	
DJ66_0427	Hypothetical protein	D427F	TGGCGTTATGATCATGTCGGA	118
		D427R	AGGTGGAGTAGGGCGTTATG	
DJ66_0429	Alcohol dehydrogenase	D429F	TTCCGCCAATGAGATCCTCG	173
		D429R	ACGCGTTTCAGATGTCGGTA	
DJ66_0840	Type III restriction-modification system DNA endonuclease	D840F	ATCCCTATCAACCGCACCTT	75
		D840R	GAACATGCTGAAAACGCTGA	
DJ66_0228	Phage terminase large subunit	D228F	CCGCCACGTCTAAACCCATA	136
		D228R	CGCCATGATGTGTACGGACA	
DJ66_0231	Hypothetical protein	D231F	TCTTTGCCCCTTCTTTCCAACT	137
		D231R	GAGAATGAAGACCCTGGTTTTCC	
DJ66_0253	Hypothetical protein	D253F	AGCAACAGCGATATCTAAAGGTT	112
		D253R	GTGCATGCAGGTGGAAAACA	
DJ66_0266	Hypothetical protein	D266F	CAAACACCTCCCCTCTCCAG	171
		D266R	TTCCCCCTTAGGTGTTGGGA	
DJ66_0267	Hypothetical protein	D267F	ACGTCTCTAAAGCTCATACTGGA	107
		D267R	AGCATGATCCTGTTTCCTTTGA	

loci from '*Ca.* L. solanacearum' NZ1 were screened using BLASTn against the NCBI nr database. Thirty loci appeared to be unique to '*Ca.* L. solanacearum' haplotype A because they had no significant similarity to any entries in the nr database (Supplementary Table S2; highlighted in gray). Most candidate loci with BLAST E-values lower than $1E^{-11}$ were against other Liberibacter spp. sequences; however, eight loci gave significant BLAST scores to non-Liberibacter bacteria species (Supplementary Table S2; highlighted in pink). In total, 38 loci differentiated the haplotype A sequences from those of other Liberibacter spp.

Investigating locus diversity of '*Ca.* L. solanacearum' samples. Primers were designed to amplify DNA sequences from the 38 loci that distinguished haplotype A from other Liberibacter spp. by qPCR. To assess primer performance, these primers were initially screened against three '*Ca.* L. solanacearum' haplotype A-positive samples, one '*Ca.* L. solanacearum' haplotype B-positive sample, and '*Ca.* L. solanacearum'-negative samples of potato and TPP in a qPCR touchdown assay (data not shown). From these results, 15 primer sets were selected and subsequently assessed against a wider panel of 43 DNA samples from '*Ca.* L. solanacearum' haplotype A- or B-positive insect or plant material from geographically diverse sources. This screen was conducted in a second laboratory, on a different qPCR platform, and with different reagents and protocol to independently validate the primer sets. In all, 8 of the 15 primer sets specifically amplified a PCR product from both haplotype A and mixed-haplotype samples, suggesting that the

