The *tufB–secE–nusG–rplKAJL–rpoB* gene cluster of the liberibacters: sequence comparisons, phylogeny and speciation

D. C. Teixeira,[1] S. Eveillard,[2,3] P. Sirand-Pugnet,[2,3] A. Wulff,[1] C. Saillard,[2,3] A. J. Ayres[1] and J. M. Bové[2,3]

Correspondence
S. Eveillard
Sandrine.Eveillard
@bordeaux.inra.fr

[1]Fundecitrus, Araraquara, SP, CEP 14807-040, Brazil

[2]Université de Bordeaux 2, UMR 1090, F-33883 Villenave d'Ornon, BP 81, France

[3]INRA, UMR 1090, F-33883 Villenave d'Ornon, BP 81, France

The *rplKAJL–rpoBC* operon or β operon is a classic bacterial gene cluster, which codes for proteins K, A, J and L of the large ribosomal subunit, as well as proteins B (β subunit) and C (β' subunit) of RNA polymerase. In the early 1990s, the operon was obtained as a 2.6 kbp DNA fragment (In-2.6) by random cloning of DNA from periwinkle plants infected with the Poona (India) strain of the huanglongbing agent, later named '*Candidatus* (*Ca.*) Liberibacter asiaticus'. DNA from periwinkle plants infected with the Nelspruit strain (South Africa) of '*Ca.* L. africanus' was amplified with a primer pair designed from In-2.6 and yielded, after cloning and sequencing, a 1.7 kbp DNA fragment (AS-1.7) of the β operon of '*Ca.* L. africanus'. The β operon of the American liberibacter, as well as the three upstream genes (*tufB*, *secE*, *nusG*), have now also been obtained by the technique of chromosome walking and extend over 4673 bp, comprising the following genes: *tufB*, *secE*, *nusG*, *rplK*, *rplA*, *rplJ*, *rplL* and *rpoB*. The sequence of the β operon was also determined for a Brazilian strain of '*Ca.* L. asiaticus', from *nusG* to *rpoB* (3025 bp), and was found to share 99 % identity with the corresponding β operon sequences of an Indian and a Japanese strain. Finally, the β operon sequence of '*Ca.* L. africanus' was extended from 1673 bp (*rplA* to *rpoB*) to 3013 bp (*nusG* to *rpoB*), making it possible to compare the β operon sequences of the African, Asian and American liberibacters over a length of ~3000 bp, from *nusG* to *rpoB*. While '*Ca.* L. africanus' and '*Ca.* L. asiaticus' shared 81.2 % sequence identity, the percentage for '*Ca.* L. americanus' and '*Ca.* L. africanus' was only 72.2 %, and identity for '*Ca.* L. americanus' and '*Ca.* L. asiaticus' was only 71.4 %. The ~3000 bp *nusG–rpoB* sequence was also used to construct a phylogenetic tree, and this tree was found to be identical to the known 16S rRNA gene sequence-based tree. These results confirm earlier findings that '*Ca.* L. americanus' is a distinct liberibacter, more distantly related to '*Ca.* L. africanus' and '*Ca.* L. asiaticus' than '*Ca.* L. africanus' is to '*Ca.* L. asiaticus'. The dates of speciation have also been estimated.

INTRODUCTION

'*Candidatus* (*Ca.*) Liberibacter americanus' was discovered in 2004 in São Paulo state, Brazil (Teixeira *et al.*, 2005a), and, together with '*Ca.* L. africanus' and '*Ca.* L. asiaticus', is one of the three liberibacters that cause huanglongbing (HLB) disease of citrus worldwide (Bové, 2006). In São Paulo state, '*Ca.* L. americanus' infects most HLB-affected citrus trees, '*Ca.* L. asiaticus' infecting the remaining trees. In Florida, where HLB appeared in 2005, only '*Ca.* L. asiaticus' is involved. HLB, an insect-vector transmitted infection, is destructive and endangers the very existence of commercial citrus farms. The liberibacters are endogenous alphaproteobacteria and are restricted to the phloem sieve tubes. They have never been obtained in culture, and liberibacter-infected plants, citrus or experimental periwinkle plants (*Catharanthus roseus*) (Garnier & Bové, 1983), are the only sources of liberibacter DNA. Therefore, because of the difficulties of getting pure liberibacter DNA, very few liberibacter genes have been characterized, the β operon gene cluster being one of the exceptions. In the early 1990s, part of the *rplKAJL–rpoBC* gene cluster or β operon was obtained as a 2.6 kbp DNA fragment (In-2.6)

Abbreviation: HLB, huanglongbing.

