

# Acquisition of uncharacterized sequences from *Candidatus* Liberibacter, an unculturable bacterium, using an improved genomic walking method

Hong Lin<sup>a,\*</sup>, Harshavardhan Doddapaneni<sup>a,b</sup>, Xianjing Bai<sup>d</sup>, Jiqiang Yao<sup>a,c</sup>, Xiaolong Zhao<sup>e</sup>, Edwin L. Civerolo<sup>a</sup>

<sup>a</sup>USDA-ARS, San Joaquin Valley Agricultural Science Center, 9611 So. Riverbend Avenue, Parlier, CA 93648, USA

<sup>b</sup>Department of Viticulture and Enology, University of California, Davis, CA 95616, USA

<sup>c</sup>California Citrus Research Board, Visalia, CA 93279, USA

<sup>d</sup>Guangxi Academy of Agricultural Sciences, Nanning, China

<sup>e</sup>Guangxi Citrus Research Institute, Guilin, China

Received 20 February 2007; accepted 5 June 2007

Available online 1 July 2007

## Abstract

An effective PCR-based genomic walking approach is described to discover previously unknown flanking genomic DNA sequences from *Candidatus* Liberibacter asiaticus, an unculturable, phloem-limited bacterium. Using this technique, 8564 bp of new DNA sequences were obtained from three genomic loci; *tufB-secE-nusG-rplKJL-rpoBC* gene cluster, *omp* gene (outer membrane protein, Omp) and 16/23S rRNA gene in *Ca. L. asiaticus*. These, together with publicly available *Ca. Liberibacter* sequences, are clustered into five contigs and two singlets representing 24,477 non-redundant base pairs. BLAST annotation predicts 12 full-length genes, two partial genes and one pseudogene among these sequences. The sequences obtained in this study provide new genome information about *Ca. Liberibacter* that will facilitate development of new genome-based detection tools. The technique described here can also be employed to acquire new genomic information for other unculturable or fastidious organisms for which available sequences are limited or for filling sequence gaps between known flanking genomic DNA sequences.

Published by Elsevier Ltd.

**Keywords:** Genomic walking; Huanglongbing; *Candidatus* Liberibacter; Citrus greening

## 1. Introduction

Huanglongbing (HLB), previously known as citrus greening, is one of the most destructive citrus diseases in tropical and subtropical regions around world, and is emerging as a threat to citrus production worldwide. The causal agent of this disease is associated with *Candidatus* Liberibacter, a Gram negative, phloem-limited bacterium [1]. Taxonomically, there are currently three species, *Candidatus* Liberibacter asiaticus, *Candidatus* Liberibacter africanus and *Candidatus* Liberibacter americanus, each based on sequence variation in 16S rDNA or 16S/23S rRNA intergenic regions, and were named after the different diseases or characteristics associated with these

diseases. [2,3]. These three *Ca. Liberibacter* species cause different forms of the disease [4,5]. Studies of HLB have been impeded by the fact that the causal agent(s) have not yet been cultured on artificial nutrient media. Therefore, obtaining genomic information from the pathogen is challenging because of the lack of pure pathogen DNA.

In citrus, the HLB agent is present in very low titers in the phloem sieve tubes [6]. Periwinkle (*Catharanthus roseus* L.), an alternative host for the HLB pathogen, was reported to have higher levels of the HLB bacteria than in citrus following transmission from HLB-infected citrus plants via a dodder (*Cuscuta campestris*) [7]. This has allowed high titers of HLB-infected periwinkle plants to be produced for various types of research; however, the HLB agents are not transmitted naturally to periwinkle by psyllids [8]. Garnier et al. [1] used HLB-infected periwinkle to characterize the nature of the bacterium by electron

\*Corresponding author. Tel.: +1 559 596 2933; fax: +1 559 596 2921.

E-mail address: [hlin@fresno.ars.usda.gov](mailto:hlin@fresno.ars.usda.gov) (H. Lin).

microscopy. HLB-infected periwinkle was also used for developing monoclonal antibodies for strain differentiation studies [9,10]. In the first attempt to obtain *Ca. Liberibacter* DNA sequences, random cloning, Southern blot and dot hybridization methods from healthy and HLB-infected periwinkle DNAs were successfully used to obtain partial sequences derived from the bacterium [11]. These sequences were the *rplKAJL-rpoBC* gene cluster and a bacteriophage-type DNA polymerase gene. The loci of *rplKAJL-rpoBC* gene cluster were subsequently extended further to the *tufB-secE-nusG-rplKAJL-rpoBC* gene cluster using thermal asymmetric interlaced polymerase chain reaction [12]. Using PCR, Jagoueix et al. identified the 16S ribosomal DNAs (rDNAs) of Asian and African *Ca. Liberibacter* strains [13,14]. Sequence analyses suggested that they are members of the alpha subdivision of the class Proteobacteria. Because there is only 87.5% sequence homology to this class, these authors suggested that they represent a new lineage and proposed the genus name *Liberobacter* [13], which was corrected to *Liberibacter* in agreement with its Latin gender [2]. More recently, using 102 random 10-mer primers to screen DNAs derived from HLB-infected periwinkle plants and healthy plants, Hocquellet et al. identified another sequence locus, which belong to *nusG*, *pgm*, *omp* and a hypothetical protein gene [15].