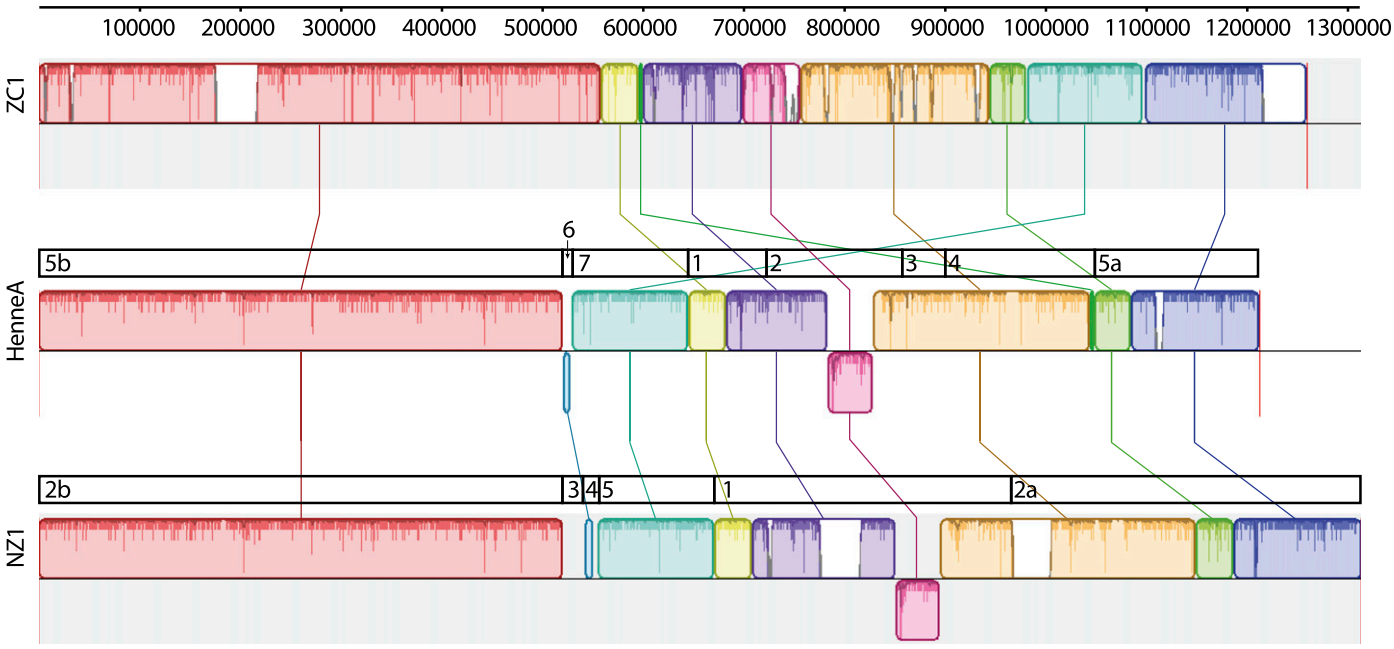

Fig. 1. Mauve alignment of the two '*Candidatus* Liberibacter solanacearum' haplotype A genomes (NZ1 and HenneA) and haplotype B genome ZC1. Two contigs (5 in HenneA, and 2 in NZ1) have been broken during alignment of the pseudomolecules to ZC1. The white blocks at 200 and 1,200 kbp in ZC1 and 550, 800, and 950 kbp in NZ1 are prophage domains; the prophage domains in HenneA are not assembled. The block at 981 to 1,095 kbp in ZC1 is relocated in both haplotype A genomes and the block at 699 to 756 kbp in ZC1 is inverted (presented below the other blocks) in both haplotype A genomes.

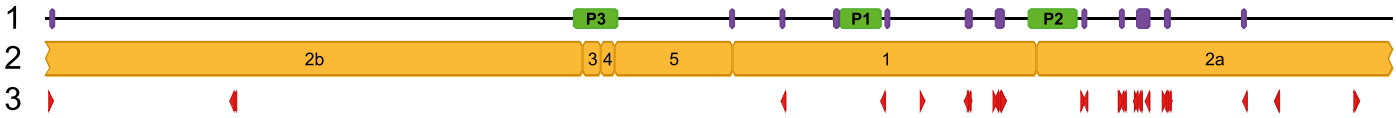

Fig. 2. Genomic location of regions specific to '*Candidatus* Liberibacter solanacearum' haplotype A. Track 1: Location of prophage (green) and phage-related regions (blue) in the '*Ca.* L. solanacearum' NZ1 genome. Track 2: Concatenated contigs of NZ1, ordered in relation to '*Ca.* L. solanacearum' ZC1. Track 3: Location of regions on the NZ1 genome that are specific to haplotype A as defined by OrthoMCL comparative genome analysis against other Liberibacter genomes, and subsequent BLAST screen.

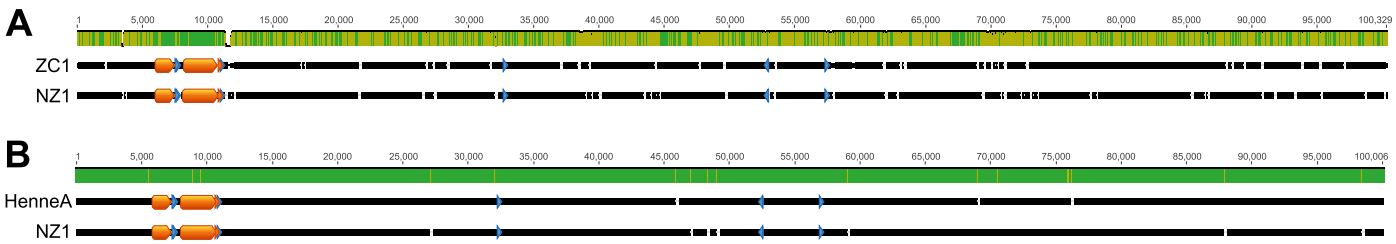

Fig. 3. Alignment of corresponding 100-kbp regions of '*Candidatus* Liberibacter solanacearum' haplotypes. **A,** Alignment of ZC1 (position 300,000 to 400,000) and corresponding region (902,948 to 1,002,948) of NZ1 reveals significant single-nucleotide polymorphisms (SNPs) between the two genomes, while **B,** the comparison of NZ1 (902,948 to 1,002,948) to HenneA (827,624 to 927,624) shows few SNPs. Red: <30% similarity, brown: 30% to <100%, green: 100%. Orange blocks indicate one rRNA operon (of the three in the genomes) and blue triangles represent tRNAs.

targeted loci were specific to haplotype A. Five of the primer sets failed to amplify putative haplotype A loci (D348, D405, D406, D840, and D231) from individual haplotype A or mixed-haplotype samples, indicating that these loci were not present in all haplotype A populations. Primer sets for three loci (D348, D388, and D427) amplified PCR products from a limited number of haplotype B DNA samples (Table 3). PCR primers to amplify locus D1204 generated nonspecific amplicons from almost all haplotype B samples.