The GenBank/EMBL/DDBJ accession numbers for the β operon sequences of '*Ca.* L. americanus' São Paulo and '*Ca.* L. africanus' Brazil are respectively EF122254 and EU078703.

by random cloning of DNA from periwinkle plants infected with the Poona (India) strain of '*Ca.* L. asiaticus' (Villechanoux *et al.*, 1992, 1993). Next, DNA from periwinkle plants infected with the Nelspruit strain (South Africa) of '*Ca.* L. africanus' was amplified with a primer pair designed from In-2.6 and yielded, after cloning and sequencing, a 1.7 kbp DNA fragment (AS-1.7) of the β operon of '*Ca.* L. africanus' (Planet *et al.*, 1995). In-2.6 and AS-1.7 have been used as specific hybridization probes for the detection of the Asian and African liberibacter strains, respectively (Villechanoux *et al.*, 1992; Hocquellet *et al.*, 1997). They have served to design PCR primers *rplA*2 and *rplJ*5 for detection of liberibacters by amplification of ribosomal protein genes (Hocquellet *et al.*, 1999). Finally, comparison of the sequences of In-2.6 and AS-1.7 confirmed that the Asian and African liberibacters represent two distinct species (Planet *et al.*, 1995).

When '*Ca.* L. americanus' was described as a new liberibacter (Teixeira *et al.*, 2005a), its β operon sequence was not available and could not be used for sequence comparisons with the corresponding gene clusters of the other two liberibacters. Characterization of the American liberibacter was only based on the sequences of the 16S rRNA gene and the 16S–23S rRNA intergenic region. Here, we describe how the β operon of the American liberibacter, as well as three upstream genes (*tufB*, *secE*, *nusG*), have now been obtained by the technique of chromosome walking and characterized (GenBank accession no. EF122254). Furthermore, the β operon sequence was determined for a Brazilian strain of '*Ca.* L. asiaticus' (GenBank accession no. EU078703), the sequences of an Indian strain (M94319; Villechanoux *et al.*, 1992, 1993) and a Japanese strain (AY342001; Okuda *et al.*, 2005) being already known. Finally, the β operon sequence of '*Ca.* L. africanus', only partially known (GenBank accession no. U09675), was completed (EF122255), making it possible to compare the β operon sequences of African, Asian and American liberibacter strains over a length of ~3000 bp. The results of these analyses will be presented and used to discuss the phylogeny and speciation of the liberibacters.

METHODS

Plant material and DNA extraction. Leaves infected with '*Ca.* L. americanus' (São Paulo strain) or '*Ca.* L. asiaticus' (Brazil strain) were from sweet orange trees in São Paulo State, Brazil. Leaves from healthy periwinkle plants and from periwinkle plants infected with '*Ca.* L. africanus' (Nelspruit strain), '*Ca.* L. asiaticus' (Poona strain) or '*Ca.* L. americanus' (São Paulo strain), as well as healthy sweet orange leaves, were from the Bordeaux laboratory greenhouse. DNA extraction was from 500 mg leaf midribs, using the cetyl trimethyl ammonium bromide (CTAB) method (Murray & Thompson, 1980). The DNA concentration was adjusted to 500 ng μl^{-1}.