Currently, there are 71 *Ca. Liberibacter* genomic DNA sequence accessions available in the NCBI GenBank, seven of which were deposited from this study. Prior to our submissions, 43 (67%) of the accessions were located in the 16S/23S rDNA region. The total non-redundant sequences for *Ca. Liberibacter* consist of 15,664 bp. Limited genomic sequence information impedes understanding molecular basis of the pathogenicity of this pathogen.

The new genomic sequences obtained in this study provide additional information for better understanding the structure and organization of the *Ca. Liberibacter* genome. This information can be useful for developing new, reliable diagnostic tools with improved specificity. Some PCR-based detection methods currently being used are based on the 16S rRNA gene region. Because the 16S conserved sequences are homologous to sequences of the host and/or citrus-associated endophytes, reliable and specific detection and diagnosis of HLB is of concern. The technique described in this study can also be employed to acquire new genomic information in any other fastidious or unculturable organism where available sequences are limited or to fill in sequence gaps between known flanking genomic DNA sequences.

## 2. Materials and methods

### 2.1. Plant materials and DNA isolation

HLB-infected leaves of several citrus cultivars were collected from groves in Guilin City, Guangxi Province, China. Asymptomatic leaves were collected from healthy plants of the same cultivars grown in a greenhouse and

Table 1

Isolates of HLB-infected citrus and healthy citrus used in this study

| Sample ID         | Location               | Host   |
|-------------------|------------------------|--|
| GL-1 <sup>a</sup> | Guilin, Guangxi        | Infected, <i>C. reticulata</i> cv Ponkan No1     |
| GL-2 <sup>a</sup> | Guilin, Guangxi        | Infected, <i>C. sinensis</i> cv Newhall Navel    |
| GL-3 <sup>a</sup> | Guilin, Guangxi        | Infected, <i>C. grandis</i> Pummelo              |
| GL-4 <sup>a</sup> | Guilin, Guangxi        | Infected, <i>C. reticulata</i> cv Satsuma        |
| GL-5 <sup>a</sup> | Guilin, Guangxi        | Healthy, <i>C. reticulata</i> cv Ponkan No1      |
| GL-6 <sup>a</sup> | Guilin, Guangxi        | Healthy, <i>C. sinensis</i> cv Newhall Navel     |
| GL-7 <sup>a</sup> | Guilin, Guangxi        | Healthy, <i>C. grandis</i> Pummelo               |
| GL-8 <sup>a</sup> | Guilin, Guangxi        | Healthy, <i>C. reticulata</i> cv Satsuma         |
|                   | USDA, ARS, Parlier, CA | Healthy, <i>C. reticulata</i> cv Clementina fina |
|                   | USDA, ARS, Parlier, CA | Healthy, <i>C. limon</i> cv Eureka Lemon         |

<sup>a</sup>DNA samples, GL-1 to GL-8 were received from Guangxi Citrus Research Institute, Guilin, China.

used as negative control tissues (Table 1). Leaves were rinsed with water. Leaf midribs were excised from leaf blade and kept at  $-20^{\circ}\text{C}$  until used. Total genomic DNA was extracted from 5.0 g of collected tissue as previously described [16]. Precipitated DNA was re-dissolved in TE buffer. DNA samples were analyzed on 1.0% agarose gels and concentrations were determined by spectrophotometry.

### 2.2. Genomic walking adaptor design

Genomic walking adaptors were modified from the previously described method [17]. Instead of using blunt end restriction enzymes, adaptor sequences were modified for sticky end restriction enzymes *HindIII*, *EcoRI* and *PstI*. The restriction site was redesigned in such a way that the restriction site is no longer recognized once it is ligated with the digested genomic DNA (Fig. 1). A shorter strand adaptor was modified with an amine group at the 3' end (modified oligo was synthesized by W.M. Keck Facility, Yale University). To prepare double strand adaptors, equal amounts of 100  $\mu\text{M}$  of single stranded complementary oligo nucleotides were mixed and heated at  $95^{\circ}\text{C}$  for 4 min and gradually cooled to room temperature.

### 2.3. Preparation genomic walking templates

Total genomic DNA from HLB-infected citrus leaf midrib tissue was adjusted to a concentration of 1  $\mu\text{g}/\mu\text{l}$ . Five microgram of this DNA was digested with each of the restriction enzymes *HindIII*, *EcoRI* and *PstI* separately as follows. Fifty microliter of  $1 \times$  digestion buffer with 40 U of enzyme was incubated at  $37^{\circ}\text{C}$  overnight. Each digestion was checked by electrophoresis of 10  $\mu\text{l}$  of digested DNA on a 1.2% gel in  $1 \times$  TBE buffer. Digested DNA was then purified once by chloroform-phenol extraction followed by ethanol precipitation. The pellet was re-dissolved in 20  $\mu\text{l}$  of water. From each type of restricted DNA, 5  $\mu\text{l}$  was then mixed with 15  $\mu\text{l}$  of ligation mixture containing 10 pmole of adaptor, 1 U of T4 DNA