DISCUSSION

The draft genome of '*Ca.* L. solanacearum' haplotype A was sequenced and assembled from two sources of TPP, one from the United States and one from New Zealand. Subhaplotype (intra A-A) and interhaplotype (inter A-B) comparisons using the two haplotype A genomes and the complete genome of '*Ca.* L. solanacearum' ZC1 (haplotype B) enabled the identification of haplotype A-specific loci. The specificity of a number of the loci to haplotype A was subsequently confirmed by qPCR using primer sets designed to amplify part of each locus. These provided sequences and genomic locations for the differentiation of haplotype A not previously described in '*Ca.* L. solanacearum'; previous studies focused on rRNA (Nelson et al. 2011, 2013; Teresani et al. 2014), SSR (Lin et al. 2012; Wen et al. 2013), or MLST (Glynn et al. 2012).

The two '*Ca.* L. solanacearum' haplotype A genomes, HenneA and NZ1, are 34.9 to 35.3% GC rich, with 1,225 and 1,236 predicted coding sequences, respectively. These values are consistent with other Liberibacter spp. genomes, including '*Ca.* L. solanacearum' ZC1, with 35.2% GC and 1,246 predicted coding sequences (Lin et al. 2011). Of interest, the three rRNA operons of the two '*Ca.* L. solanacearum' haplotype A draft genomes were identical. Lin et al. (2011) reported SNPs between the three operons in ZC1. No SNPs were apparent when all reads from NZ1 were mapped to a single operon. '*Ca.* L. americanus' Sao Paulo also has three identical operons (Wulff et al. 2014.) although, in most Liberibacter genomes, the operon sequences vary. The evolutionary significance of this difference is not known.

In total, 839 core ortholog clusters were identified for '*Ca.* L. solanacearum' haplotypes A and B by OrthoMCL analysis, containing 840 HenneA CDS and 854 NZ1 CDS. The nucleotide sequences encoding these core orthologs are highly conserved between NZ1 and HenneA, with relatively few SNPs differentiating the two strains. Such conservation between the two haplotype A sequences from such geographically distinct sources suggests that haplotype

TABLE 3. Cycle threshold values for screen of 15 loci quantitative polymerase chain reaction assays against 43 samples[a]

Sample	Type[b]	Host	Collection location	Year	LsoF-HLBr[c]	D348	D385	D388	D405	D406
Olsen Kt5	A	Potato	Idaho	2012	25.4	–	**27.9**	**27.3**	**28.9**	**29.1**
Olsen Kt2	A	Potato	Idaho	2012	27.8	–	**30.1**	**30.2**	**31.1**	**31.2**
ZC675	A	Potato	Nebraska	2014	22.0	**23.7**	**22.9**	**22.7**	**23.5**	**23.2**
ZC675-5	A	Potato	Nebraska	2014	24.7	**26.7**	**25.6**	**25.7**	**27.2**	**26.3**
ID258-2	A	Potato	Texas	2008	25.8	–	**27.7**	**27.4**	**29.1**	**28.9**
450T15	A	Potato	Texas	2011	23.4	**25.5**	**24.5**	**24.3**	**26.5**	**25.5**
ZC369	A	Potato	Texas	2010	27.8	**30.4**	**29.4**	**28.9**	**30.9**	**29.9**
WA471-3	A	Potato	Washington	2011	21.8	**23.8**	**23.3**	**23.2**	**25.7**	**24.8**
Tunnel2	A	Psyllids	Guatemala	2013	24.2	**27.7**	**26.1**	**26.1**	**28.9**	**28.5**
HenB2	A	Psyllids	Texas	2013	15.1	**17.9**	**16.6**	**16.7**	**18.4**	**17.9**
HenneB	A	Psyllids	Texas	2013	20.2	**27.3**	**21.3**	**22.7**	–	–
EggRootSe1	AB	Eggplant	Guatemala	2013	17.2	**21.2**	**19.4**	**19.2**	**22.6**	**21.3**
EggRootSe3	AB	Eggplant	Guatemala	2013	17.1	**20.4**	**19.4**	**19.1**	**21.9**	**20.6**
Olsen817	AB	Potato	Idaho	2013	24.2	**26.8**	**26.0**	**26.0**	**27.4**	**27.0**
Olsen859	AB	Potato	Idaho	2013	24.7	**27.4**	**26.7**	**26.2**	**28.0**	**27.6**
ZC457T1	AB	Potato	Washington	2011	26.6	**34.0**	**32.1**	**32.3**	**34.1**	–
TunnelVV	AB	Psyllids	Guatemala	2013	18.2	**21.4**	**20.0**	**19.9**	**22.1**	**21.7**
PO	AB	Psyllids	Texas	2013	16.6	**19.1**	**18.4**	**18.1**	**18.5**	**18.4**
HenF	AB	Psyllids	Texas	2013	13.9	**18.8**	**26.1**	**25.7**	**27.7**	**27.6**
HenL	AB	Psyllids	Texas	2014	17.8	**19.6**	**18.7**	**19.3**	**18.8**	**19.1**
ZC398-2	B	Potato	Colorado	2010	25.6	–	–	–	–	–
FM7wSto	B	Potato	Colorado	2012	21.2	–	*	–	*	–
ZC673-14	B	Potato	Colorado	2014	23.8	–	–	–	–	–
ZC673-12	B	Potato	Colorado	2014	23.8	–	–	–	–	–
ZC673-7	B	Potato	Colorado	2014	23.3	–	–	–	–	–
TX418Sto2	B	Potato	Texas	2011	19.4	–	–	**29.4**	–	–
ZC674-2	B	Potato	Texas	2014	25.2	–	–	–	–	–
ZC672-5	B	Potato	Texas	2014	26.8	–	–	–	–	–
ZC431T8	B	Potato	Texas	2011	22.5	–	–	–	–	–
ZC429T3b	B	Potato	Texas	2011	19.5	–	–	–	–	–
ZC367-adj1	B	Potato	Texas	2010	22.0	–	–	*	–	–
ZC416T4b	B	Potato	Texas	2011	24.2	–	–	–	–	–
ZC414T1	B	Potato	Texas	2011	24.8	–	–	–	–	–
ZC421 T7	B	Potato	Texas	2011	25.7	–	–	–	–	–
ZC456T1	B	Potato	Washington	2011	24.6	–	–	–	–	–
Herm	B	Psyllids	Oregon	2012	13.7	**19.9**	–	*	–	*
HenneA1	B	Psyllids	Texas	2013	26.5	–	–	–	–	–
Edbg3	B	Psyllids	Texas	2011	17.3	–	*	–	–	–
551Bulk	B	Psyllids	Texas	2012	16.4	–	–	–	–	–
572C	B	Psyllids	Texas	2012	16.1	–	–	–	–	–
ZC572b	B	Psyllids	Texas	2012	17.7	–	*	–	–	–
HenW	B	Psyllids	Texas	2012	25.3	–	–	–	–	–
HenT	B	Psyllids	Texas	2012	19.1	**24.8**	–	–	–	–