PCR amplification, cloning and sequencing. For amplification of a 224 bp DNA fragment of the *rplKAJL–rpoB* gene cluster of '*Ca.* L. americanus' (São Paulo strain), conventional PCR was carried out in a 40 µl reaction mixture containing 500 ng total DNA, 1× PCR buffer (Promega), 3 mM $MgCl_2$, 0.2 mM of each dNTP, 500 nM of each primer (KF1, 5′-CGCCGCCGGTTGGTCCTGC-3′; KR1, 5′-CTTACCAGGAAGTTTAGATCC-3′) and 1.25 U *Taq* DNA polymerase (Promega). The program consisted of 35 cycles of 95 °C for 45 s, 60 °C for 45 s, 72 °C for 45 s and a final extension of 72 °C for 5 min. The amplified fragment was ligated into the pGEM-T easy vector following the supplier's protocol (Promega). Two microlitres of the ligation mixture was then used to transform competent *Escherichia coli* DH5-α cells by electroporation (Dower *et al.*, 1988). The cloned DNA was sequenced using Genome Express facilities (https://www.gexbyweb.com/gexbyweb).

Chromosome walking. The 224 bp fragment amplified from the *rplKAJL* gene cluster of '*Ca.* L. americanus' was used to design primers GSP1 and GSP2, as well as GSP3inv and GSP4inv (Table 1, Fig. 1), for the chromosome walking method, according to the Universal Genome Walker kit (BD Biosciences). Four batches of total DNA extracted from periwinkle leaves infected with '*Ca.* L. americanus' (São Paulo strain) were each digested with one of the following restriction enzymes: *Sna*I, *Sna*BI, *Dra*I and *Eco*RV. For each batch of digested DNA, a genomic library was obtained by ligation to the genome walker adaptors. The sequence at the 5′ end of the 224 bp fragment (sequence #1) was obtained with the *Sna*I library. A first PCR amplification with primer AP1 (on the adaptor) and GSP3inv (on the 224 bp fragment), followed by a nested PCR with primer AP2 (on the adaptor) and GSP4inv (on the 224 bp fragment), gave a fragment of about 1600 bp. The sequence at the 3′ end was obtained in two stages. In the first stage, sequence #2, immediately following the 224 bp fragment, was obtained with the *Sna*BI library. A first PCR amplification using primers GSP1 (on the 224 bp fragment) and AP1 (on the adaptor) followed by a second, nested PCR with primers

Table 1. Sequences of the primers used for chromosome walking

Primer	Location	Sequence (5′–3′)
AP1	Adaptor	GTAATACGACTCACTATAGGGC
AP2	Adaptor	ACTATAGGGCAGGCGTGGT
GSP3inv	224 bp rpl fragment	GGAAAAAACTGACTGGAGGCTGACTCA
GSP4inv	224 bp rpl fragment	ATCACTGTAGTTGGAACAGGAATACCC
GSP1	224 bp rpl fragment	TGGTCCTGCTATTGGTCAAACTGGTGT
GSP2	224 bp rpl fragment	GGGTATTCCTGTTCCAACTACAGTGAT
GSP5d	Sequence #2 1700 bp	TAGCCTCTCTTCCGAGCATTGATGTAC
GSP6d	Sequence #2 1700 bp	CTTAGAACTCTTAATGCGCCTCATGCT
KF1	224 bp rpl fragment	CGCCGCCGGTTGGTCCTGC
KR1	224 bp rpl fragment	CTTACCAGGAAGTTTAGATCC

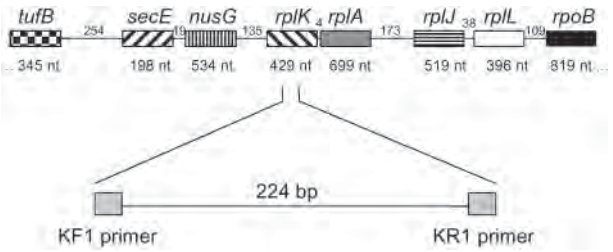


Fig. 1. Localization of the 224 bp fragment corresponding to the sequence of the *rpl* operon for '*Ca.* L. americanus'.