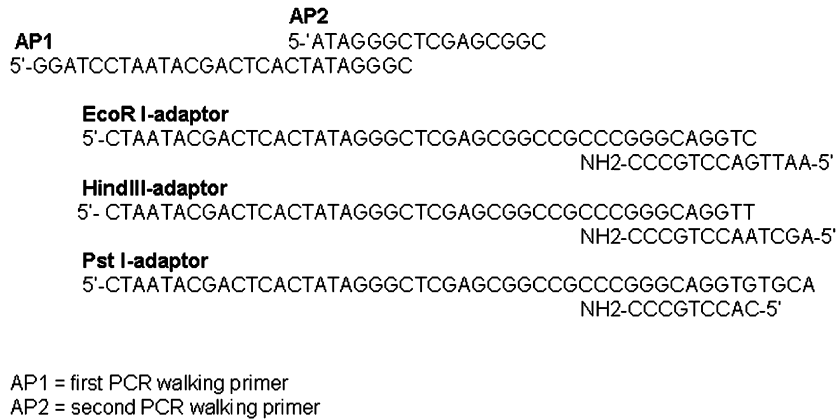


Fig. 1. Details of adaptor and genomic walking primer sequences used in this study.

Table 2  
Sequences of locus specific primers (LSP-1 and LSP-2)

| Locus specific primer-1 (first walking primer) | Locus specific primer-2 (nested primer) | Walking (bp) |
|--|---|--------------|
| GC-1-L GTCGCAATCGTAATACCCCGTAATTTCTCTTC        | Nested-GC-1-L TGGAGCACTATCAATATCGC      | 1506         |
| GC-2-L CACAACAAGAACCCTTCTCTTTGACGCC            | Nested-GC-2-L TAACTGCACTGTGACTCAA       | 1528         |
| GC-1-R GGAGCGTTCTATTAAGGAACGCATTTTCATCAG       | Nested-GC-1-R ATTCAGTGATGCCTCAAGATC     | 942          |
| GC-2-R TCGCAGAAATGATATTATAGCAGACGGTCC          | Nested-GC-2-R GATTTAGGAGATCTGGCTCT      | 1141         |
| 16S-1-L AAAATCTTTCCCCCAATAGGGCGTATACGG         | Nested-16S-1-L CTCACCCGTCTGCCGCTCGTA    | 302          |
| 23S-1-R TTCCGAATGGGGTAACCCGCCATTATGTGC         | Nested-23S-1-R CTAAGTTGGTTTAATTTCTAGG   | 1152         |
| OMP-1-L CCTCTCGTAATTGAACTAACCTCATCTAATCCTC     | Nested-OMP-1-L GACTGCAAAAACCGTCTACTGT   | 756          |
| OMP-1-R ATTGATCGTGGTACAATGGGTGATACAATCATTGG    | Nested-OMP-1-R CGAAGATTGATAATCAGGTGC    | 1237         |
|  |   | 8564         |

ligase, 4  $\mu$ l of 5  $\times$  ligation buffer and 5 U of each restriction enzyme. Ligation tubes were maintained at 16  $^{\circ}$ C overnight. The ligation mixture was diluted 10-fold with water and used for PCR reaction.

#### 2.4. PCR amplification

PCR amplifications were performed from each restricted-adaptor-ligation mixture. The 20  $\mu$ l PCR volume contained 1  $\mu$ l of 10  $\times$  dilution of restricted-adaptor-ligation mixture, 1  $\times$  reaction buffer, 2.0 mM MgOAc, 0.2 mM dNTPs, 0.5  $\mu$ M adaptor primer AP1 (Fig. 1) and locus-specific primer LSP1 (Table 2), and 0.5 U XL DNA *Tth* polymerase (ABI, Foster city, CA). The thermocycler parameters (ABI 7700, Foster city, CA) were denatured at 94  $^{\circ}$ C for 1 min following by 7 cycles of 94  $^{\circ}$ C for 10 s, annealing/extension at 72  $^{\circ}$ C for 6 min then followed by another 32 cycles of 94  $^{\circ}$ C for 10 s and 67  $^{\circ}$ C for 6 min. After the first round of PCR, amplified products were diluted 10  $\times$  with water and 2  $\mu$ l of this diluent was used for a second round of PCR with nested primers, AP2 (Fig. 1) and LSP2 (Table 2). The PCR conditions were 94  $^{\circ}$ C for 1 min followed by 5 cycles of 94  $^{\circ}$ C for 10 s, and 62  $^{\circ}$ C for 6 min and then followed by another 20 cycles of 94  $^{\circ}$ C for 10 s, 65  $^{\circ}$ C for 6 min. PCR products were

separated and examined on 1% agarose gels containing ethidium bromide.

#### 2.5. Cloning and sequencing

Nested PCR products ranging from 500 to 2000 bp were collected from the gels and purified. DNA fragments were ligated to pGEM<sup>®</sup>-T Easy Vector (Promega, Madison, WI, USA) and transformed into JM109 competent cells per the manufacturer's instructions. Three colonies per transformation were selected. The plasmid DNA was isolated and purified using QIAprep 96 well kit (Qiagen, Valencia, CA) per the manufacturer's specifications. The purified plasmid DNA was used for sequencing reaction with BigDye Terminator v3.1 Cycle Sequencing Kits (ABI) run by ABI 3100 DNA analyzer.