(continued on next page)

[a] NT = not tested, * = alternative amplicon (different melt curve), – = no amplicon, and bold = positive quantitative polymerase chain reaction (qPCR) result.
[b] Haplotype.
[c] '*Ca.* L. solanacearum'-specific LsoF-HLBr qPCR primers (Li et al. 2009b).

A is highly homogeneous within the core genome. However genome analyses of other bacteria have shown that the sequencing of one or two genomes is not sufficient to understand subspecific diversity and that sequencing of multiple strains is required to present a more consistent definition of the species (or, in this case, the haplotype) (Mann et al. 2013; Medini et al. 2005; Tettelin et al. 2008). A second '*Ca.* L. solanacearum' genome (R1) has recently become available in GenBank (GCA_000756225.1). Because of the number of contigs (99), it is difficult to undertake a robust analysis to define the haplotype or to compare it with the NZ1 and HenneA haplotype A genome sequences presented here. The sequencing of additional '*Ca.* L. solanacearum' haplotype A genomes is required to define the extent of homogeneity of the core genome.

Although the core genome of the two haplotype A sequences is highly homogeneous, there are significant differences in the accessory genome. The key difference between the two haplotype A sequences is the location of the prophage sequences. Although the complete assembly and placement of the prophages in HenneA is yet to be elucidated, it appears that prophage 1 of '*Ca.* L. solanacearum' NZ1 is not present in HenneA at the corresponding genomic location. In addition, few HenneA reads map to the NZ1 prophage 1 (data not shown). Thus, NZ1 contains at least one prophage that differentiates it from HenneA. Phage typing has been used to differentiate strains of bacterial species (Baggesen et al. 2010). Profiling the prophage in '*Ca.* L. solanacearum' isolates may be useful in identifying the sources of disease outbreaks or incursions such as that of '*Ca.* L. solanacearum' in New Zealand in 2008.

The genome structure and organization of ZC1 was distinct from that of HenneA and NZ1. SNP analysis also differentiated the haplotype B genome from the two haplotype A genomes. The existence of two haplotypes of '*Ca.* L. solanacearum' associated with ZC was revealed in previous studies (Glynn et al. 2012; Nelson et al. 2011). Our findings confirm that differences between the haplotypes are present throughout the genome and not restricted to the rRNA operon or a limited number of housekeeping genes.

The composition and genomic locations of the prophages are also a major difference between ZC1 and the haplotype A samples. Only two prophages are present in ZC1. NZ1 prophage 1 has 84.3 to 87.7% identity to the ZC1 prophages (that is, below the threshold for the Mauve comparison in Figure 1 to indicate an alignment), suggesting that these may be related but distinct. NZ1 prophages 2 and 3 are similar to each other but divergent from both the NZ1 prophage 1 and the ZC1 prophages. They may represent a different

TABLE 3. (*continued from preceding page*)