GSP2 (on the 224 bp fragment) and AP2 (on the adaptor) gave a 1700 bp product, which was sequenced. In the second stage, sequence #3 was obtained with the *Eco*RV library. A PCR amplification with primer GSP5d (on the 1700 bp sequence #2 from the first stage) and AP1 (on the adaptor), followed by nested PCR with primers GSP6d (on the 1700 bp sequence # 2) and AP2 (in the adaptor), gave a 1400 bp fragment. PCR conditions were as follows. The first amplification consisted of 7 cycles of 25 s at 94 °C and 3 min at 72 °C, followed by 32 cycles of 25 s at 94 °C and 3 min at 67 °C and a final extension of 7 min at 67 °C. The second (nested) amplification consisted of 5 cycles of 30 s at 94 °C and 3 min at 67 °C followed by 20 cycles of 30 s at 94 °C and 3 min at 62 °C and a final extension of 7 min at 62 °C. Products from the nested PCRs were purified on columns (Qiagen), cloned in the pGEM-T easy vector and sequenced. The sequence of each primer is shown in Table 1.

Sequence analyses. DNA and deduced protein sequence analyses were carried out with several programs. Searches for identities and similarities were by LALIGN (http://bioinfo.hku.hk/FASTA/lalign.htm). Multiple alignments of nucleotide sequences were performed with CLUSTAL W (http://www.ebi.ac.uk/clustalw; Thompson *et al.*, 1994). The *nusG–rplKAJL–rpoB* sequences from the three liberibacters '*Ca.* L. americanus' São Paulo strain (GenBank accession no. EF122254), '*Ca.* L. asiaticus' Poona strain (M94319) and '*Ca.* L. africanus' Nelspruit strain (U09675, EF122255) were compared using the FASTA program. Nucleotide sequences, as well as open reading frames (ORFs) as deduced from the nucleotide sequences (http://www.expasy.ch/tools/dna.html), were compared to determine percentages of identity and similarity (LALIGN).

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4 (Tamura *et al.*, 2007). Deduced amino acid sequences corresponding to the product of *rplJ* used were those of '*Ca.* L. asiaticus' (GenBank accession nos AAZ81564, ABI35994, 2002224D, AAP22421 and AAR13469), '*Ca.* L. africanus' (P41191), '*Ca.* L. africanus subsp. capensis' (Q9L5W5), *Bartonella henseleae* Houston-1^{T} (YP_033437), *Mesorhizobium* sp. BNC1 (YP_674379), *Brucella suis* 1330^{T} (AAN30165), *Rhizobium etli* CFN 42^{T} (YP_469187), *Sinorhizobium meliloti* 1021 (NP_385452), *Agrobacterium tumefaciens* C58 (NP_354933) and *E. coli* K-12 W3110 (P0A7J3). Nucleotide sequences corresponding to the *nusG–rplKAJL–rpoB* genes used were those from '*Ca.* L. asiaticus' Poona strain (GenBank accession no. M94319), '*Ca.* L. asiaticus' Japan strain (AY342001), '*Ca.* L. americanus' São Paulo strain (EF122254) '*Ca.* L. asiaticus' Brazil strain (EU078703) and '*Ca.* L. africanus' (U09675, EF122255).

For gene organization analysis, the STRING database was used (http://string.embl.de/; von Mering *et al.*, 2007).

Dot-blot hybridization. Total DNA was extracted from periwinkle or sweet orange leaf midribs collected on healthy plants or plants infected with one of the three liberibacters. The DNA was dot-blotted onto nylon N+ membranes according to Villechanoux *et al.* (1992). Two probes were used. One corresponded to a PCR product of 878 bp from '*Ca.* L. americanus' São Paulo strain, comprising the 3′ half of *rplA*, the *rplA–rplJ* intergenic region and most of *rplJ* (Fig. 2). Primers used for the amplification were *rplA*-Am-PrF1 (5′-TAGCAACTCCTGATATGATGCC) and *rplJ*-Am-PrR1 (5′-GGACAAGGGGATATTGGATAATG). The second probe was the 1027 bp amplicon amplified from the 16S rRNA gene of '*Ca.* L. americanus' with primers GB1 and GB3 (Teixeira *et al.*, 2005b). This 16S rRNA gene probe is not specific to liberibacter and hybridized with bacterial and plant ribosomal genes. The probes were labelled by random priming with the Promega kit in the presence of 1 μl [α-^{32}P]dATP. The probes were purified on Promega columns before use. Hybridization was according to Villechanoux *et al.* (1992).