#### 2.6. Sequence data analyses

Sequence data were first masked to remove plasmid sequences. Clean sequence files amplified from the three restricted-ligation DNA templates were analyzed by CAP3 software [18]. The consensus sequences were generated from contigging multiple overlapping sequencing files. BLAST analysis was performed against nucleotide

sequences in NCBI GenBank with cut-off *E* value of 1E05 to check for eukaryotic DNA contamination.

### 2.7. Sequence validation

To confirm that the newly obtained DNA sequences have linearly extended from the previously known sequences, we designed confirmation primers. These primers were designed at locations that will amplify DNA covering both known and newly obtained sequence regions. Four HLB-infected DNA samples collected from various orchards along with six HLB-free citrus DNA samples were used for PCR tests (Table 1). The tests were carried out to confirm that the expected size of PCR products were amplified only from HLB-infected DNA samples and not from the HLB-free DNA samples. The bands amplified from HLB-infected DNA samples were collected and re-sequenced to confirm their sequence identities. The confirmed sequences were then used for designing the next round of genomic walking primers.

## 3. Results

### 3.1. Genomic walking

Three loci, *tufB-secE-nusG-rplKAJL-rpoBC* gene cluster, outer membrane protein, and 16/23S rRNA gene were chosen for genomic walking in this study. In the gene cluster region, a locus specific primer, GC-1-L, and a nested primer, Nested-GC-1-L, were used to walk from the 5' end of the known sequence. This round of genomic walking yielded 1506 bp of new sequence. The second round of genomic walking using primers GC-2-L and nested-GC-2-L extended another 1528 bp of new sequence. The second round of genomic walking using primers GC-2-L and Nested-GC-2-L extended another 1528 bp of additional sequences. Thus, new genomic DNA sequences of more than 3000 bp were obtained from 5' end of the gene cluster (Table 2). At the 3' downstream end of the gene cluster, the locus specific primer GC-1-R and Nested-GC-1-R were used to generate new sequences of 942 bp in the first round of genomic walking. In the second round of genomic walking, 1141 bp of new sequences were obtained

using primers GC-2-R and Nested-GC-2-R. Thus, a total of 2083 bp of new sequences were extended at the 3' downstream end of the gene cluster region (Table 2). Similarly, in outer membrane protein region, 756 bp of new sequences were acquired from 5' upstream end using primer OMP-1-L and Nested-OPM-1-L and 1237 bp of new sequence was obtained from 3' downstream end using OMP-1-R and Nested-OMP-1-R. For the 16S/23S rRNA gene region, primers 16S-1-L and Nested-16S-1-L were used to extend 302 bp of new sequences upstream from the 5' end of the 16S locus while 1152 bp of new sequences were extended downstream from the 3' end using primers 23S-1-R and Nested-23S-1-R. Thus, overall, a total of approximately 8564 bp of new genomic DNA sequences were obtained for three loci of the *Ca. L. asiaticus* using our modified genomic walk method (Table 2).

### 3.2. Sequence validation

Each newly obtained sequence was confirmed by confirmation PCR primers that extended from known to newly cloned sequences (Table 3). Amplicons from all four DNA samples from HLB-infected DNA samples were of the expected sizes (Fig. 2). No amplified DNA product was observed in six HLB-free DNA samples. The sequence identity of each PCR product was further confirmed by re-sequencing the amplicons. Sequence analysis of each PCR fragment showed that each walking region started with the expected length of known sequences and was contiguous with the uncharacterized sequence region suggesting that the digestion, ligation and PCR-based locus specific primer worked properly.

### 3.3. Sequence analyses

#### 3.3.1. Gene cluster region

Villechanoux et al. first reported a gene cluster locus with 3197 bp (GenBank accession nos. M94319) in *Ca. Liberibacter-HLB* and identified it as the *rplKAJL-rpoBC* gene cluster, a well-known operon in eubacteria [11,19]. Okuda et al. [12] extended this locus at the 5' end to include a total of 6145 bp (GenBank accession nos. AY342001) and characterized it as the “*tufB-secE-nusG-rplKAJL-rpoB*”

Table 3  
Genomic walking confirmation primers

|              | Forward primer         | Reverse primer           | bp   |
|--------------|------------------------|--------------------------|------|
| Conf-GC-1-L  | GACGCACGTCGTGTGTGTGAT  | GACTCTCCTTGTTACGAACATATC | 1326 |
| Conf-GC-2-L  | TTGTAGTCATGAATACTCCTT  | CGACGTGCGTCATACACATC     | 1585 |
| Conf-GC-R    | AATTTTTCTGTTCTCGCAGC   | TTCTTTCCCAAGACCAACACATGC | 1831 |
| Conf-16S-1-L | TTGAGTATTATTGGTTGGAG   | CTCCTAAACTCTAGACAACC     | 861  |
| Conf-16S-1-R | GAAGCGAACGCAGGGAAGCTG  | TTACGGAACGCTCCGCTACC     | 1022 |
| Conf-OMP-1-L | TGGAAGAAGATATAGAGTCCG  | CCTACTGTGCAATTTGGTC      | 678  |
| Conf-OMP-1-R | TAACGTTTCATATAGGTTGTGG | GGTTAGCTTACAATCATGCGC    | 1141 |