D411	D426	D427	D429	D840	D228	D231	D253	D266	D267
27.8	**27.9**	**27.8**	**28.4**	**26.9**	**27.5**	**28.1**	**27.4**	**28.4**	**27.7**
29.7	**30.1**	**30.1**	**31.3**	**29.0**	**30.2**	**30.2**	**29.5**	**31.2**	**30.2**
23.4	**22.8**	**22.9**	NT	**28.2**	**22.8**	**23.3**	**23.0**	**22.9**	**23.1**
26.3	**26.0**	**25.8**	**25.6**	**25.2**	**25.6**	**26.4**	**25.9**	**26.1**	**26.2**
28.4	**28.5**	**28.1**	**28.2**	**27.0**	**28.1**	**28.7**	**27.4**	**28.8**	**28.1**
25.1	**24.6**	**24.6**	**24.6**	–	**24.6**	**25.3**	**24.6**	**25.1**	**24.6**
29.5	**29.5**	**30.2**	**29.3**	**28.7**	**29.7**	**30.6**	**29.7**	**31.1**	**29.5**
23.5	**23.8**	**23.4**	**24.1**	**22.5**	**24.0**	**24.1**	**23.4**	**24.5**	**23.6**
26.3	**27.4**	**26.1**	**28.0**	**25.1**	**27.4**	**27.0**	**26.8**	**28.3**	**26.5**
17.2	**17.3**	**17.0**	**17.4**	**16.2**	**17.3**	**17.1**	**16.8**	**18.1**	**17.2**
20.4	**29.3**	**21.1**	**24.8**	**21.2**	**22.3**	**22.0**	**24.3**	**30.3**	**22.3**
19.6	**20.1**	**19.3**	**19.7**	**19.0**	**20.0**	**20.5**	**19.7**	**20.6**	**19.9**
19.6	**19.7**	**19.2**	**19.6**	**18.8**	**19.7**	**20.0**	**19.5**	**20.4**	**19.6**
26.5	**27.7**	**26.3**	**27.1**	**25.5**	**26.5**	**27.0**	**26.2**	**27.0**	**26.4**
26.6	**26.7**	**26.5**	**26.8**	**25.8**	**26.8**	**27.1**	**26.8**	**27.2**	**26.8**
33.7	**32.2**	**31.1**	**33.4**	**32.2**	**32.7**	–	**33.1**	**31.4**	**32.3**
20.1	**20.7**	**20.2**	**21.2**	**19.1**	**21.0**	**20.6**	**20.5**	**21.5**	**20.4**
18.8	**18.1**	**18.1**	**17.8**	**18.1**	**18.3**	**18.5**	**18.3**	**18.3**	**18.5**
26.0	**26.4**	**20.4**	**26.8**	**25.2**	**26.8**	**26.6**	**26.1**	**26.9**	**26.2**
19.3	NT	**18.6**	NT	**18.4**	**19.0**	**19.0**	NT	**18.5**	**19.0**
–	–	–	–	–	–	–	–	*	–
–	–	**25.9**	–	–	–	–	–	*	–
–	–	**29.3**	–	–	–	–	–	–	–
–	–	**29.3**	–	–	–	–	–	–	–
*	–	**28.9**	–	*	–	–	–	–	–
–	–	**25.4**	*	–	–	–	–	*	–
–	–	**31.8**	–	–	–	–	–	–	–
–	–	**31.0**	–	–	–	–	–	–	–
*	–	**27.4**	–	–	–	–	–	*	–
*	–	**25.8**	–	–	–	–	–	–	–
–	–	**32.6**	*	–	–	–	–	–	–
–	–	**29.4**	–	–	–	–	–	*	–
–	–	**31.0**	–	–	–	*	–	*	–
–	–	**32.0**	–	–	–	–	–	*	–
–	–	–	–	–	–	–	–	*	*
*	*	**20.0**	–	–	*	–	*	*	*
–	–	–	–	–	–	–	–	*	–
–	–	**23.6**	–	–	–	–	*	–	–
–	–	**23.0**	–	–	–	–	–	*	–
–	–	**22.3**	–	–	–	–	–	–	–
–	–	**23.6**	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–
–	–	**24.6**	–	–	–	–	–	*	–

lineage of prophage. The divergence of prophage sequence in haplotype A in comparison with haplotype B is consistent with the two haplotypes being genetically distinct. This variability in prophage sequence is not unexpected. In '*Ca.* L. asiaticus', at least seven prophage variants have been described (Gao et al. 2011; Zhou et al. 2013). In addition, in the assembled genomes, the number of prophage present varies: Psy62 has one, Gxpsy has two tandem prophages, while the Japanese strain Ishi has none (Duan et al. 2009; Katoh et al. 2014; Lin et al. 2013b). However the variability in location of the prophage sets '*Ca.* L. solanacearum' apart from '*Ca.* L. asiaticus': in the two fully assembled '*Ca.* L. asiaticus' genomes with prophage (Psy62 and Gxpsy), the prophages are present in the same location within the genome. The number and diversity of the prophages in '*Ca.* L. solanacearum' may be important in the variation of virulence observed between haplotype A and haplotype B (or even between strains), because the genotype of '*Ca.* L. asiaticus' prophages appears to correlate with disease development in plant hosts (Zhou et al. 2013). If prophages have a similar role in pathogenicity in '*Ca.* L. solanacearum', then the differences in prophages within and between haplotypes may result in differences in virulence. We speculate that such differences may contribute to the variation in severity of ZC observed in New Zealand and the United States. A better understanding is required of the diversity of prophage genomes in '*Ca.* L. solanacearum' and how they contribute to the biology of the pathogen.