Estimation of evolutionary divergence and speciation dating. The number of base substitutions per site was determined by pairwise analysis of the three liberibacter 16S rRNA gene sequences. Other 16S rRNA gene sequences used for comparison were from *Salmonella enterica* subsp. *enterica* serovar Paratyphi A ATCC 9150 (GenBank accession no. CP000026), *Escherichia coli* (unknown strain) (V00348), *Bartonella henselae* 882_ANT5 (AY513504), *Brucella abortus* 11-19 (X13695), *Mesorhizobium loti* MAFF303099 (BA000012), *Afipia felis* 2002033830 (AY513503), *Xanthomonas campestris* pv. *vesicatoria* 85-10 (AM039952), *Xylella fastidiosa* 9a5c (AE003849), *Mycoplasma pneumoniae* M129 (U00089) and *Mycoplasma genitalium* G-37^{T} (X77334). Analyses were conducted using the maximum composite likelihood method in MEGA4 (Tamura *et al.*, 2004, 2007). All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 1264 positions in the final dataset. The phylogenetic tree was inferred using the neighbour-joining method, and a bootstrap test of 500 replicates was performed. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages. To estimate the time of divergence between liberibacters, the molecular clock was calibrated using the time of divergence between the genera *Escherichia* and *Salmonella*. This time was estimated as 102 million years (Myr) by Battistuzzi *et al.* (2004), in agreement with previous work by others (Ochman & Wilson, 1987; Doolittle *et al.*, 1996).

RESULTS AND DISCUSSION

Sequence of the *tufB*, *secE*, *nusG*, *rplKAJL–rpoBC* gene cluster of '*Ca.* L. americanus'

To start chromosome walking, a short sequence of the *rplKAJL–rpoBC* gene cluster of '*Ca.* L. americanus' was required. For that purpose, the *β* operon sequences of '*Ca.* L. africanus' strain Nelspruit (GenBank accession no. U09675; partial) and '*Ca.* L. asiaticus' strain Poona

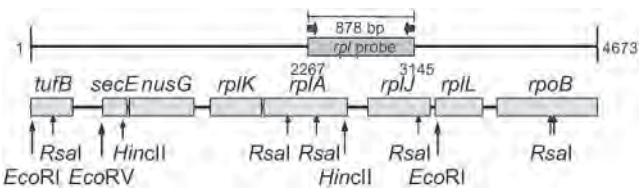


Fig. 2. Position of the rpl probe from '*Ca.* L. americanus' used in the hybridization experiments.

(M94319) were aligned and 16 forward and reverse primers for conventional PCR amplification were designed from conserved regions. Only one primer pair, KF1/KR1 in the *rplK* gene (Fig. 1), gave an amplicon, of 224 bp, with DNA isolated from sweet orange leaves infected with '*Ca.* L. americanus' São Paulo strain. After cloning and sequencing, this fragment turned out to be, indeed, a fragment of a liberibacter *rplK* gene, on the basis of BLAST scores, but, as expected, it did not correspond to the *rplK* gene of '*Ca.* L. africanus' or '*Ca.* L. asiaticus'.