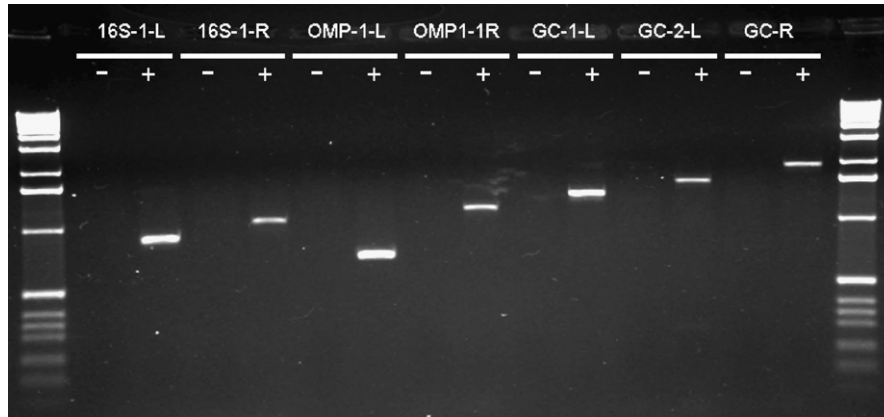


Fig. 2. Seven pairs of sequence confirmation primers (see Table 3) were used to validate new sequences obtained by genomic walking. PCR products were separated on a 1.2% agarose gel along with 1 kb molecular marker. Picture above showed expected amplicon sizes from HLB-infected DNA (+) but no products were observed in HLB-free DNA samples (-).

gene cluster. Using the genomic walking technique described herein, we extended this region further up to a total of 2920 bp at the 5' end and 2329 bp at the 3' end. Accordingly, this region is now 10,989 bp long. BLAST analysis suggested that a partial D-3-phosphoglycerate dehydrogenase gene is located from 64 to 1296 bp. Since it contains internal stop codons, it is likely a pseudogene. This was followed by a full-length gene coding for tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (1353–2579 bp) (GenBank accession nos. [DQ778017](#)). In addition, 57 bp from the beginning of this genomic walking combined with previously reported partially coding gene, the protein chain elongation factor, EF-Tu, is now full length (2864–4042) (GenBank accession nos. [EF164805](#)). At the 3' downstream end of this gene cluster region, 2329 bp of new sequences were obtained. This new sequence along with previous partial sequence of RNA polymerase beta subunit belongs to *rpoB* gene ranging from 7487 to 10,945 (GenBank accession nos. [EF164807](#)) (Table 4).

### 3.3.2. OMP gene region

Hocquellet et al. [15] used 10-mer RAPD primers (Operon Technologies Inc., Alameda, CA, USA) to screen HLB-infected periwinkle (*C. roseus* L.) and sweet orange [*Citrus sinensis* (L.) Osb.] [15]. From 3 of 102 primer screens, six fragments which were present in HLB-infected plants but not in healthy plants were identified and cloned. Sequence analysis showed that they are part of an *omp* gene locus. Chromosome walking [20] extended the 5' and 3' uncharacterized sequences flanking the *Ca. L. asiaticus omp* locus. Further analysis showed that the *Ca. L. africanus* and *Ca. L. asiaticus omp* genes were 2340 and 2346 bp long, respectively (GenBank accession nos. [AY642159](#)). Attempts to amplify the 3' part of the *Ca. L. africanus* gene by chromosome walking failed [20]. Using our improved genomic walking approach, we extended this region by 692 bp at the 5' upstream end and 1272 bp at the 3' downstream end. This resulted in generating full-length

zinc metalloprotease (47–1096 bp) (GenBank accession nos. [EF164804](#)) at 5' the upstream end and full-length 3-hydroxydecanoyl-[acyl carrier protein] at the 3' downstream end (4583–5068 bp) (GenBank accession nos. [EF164806](#)) (Table 4).

### 3.3.3. 16/23S rDNA region

Using PCR, Jagoueix et al. was able to amplify prokaryotic 16S rRNA from HLB-infected periwinkle plants with universal primers f-D1/r-P1 [21]. This 1450–1500 bp fragment was cloned and sequenced [13]. Hybridization and PCR experiments performed with oligonucleotides specific for the amplified sequences revealed that DNAs obtained were indeed the 16S rDNAs of the HLB bacteria [13]. In this study, we cloned 356 bp of new sequence from the 5' upstream end and 1260 bp from 3' downstream end. Thus, there are now a total of 3712 bp representing this region. BLAST analysis identified ribosomal rRNA extending from 377 to 1830 bp followed an intergenic region and sequence coding for miscellaneous RNAs such as tRNA-Alanine and tRNA-Isoleucine. The newly cloned 1260 bp sequence at 3' downstream end (2451–3712 bp) GenBank accession nos ([EF438153](#)) represents the 23S ribosomal RNA region.