Forty-six haplotype A orthologs were identified as unique to haplotype A using OrthoMCL and BLAST analysis, because no equivalent ortholog was identified in the genome of ZC1. Many (37 of 46) of the orthologs identified as specific to haplotype A were annotated by RAST as hypothetical proteins and were less than 70 amino acid residues in size. In total, 24 orthologs from '*Ca.* L. solanacearum' ZC1 were also predicted to be specific to this genome and annotated by RAST as hypothetical proteins (again, most were less than 70 amino acid residues in length). The lack of clarity surrounding the predicted function of many of these haplotype- or species-specific orthologs is consistent with Kogenaru et al. (2014), who identified 18 coding sequences in '*Ca.* L. asiaticus' that were not present in '*Ca.* L. americanus' and '*Ca.* L. africanus', the other two citrus-infecting Liberibacter spp. All 18 sequences were annotated as hypothetical proteins. The extent of hypothetical proteins that are unique between the '*Ca.* L. solanacearum' A and B genomes makes interpreting biological differences between the haplotypes from genomic sequence impossible. One potential contribution to differences in pathology between the A and B haplotypes, however, is the presence of a shorter coding domain predicted to encode a serralysin (or hemolysin-related protein) in the haplotype B genome compared with that in the haplotype A genomes. The coding domain is shorter because two units of a N-terminal tandem repeat is absent. Hemolysin-like proteins are required for virulence of both animal pathogens (Nishibuchi and Kaper 1995) and plant pathogens (Agrios 2004).

Further, a number of the putative haplotype A unique orthologs cluster in what appear to be phage-related regions of the bacterial genome, which perhaps represent remnants of chromosomal insertion of phages or the presence of other mobile elements with phage homologs. This would support the suggestion by Lin et al. (2011) that phage integration is involved in genome variation in '*Ca.* L. solanacearum'. For the development of haplotype-specific diagnostics, it is important to understand whether these loci are mobile or fixed.

In this study, 15 of the loci were selected to further investigate genetic variability among haplotypes. Screening of a panel of insect and plant material known to be positive for '*Ca.* L. solanacearum' using qPCR revealed that eight loci differentiated '*Ca.* L. solanacearum' haplotype A from haplotype B. These appear to be fixed in the '*Ca.* L. solanacearum' haplotype A genome and may prove to be useful as diagnostic targets. The other seven showed variability. These loci were present in PacBio reads that confirmed the assembly of the Illumina data and their presence in the '*Ca.* L. solanacearum' NZ1 genome. This variability of loci, particularly in regions in proximity to genes such as transposase and integrase, suggests the presence of mobile genetic elements. These results reiterated that intra-haplotype variation is probably significant, consistent with the findings of previous studies (Glynn et al. 2012; Lin et al. 2012).

In summary, this study provides genome-wide evidence for genetic diversity within '*Ca.* L. solanacearum' populations. Together, the genomic rearrangements, the presence of three prophage sequences in '*Ca.* L. solanacearum' NZ1 (including one that is highly similar to that in '*Ca.* L. solanacearum' ZC1), and the variation in the presence or absence of other loci in '*Ca.* L. solanacearum' samples confirm genome plasticity in '*Ca.* L. solanacearum'. Our analysis indicates that major variation in the accessory genome occurs within haplotypes as well as between haplotypes. We propose that differences in the disease symptoms described in the United States and New Zealand, and even within Texas, may be the result of this genetic variation and, hence, disease symptom variability may occur not just between haplotypes but also within haplotypes. This highlights the need to characterize genotypes used in research, and the tools we have developed based on our genome comparisons allow us to begin addressing this.

ACKNOWLEDGMENTS

We thank our colleagues at Plant & Food Research, North Dakota State University, the Plant Biosecurity CRC, and the Liberibacter research community for their assistance and samples to undertake this research; D. Gibson for assistance with the figures; N. Berry and N. Jorgensen for psyllid colony establishment and maintenance; and Plant & Food Research via CAP10-72, the Australian Government's Cooperative Research Centres Program via CRC2002, and the USDA via USDA-NIFA-SCRI 2009-51181-20176 funds for their support.

LITERATURE CITED

Agrios, G. N. 2004. Plant Pathology, 5th ed. Academic Press, San Diego.

Baggesen, D. L., Sørensen, G., Nielsen, E. M., and Wegener, H. C. 2010. Phage typing of *Salmonella* Typhimurium—Is it still a useful tool for surveillance and outbreak investigation? Euro Surveill. 15:19471.

Bastianel, C., Garnier-Semancik, M., Renaudin, J., Bové, J. M., and Eveillard, S. 2005. Diversity of "*Candidatus* Liberibacter asiaticus," based on the *omp* gene sequence. Appl. Environ. Microbiol. 71:6473-6478.

Beard, S. S., Pitman, A. R., Kraberger, S., and Scott, I. A. W. 2013. SYBR Green real-time quantitative PCR for the specific detection and quantification of '*Candidatus* Liberibacter solanacearum' in field samples from New Zealand. Eur. J. Plant Pathol. 136:203-215.