Next, the sequences to the left (5′ end) and right (3′ end) of the 224 bp fragment could be determined by chromosome walking, as described in Methods. The 5′ end sequence was determined in one step, the ~1600 bp sequence being obtained with the *Sna*I library, and confirmed with a 1100 bp stretch from a *Dra*I library. Two steps were necessary for the 3′ end sequence. The first step involved the *Sna*I library and resulted in a ~1700 bp sequence, which was confirmed with a 750 bp sequence from the *Eco*RV library. The second step led to a 1400 bp sequence using the *Eco*RV library, and was confirmed by a ~1100 bp sequence obtained with the *Sna*I library. The combined sequences reached 4673 bp and comprised several ORFs, as determined by Frame D (Schiex *et al.*, 2003), and identified as *tufB*, *secE*, *nusG*, *rplKAJL* and *rpoB* (Fig. 1). The gene order *secE* to *rpoB*, as determined for '*Ca.* L. americanus', is characteristic of many alpha and gammaproteobacteria, including *E. coli*, as well as non-proteobacteria, such as certain members of '*Candidatus* Phytoplasma' (von Mering *et al.*, 2007). It can be assumed that the two upstream genes, *tufB* and *secE*, which have been identified in '*Ca.* L. americanus' (GenBank accession no. EF122254) and '*Ca.* L. asiaticus' (AY342001) also occur in '*Ca.* L. africanus'.

The 4673 bp gene cluster, when aligned with the corresponding sequences of '*Ca.* L. asiaticus' strain Poona and '*Ca.* L. africanus' strain Nelspruit, shared only ~72 % identity. Proof that the 4673 bp gene cluster was that of '*Ca.* L. americanus' was obtained by (i) DNA hybridization and (ii) PCR experiments as follows. (i) An 878 bp DNA fragment was amplified from the *rplA–rplJ* gene region of the 4673 bp sequence (Fig. 2), radioactively labelled with ^{32}P and used as an 'rpl' probe for dot-blot hybridization. Fig. 3 shows that the 'rpl' probe gave positive hybridizations only with DNA from citrus and periwinkle plants infected with '*Ca.* L. americanus' and not with DNA from healthy plants or plants infected with '*Ca.* L. asiaticus' or '*Ca.* L. africanus'. In contrast to the 'rpl' probe, a 16S rRNA gene probe hybridized, as expected, with DNA from healthy plants or plants infected with any of the three liberibacters. (ii) PCR experiments also proved that the 4673 bp gene cluster originated from the American liberibacter. In these experiments (Teixeira *et al.*, 2008), PCR primers f-*rplJ*Am (5′-GGACAAGGGGATATTGGA-TAATGATG-3′) and r-*rplJ*Am (5′-ATTAAGAGTTCT-AAGCAACCTGACAG-3′) were designed from the *rplJ* gene of the 4673 bp gene cluster and led to a 127 bp amplicon. They were tested in conventional and real-time PCRs (Teixeira *et al.*, 2008). No amplification was observed when the template was DNA from healthy citrus or periwinkle plants or from plants infected with '*Ca.* L. asiaticus' or '*Ca.* L. africanus'; amplification was observed only with DNA from '*Ca.* L. americanus'-infected citrus or periwinkle leaves. These results confirmed that the 4673 bp sequence was that of '*Ca.* L. americanus'.