## 4. Discussion

Various DNA walking techniques have been developed to obtain unknown sequences adjoining known regions. These include self-ligation of restricted fragments also known as inverse PCR [22,23], ligation to specifically designed adaptors [24–26], and direct PCR using locus specific primers and degenerate primers [27,28]. These methods are useful for many applications, such as acquiring promoter sequences from ESTs [29,30], filling gaps in genome maps [31], and mapping intron/extron junctions [32]. However, these methods are usually used for homogenous DNA samples. In the case of mixed DNA, particularly in the case where the target DNA is only a

Table 4  
Summary of gene predictions of *Ca. Liberibacter* bacteria

| S. no.                                | Gene name and characteristics  | Size   | Citation                            |
|---------------------------------------|--|--|-------------------------------------|
| <b>A. Gene cluster region (5'–3')</b> |  |  |                                     |
| 1                                     | D-3-phosphoglycerate dehydrogenase (partial, has internal stop codons, a definite pseudogene; first in this organism).   | 397 aa gene co-ordinates, 64–1296 bp                     | This study                          |
| 2                                     | tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase). Full-length.  | 408 aa; 1,227 bp gene co-ordinates, 1353–2579 bp         | This study DQ778017                 |
| 3                                     | Protein chain elongation factor EF-Tu; We filled the first 19 aa and now its full-length.  | 392 aa; 1179 bp gene co-ordinates, 2864–4042 bp          | [12] AY342001, this study EF164805  |
| 4                                     | Preprotein translocase   | 67 aa; 204 bp gene co-ordinates, 4277–4480bp             | [12] AY342001                       |
| 5                                     | Transcription antitermination component  | 177 aa; 534 bp gene co-ordinates, 4502–5035 bp           | [12,19], M94319                     |
| 6                                     | 50S ribosomal subunit protein L11  | 142 aa; 429 bp; gene co-ordinates, 5126–5554 bp          | [12,19], M94319                     |
| 7                                     | 50S ribosomal subunit protein L1   | 232 aa; 699 bp gene co-ordinates, 5556–6254 bp           | [12,19], M94319                     |
| 8                                     | 50S ribosomal subunit protein L10  | 130 aa; 393 bp gene co-ordinates, 6553–6945 bp           | [12,19], M94319                     |
| 9                                     | 50S ribosomal subunit protein L7/L12   | 126 aa; 381 bp gene co-ordinates, 6994–7374 bp           | [12,19], M94319                     |
| 10                                    | RNA polymerase beta subunit. Partial protein submitted by Okuda et al., 2005. We cloned the full-length protein (added 1880 aa at the 3'-end from our strain.) | 1258 aa; 3774 bp gene co-ordinates, 7487–10,945 bp       | [12,19] this study EF164807         |
| <b>B. OMP region</b>                  |  |  |                                     |
| 11                                    | Zinc metalloprotease. We added the 5' 209 aa. Now full-length.   | 349 aa; 1050 bp gene co-ordinates, 47–1096 bp, 1–1147 bp | [20], AY642159, this study EF164804 |
| 12                                    | Putative group 1 outer membrane protein  | 781 aa; 2346 bp gene co-ordinates, 1163–3508 bp          | [20], AY642159                      |
| 13                                    | UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase protein.  | 374 aa; 1044 bp; gene co-ordinates, 3538–4326 bp         | [20], AY642159, this study          |
| 14                                    | 3-hydroxymyristoyl/3-hydroxydecanoyl-[acyl carrier protein] most probably full-length.   | 161 aa; 486 bp gene co-ordinates, 4583–5068 bp           | This study EF164806                 |
| 15                                    | Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase protein (partial sequence)   | 132 aa; 396 bp gene co-ordinates, 5070–5447 bp           | This study                          |
| <b>C. 16/23S region</b>               |  |  |                                     |
| 16                                    | rRNA product 16S ribosomal RNA   | 377–1830 bp  | This study, DQ778016                |
| 17                                    | Misc_RNA product 16S-23S ribosomal RNA intergenic spacer   | 1831–2450 bp   | [13,14], this study, DQ778016       |
| 18                                    | rRNA product 23S ribosomal RNA, partial sequence (not complete yet)  | 2451–3712 bp   | This study, EF438153                |

small fraction of the total DNA, a more sensitive and more specific method is needed. While direct PCR with degenerate primers is more sensitive than other methods, this method, in many cases, results in high numbers of non-specific PCR products, which makes it difficult to distinguish a target product from the background. In order to selectively amplify the pathogen's target locus mixed with the host DNA, we modified the genome walking method based on the protocol described by [17]. We adapted this method by using an amine group at the 3' end of the short strand of oligo adaptor. The amine group serves as "blocker". This method effectively eliminates amplification that occurs from both AP primers, which would otherwise most likely amplify host DNA. To effectively eliminate the occurrence of such amplification, all synthesized oligo adaptors were HPLC-purified to ensure that oligos at the 3' ends were uniformly modified

with an amine group. In case such amplification does occur, amplified products will form a panhandle structure because of complementary sequences at both ends of the adaptor. This "suppression effect" of the panhandle structure will prevent further amplifications. Therefore, only the strand that was extended from the LSP primer will be paired with AP primer for exponential amplification. These modifications appear to improve specificity of PCR amplification. The protocol successfully extended the sequence of the specific locus of the pathogen genomic DNA by about 500–2000 bp in each round of genomic walking.