Berry, N., Thompson, S., Taylor, N., Wright, P., Shah, F., Walker, M., Beard, S., Jorgensen, N., Butler, R., Scott, I., and Pitman, A. R. 2011. The impact of *Ca.* Liberibacter infected seed tubers on potato production in New Zealand. Pages 178-182 in: Proc. 11th Annu. 2011 Zebra Chip Rep. Sess.

Darling, A. E., Mau, B., and Perna, N. T. 2010. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

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.

Fagen, J. R., Leonard, M. T., McCullough, C. M., Edirisinghe, J. N., Henry, C. S., Davis, M. J., and Triplett, E. W. 2014. Comparative genomics of cultured and uncultured strains suggests genes essential for free-living growth of *Liberibacter.* PLoS One 9:e84469.

Gao, Y., Deng, X., Chen, J., and Li, H. 2011. Prophages of "*Candidatus* Liberibacter asiaticus" and their distribution in southern China. (Abstr.) Phytopathology 101:S58.

Garnier, M., Gao, S. J., He, Y. L., Villechanoux, S., Gandar, J., and Bové, J. M. 1991. Study of the greening organism (GO) with monoclonal antibodies: Serological identification, morphology, serotypes and purification of the GO. Pages 428-435 in: Proc. 11th Conf. Int. Organ. Citrus Virol. R. H. Brlansky, R. F. Lee, and L. W. Timmer, eds. International Organization of Citrus Virologists,Riverside, CA.

Gill, G. 2006. Tomato psyllid detected in New Zealand. Biosecurity 69:10-11.

Glynn, J., Islam, M., Bai, Y., Lan, S., Wen, A., Gudmestad, N., Civerolo, E., and Lin, H. 2012. Multilocus sequence typing of '*Candidatus* Liberibacter solanacearum' isolates from North America and New Zealand. J. Plant Pathol. 94:223-228.

Gudmestad, N. C., and Secor, G. A. 2007. Zebra Chip: A new disease of potato. Nebraska Potato Eyes 19:1-4.

Haapalainen, M. 2014. Biology and epidemics of *Candidatus* Liberibacter species, psyllid-transmitted plant-pathogenic bacteria. Ann. Appl. Biol. 165:172-198.

Hansen, A. K., Trumble, J. T., Stouthamer, R., and Paine, T. D. 2008. A new Huanglongbing Species, "*Candidatus* Liberibacter psyllaurous", found to infect tomato and potato, is vectored by the psyllid *Bactericera cockerelli* (Sulc). Appl. Environ. Microbiol. 74:5862-5865.

Hu, W. Z., Wang, X. E., Zhou, Y., Li, Z. A., Tang, K. Z., and Zhou, C. Y. 2011. Diversity of the omp gene in '*Candidatus* Liberibacter asiaticus' in china. J. Plant Pathol. 93:211-214.

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.

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.

Kogenaru, S., Yan, Q., Riera, N., Roper, M. C., Deng, X., Ebert, T. A., Rogers, M., Irey, M. E., Pietersen, G., Rush, C. M., and Wang, N. 2014. Repertoire of novel sequence signatures for the detection of *Candidatus* Liberibacter asiaticus by quantitative real-time PCR. BMC Microbiol. 14:39.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Larsen, H. H., Masur, H., Kovacs, J. A., Gill, V. J., Silcott, V. A., Kogulan, P., Maenza, J., Smith, M., Lucey, D. R., and Fischer, S. H. 2002. Development and evaluation of a quantitative, touch-down, real-time PCR assay for diagnosing *Pneumocystis carinii* pneumonia. J. Clin. Microbiol. 40:490-494.

Leonard, M. T., Fagen, J. R., Davis-Richardson, A. G., Davis, M. J., and Triplett, E. W. 2012. Complete genome sequence of *Liberibacter crescens* BT-1. Stand. Genomic Sci. 7:271-283.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009a. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Li, L., Stoeckert, C. J., and Roos, D. S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178-2189.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., and Wang, J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20:265-272.

Li, W., Abad, J. A., French-Monar, R. D., Rascoe, J., Wen, A., Gudmestad, N. C., Secor, G. A., Lee, I. M., Duan, Y., and Levy, L. 2009b. Multiplex real-time PCR for detection, identification and quantification of '*Candidatus* Liberibacter solanacearum' in potato plants with zebra chip. J. Microbiol. Methods 78:59-65.

Liefting, L. W., Perez-Egusquiza, Z. C., Clover, G. R. G., and Anderson, J. A. D. 2008. A new '*Candidatus* Liberibacter' species in *Solanum tuberosum* in New Zealand. Plant Dis. 92:1474.

Liefting, L. W., Weir, B. S., Pennycook, S. R., and Clover, G. R. 2009. '*Candidatus* Liberibacter solanacearum', associated with plants in the family Solanaceae. Int. J. Syst. Evol. Microbiol. 59:2274-2276.

Lin, H., Coletta-Filho, H. D., Han, C. S., Lou, B., Civerolo, E. L., Machado, M. A., and Gupta, G. 2013a. Draft genome sequence of "*Candidatus* Liberibacter americanus" bacterium associated with citrus huanglongbing in Brazil. Genome Announce. 1:e00275-13.

Lin, H., Han, C. S., Liu, B., Lou, B., Bai, X., Deng, C., Civerolo, E. L., and Gupta, G. 2013b. Complete genome sequence of a Chinese strain of "*Candidatus* Liberibacter asiaticus". Genome Announce. 1:e00184-13.

Lin, H., Islam, M. S., Bai, Y., Wen, A., Lan, S., Gudmestad, N. C., and Civerolo, E. L. 2012. Genetic diversity of '*Candidatus* Liberibacter solanacearum' strains in the United States and Mexico revealed by simple sequence repeat markers. Eur. J. Plant Pathol. 132:297-308.

Lin, H., Lou, B., Glynn, J. M., Doddapaneni, H., Civerolo, E. L., Chen, C., Duan, Y., Zhou, L., and Vahling, C. M. 2011. The complete genome sequence of '*Candidatus* Liberibacter solanacearum', the bacterium associated with potato zebra chip disease. PLoS One 6:e19135.

Mann, R. A., Smits, T. H. M., Bühlmann, A., Blom, J., Goesmann, A., Frey, J. E., Plummer, K. M., Beer, S. V., Luck, J., Duffy, B., and Rodoni, B. 2013. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. PLoS One 8:e55644.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15:589-594.

Munyaneza, J. E., Crosslin, J. M., and Upton, J. E. 2007. Association of *Bactericera cockerelli* (Homoptera: Psyllidae) with "Zebra Chip", a new potato disease in Southwestern United States and Mexico. J. Econ. Entomol. 100:656-663.

Nelson, W. R., Fisher, T. W., and Munyaneza, J. E. 2011. Haplotypes of "*Candidatus* Liberibacter solanacearum" suggest long-standing separation. Eur. J. Plant Pathol. 130:5-12.

Nelson, W. R., Sengoda, V. G., Alfaro-Fernandez, A. O., Font, M. I., Crosslin, J. M., and Munyaneza, J. E. 2013. A new haplotype of "*Candidatus* Liberibacter solanacearum" identified in the Mediterranean region. Eur. J. Plant Pathol. 135:633-639.

Nishibuchi, M., and Kaper, J. B. 1995. Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: A virulence gene acquired by a marine bacterium. Infect. Immun. 63:2093-2099.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42:D206-D214.

Pitman, A. R., Berry, N., Thompson, S., Taylor, N., Wright, P., Shah, F., Walker, M., Read, S., Beard, S., Jorgensen, N., Butler, R., and Thompson, S. 2012. Yield and quality consequences of TPP and Lso: Year 2 of a New Zealand study. Pages 26-30 in: Proc. 12th Annu. 2012 Zebra Chip Rep. Sess.

Pitman, A. R., Drayton, G. M., Kraberger, S. J., Genet, R. A., and Scott, I. A. 2011. Tuber transmission of "*Candidatus* Liberibacter solanacearum" and its association with zebra chip on potato in New Zealand. Eur. J. Plant Pathol. 129:389-398.

Secor, G. A., Rivera, V. V., Abad, J. A., Lee, I.-M., Clover, G. R. G., Liefting, L. W., Li, X., and De Boer, S. H. 2009. Association of '*Candidatus* Liberibacter solanacearum' with Zebra Chip disease of potato established by graft and psyllid transmission, electron microscopy, and PCR. Plant Dis. 93: 574-583.

Teresani, G. R., Bertolini, E., Alfaro-Fernandez, A., Martínez, C., Tanaka, F. A. O., Kitajima, E., Rosello, M., Sanjuan, S., Ferrandiz, J. C., López, M. M., Cambra, M., and Font-San-Ambrosio, M. I. 2014. Association of '*Candidatus* Liberibacter solanacearum' with a vegetative disorder of celery in Spain and development of a real-time PCR method for its detection. Phytopathology 104:804-811.

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. 2008. Comparative genomics: The bacterial pan-genome. Curr. Opin. Microbiol. 11:472-477.

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.

Wen, A., Johnson, C., and Gudmestad, N. C. 2013. Development of a PCR assay for the rapid detection and differentiation of '*Candidatus* Liberibacter solanacearum' haplotypes and their spatiotemporal distribution in the United States. Am. J. Potato Res. 90:229-236.

Wulff, N. A., Zhang, S., Setubal, J. C., Almeida, N. F., Martins, E. C., Harakava, R., Kumar, D., Rangel, L. T., Foissac, X., Bove, J., and Gabriel, D. W. 2014. The complete genome sequence of *Candidatus* Liberibacter americanus, associated with citrus Huanglongbing. Mol. Plant-Microbe Interact. 27:163-176.

Zheng, Z., Deng, X., and Chen, J. 2014. Whole-genome sequence of "*Candidatus* Liberibacter asiaticus" from Guangdong, China. Genome Announce. 2:e00273-14.

Zhou, L., Powell, C. A., Li, W., Irey, M., and Duan, Y. 2013. Prophage-mediated dynamics of '*Candidatus* Liberibacter asiaticus' populations, the destructive bacterial pathogens of citrus huanglongbing. PLoS One 8:e82248.