One of critical precautions in dealing with mixed DNA samples is to avoid the formation of chimeric molecules or artificial recombination of molecules during DNA manipulations, such as restriction, ligation and amplification processes. Kopczynski et al. [33] reported chimeras of 16S

rDNA sequences in a cloned library. These chimeras seemed to be derived from environmental 16S rRNA genes of uncultivated microorganisms as judged by different sequence homologies of 5' and 3' halves. In the case of mixed DNA, this is a special concern because HLB DNA is usually present as a small proportion of the total DNA from citrus. Using quantitative PCR, the ratio of copies of target pathogen DNAs to the copies of host has been estimated to be approximately 1:1000 or less [34]. Recombinants between host and pathogen DNA fragments and/or authentic DNA fragments from bacteria but in wrong linear order during the ligation will cause errors and be difficult to detect since the pathogen's genome is unknown. To minimize potential error and to further improve the method's specificity, we re-designed adaptor sequences using sticky end restriction enzymes. Sequences of adaptors were designed such that restriction enzymes would no longer cut once adaptors were ligated to restricted genomic DNA. By maintaining restriction enzyme activity during the ligation, conditions favor ligation occurrence between adaptors and genomic DNA fragments and minimize chimerical ligations between genomic fragments.

Our genomic walking sequence data showed that the AP2 and LSP2 primer sequences were aligned as expected. Another important point is that primers LSP1 and LSP2 were designed at about 60–100 bp prior to the end of known sequences. The beginning of the known sequences provides a linear reference before walking into uncharacterized regions. These modifications appear to be critical to ensure specific extension of HLB pathogen genomic DNA. Using this approach, we successfully obtained 8564 bp new sequences from three genomic loci (16S/23S rRNA, gene cluster and the OMP regions) of *Ca. L. asiaticus*. These, together with current publicly available genomic DNA sequence, are clustered into five contigs and two singlets representing 24,477 non-redundant base pairs. BLAST analyses predict 12 full length, two partial genes, one pseudogene, and intergenic regions among these sequences.

## 5. Conclusion

The improved genomic walking method described here is valuable for identifying new, uncharacterized DNA sequences flanking known sequence regions in samples containing mixed target and non-target DNA templates. In this study, we successfully extended the genomic sequences of three loci of upstream and downstream in *Ca. L. asiaticus*. Multiple genome walking runs can be strung together to acquire longer sequences. This method, however, is still relatively time consuming since it involves designing, synthesizing genome walking primers, performing the first and nested PCR, cloning and sequencing amplicons for each walking run. The method may encounter difficulty when walking through low complex sequence regions such as long-stretch simple sequence

repeats or walking on multiple copy regions in a genome. Regardless of these limitations, this technique is a valuable tool for rapidly finding upstream and downstream uncharacterized genomic DNA sequences from known tagged sites.

## Acknowledgments

We thank Sahota Parminder and Huiqing Cao for technical support. We would also like to thank Dr. Ryokomi Yokomi for his critical comments on the manuscript. This work was funded by the United States Department of Agriculture, Agricultural Research Service.

## References

- [1] Garnier M, Danel N, Bove JM. The greening organism is a Gram negative bacterium. In: Proceedings of ninth conference IOCV. Riverside: IOCV; 1984, p. 115–24.
- [2] Garnier M, Jagoueix-Eveillard S, Cronje PR, Le Roux HF, Bove JM. Genomic characterization of a Liberibacter present in an ornamental rutaceous tree, *Calodendrum capense*, in the Western Cape province of South Africa. Proposal of '*Candidatus* Liberibacter africanus subsp. capensis'. Int J Syst Evol Microbiol 2000;50(6):2119–25.
- [3] Teixeira DdC, Danet JL, Eveillard S, Martins EC, Junior WCdJ, Yamamoto PT, et al. Citrus huanglongbing in Sao Paulo state, Brazil: PCR detection of the '*Candidatus*' Liberibacter species associated with the disease. Mol Cell Probes 2005;19(3):173–9.
- [4] Bove JM, Calavan EC, Capoor SP, Cortez RE, Schwarz RE. Influence of temperature on symptoms of California stubborn, South African greening, Indian citrus decline and Philippines leaf mottling diseases. In: Proceedings of the sixth conference of the international organization of citrus virology. Weathers LG, Cohen M, (editors), Department of Plant Pathology, University of California Riverside, CA, USA, 1974. p. 12–15.
- [5] Halbert SE, Manjunath KL. Asian citrus psyllids (Sternorrhyncha: Psyllidae) and greening disease of citrus: a literature review and assessment of risk in Florida. Fla Entomol 2004;87(3):330–53.
- [6] Ghosh SK, Giannotti J, Louis C. Intense multiplication of prokaryotes associated with greening disease of citrus in phloem cells of dodder. Ann Phytopathol 1978;9(4):525–30.
- [7] Garnier M, Bove JM. Transmission of the organism associated with citrus greening disease from sweet orange *Citrus-sinensis* cultivar Madame-vinosa to periwinkle Vinca-Rosea by dodder *Cuscuta-Campestris*. Phytopathology 1983;73(10):1358–63.
- [8] Bove JM. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. J Plant Pathol 2006;88(1):7–37.
- [9] Villechanoux S, Garnier M, Bove JM. Purification of the bacterium-like organism associated with greening disease of citrus by immunoaffinity chromatography and monoclonal antibodies. Curr Microbiol 1990;21(3):175–80.
- [10] Gao S, Garnier M, Bove JM. Production of monoclonal antibodies recognizing most strains of the greening BLO by in vitro immunization with an antigenic protein purified from BLO. In: Proceedings of 12th conference IOCV, Riverside: IOCV; 1993, p. 244–9.
- [11] Villechanoux S, Garnier M, Renaudin J, Bove JM. Detection of several strains of the bacterium-like organism of citrus greening disease by DNA probes. Curr Microbiol 1992;24(2):89–96.
- [12] Okuda M, Matsumoto M, Tanaka Y, Subandiyah S, Iwanami T. Characterization of the *tufB-secE-nusG-rplKAL-rpoB* gene cluster of the citrus greening organism and detection by loop-mediated isothermal amplification. Plant Dis 2005;89(7):705–11.
- [13] Jagoueix S, Bove J-M, Garnier M. The phloem-limited bacterium of greening disease of citrus is a member of the alpha subdivision of the Proteobacteria. Int J Syst Bacteriol 1994;44(3):379–86.

- [14] Jagoueix S, Bove J-M, Garnier M. Comparison of the 16S/23S ribosomal intergenic regions of “*Candidatus liberobacter asiaticum*” and “*Candidatus liberobacter africanum*,” the two species associated with citrus huanglonging (Greening) disease. *Int J of Systemat Bacteriol* 1997;47(1):224–7.
- [15] Hocquellet A, Bove JM, Garnier M. Isolation of DNA from the uncultured “*Candidatus Liberobacter*” species associated with citrus huanglongbing by RAPD. *Curr Microbiol* 1999;38(3):176–82.
- [16] Lin H, Walker MA. Extracting DNA from cambium tissue for analysis of grape rootstocks. *HorScience* 1997;32(7):1264–6.
- [17] Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res* 1995;23(6):1087–8.
- [18] Huang X, Madan A. CAP3: a DNA sequence assembly program. *Genome Res* 1999(9):868–77.
- [19] Villechanoux S, Garnier M, Laigret F, Renaudin J, Bove J-M. The genome of the non-cultured, bacterial-like organism associated with citrus greening disease contains the *nusG-rplKAL-rpoBC* gene cluster and the gene for a bacteriophage type DNA polymerase. *Curr Microbiol* 1993;26(3):161–6.
- [20] Bastianel C, Garnier-Semancik M, Renaudin J, Bove JM, Eveillard S. Diversity of “*Candidatus Liberibacter asiaticus*,” based on the omp gene sequence. *Appl Environ Microbiol* 2005;71(11):6473–8.
- [21] Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J bacteriol* 1991(173):697–703.
- [22] Pham H-S, Kiuchi A, Tabuchi K. Methods for rapid cloning and detection for sequencing of cloned inverse PCR-generated DNA fragments adjacent to known sequences in bacterial chromosomes. *Microbiol Immunol* 1999;43(9):829–36.
- [23] Ren M, Chen Q, Li L, Zhang R, Guo S. Successive chromosome walking by compatible ends ligation inverse PCR. *Mol Biotechnol* 2005;30(2):95–101.
- [24] Padegimas LS, Reichert NA. Adaptor ligation-based polymerase chain reaction-mediated walking. *Anal Biochem* 1998;260(2):149–53.
- [25] Lin J, Liu J, Sun X, Zhou X, Fei J, Tang K. An efficient method for rapid amplification of *Arisaema heterophyllum* agglutinin gene using a genomic walking technique. *Prep Biochem Biotechnol* 2005;35(2):155–67.
- [26] Xu P, Hu R-Y, Ding X-Y. Optimized adaptor polymerase chain reaction method for efficient genomic walking. *Acta Biochim Biophys Sin* 2006;38(8):571–6.
- [27] Chimwamurombe PM, Wingfield BD, Botha A-M, Wingfield MJ. Cloning and sequence analysis of the endopolygalacturonase gene from the pitch canker fungus, *Fusarium circinatum*. *Curr Microbiol* 2001;42(5):350–2.
- [28] Ridlon J, Hylemon P. Genome-walking by PCR and characterization of a bile acid 7 alpha-dehydroxylating operon in *clostridium hylemonae* Tn271 involved in secondary bile acid production in the human colon. *Gastroenterology* 2006;130(4, Suppl. 2).
- [29] Wu A-M, Liu J-Y. An improved method of chromosome walking for promoter sequences cloning. *Zhongguo Shengwu Huaxue yu Fenzi Shengwu Xuebao* 2006;22(3):243–6.
- [30] Oubrahim H, Wang J, Stadtman E, Chock PB. Identification of caspase-12 promoter region. *FASEB J* 2004;18(8, Suppl. S):C203–4.
- [31] Wang C-l, Chen L-t, Zeng C-z, Zhang Q-y, Liu P-q, Liu Y-g, et al. Chromosome walking for fine mapping of Xa23 gene locus by using genomic libraries. *Zhongguo Shuidao Kexue* 2006;20(4):355–60.
- [32] Chawla R, DeMason D. Genome walking in pea: an approach to clone unknown flanking sequences. *Pi sum Genet* 2003;35:1–2.
- [33] Kopeczynski ED, Bateson MM, Ward DM. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. *Appl Environ Microbiol* 1994(60):746–8.
- [34] Li W, Hartung JS, Levy L. Quantitative real-time PCR for detection and identification of *Candidatus Liberibacter* species associated with citrus huanglongbing. *J Microbiol Meth* 2006;66(1):104–15.