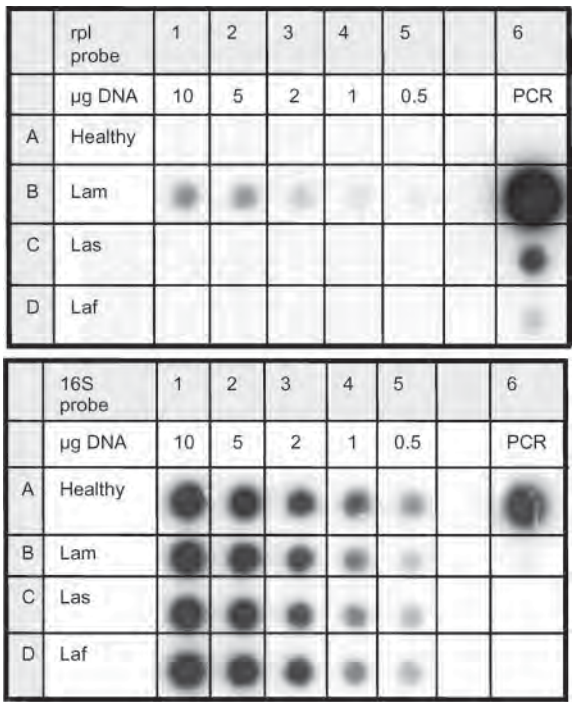


Fig. 3. Dot-blot hybridizations between ^{32}P-labelled probes and DNA from healthy periwinkles or periwinkles infected with either '*Ca.* L. asiaticus', '*Ca.* L. africanus' or '*Ca.* L. americanus' (Las, Laf, Lam). Upper panel: hybridization with the rpl probe corresponding to the *rpl* region of '*Ca.* L. americanus' (see Fig. 2). Lower panel: hybridization with the probe corresponding to the 16S rRNA gene of '*Ca.* L. americanus'. The amounts (in µg) of periwinkle DNA deposited on the membranes are indicated. PCR: diluted PCR products corresponding to the rpl probe (upper) or 16S rRNA gene probe (lower), used as positive controls. Aliquots containing 1.0, 0.1 and 0.01 ng PCR product were deposited.

Sequence of the *nusG–rplKAJL–rpoB* gene cluster of a Brazilian strain of '*Ca.* L. asiaticus' and completion of the corresponding sequence of '*Ca.* L. africanus'

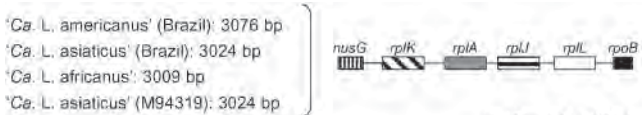
In São Paulo state, Brazil, citrus HLB is caused by one of two liberibacters, '*Ca.* L. americanus' and '*Ca.* L. asiaticus'. The β operon sequence of '*Ca.* L. asiaticus' is known for an Indian strain (Poona strain) (Villechanoux *et al.*, 1993) and a Japanese strain (GenBank accession no. AY342001). The sequence from a Brazilian strain of the Asian liberibacter was thought to be useful for comparison with the Brazilian

'*Ca.* L. americanus' São Paulo strain. Since the sequence of the *nusG–rplKAJL–rpoB* gene cluster was already known for two strains of '*Ca.* L. asiaticus', the sequence for the Brazilian strain (Brazil strain) was obtained by PCR amplification with a series of primers defined from the *nusG*, *rplA*, *rplJ* and *rpoB* sequences of these two '*Ca.* L. asiaticus' strains (GenBank accession nos M94319 and AY342001). In this way, a sequence of 3025 bp was obtained, and it extended from the 3′ end of *nusG* to the 5′ end of *rpoB* (Fig. 4). Similarly, the gene cluster of '*Ca.* L. africanus' strain Nelspruit was increased from a 1673 bp sequence (*rplA* to *rpoB*) to a 3013 bp sequence extending from the 3′ end of *nusG* to the middle of *rpoB* (Fig. 4).

Fig. 4 represents the gene organization of the *nusG–rplKAJL–rpoB* gene clusters, and the percentage of identity, between '*Ca.* L. americanus' (São Paulo strain), '*Ca.* L. asiaticus' (Brazil strain and Poona strain) and '*Ca.* L. africanus' (Nelspruit strain). The gene clusters of the American liberibacter strain and the Asian liberibacter strain from Japan have two upstream genes, *tufB* and *secE*, which are probably also present in the African liberibacter gene cluster but have not yet been investigated. This explains why the '*Ca.* L. americanus' gene cluster has a size of 4673 bp, compared with 3025 bp for '*Ca.* L. asiaticus' strain Brazil and 3013 bp for '*Ca.* L. africanus' strain Nelspruit.

Comparison of liberibacter *nusG* and *rplKAJL–rpoB* nucleotide sequences

When *β* operon sequences of '*Ca.* L. asiaticus' strain Poona and '*Ca.* L. africanus' strain Nelspruit were compared for the first time (Planet *et al.*, 1995), the comparison involved ~1700 bp, since the available sequence for the African liberibacter reached only 1676 bp, from the 3′ half of *rplA* to the 5′ half of *rpoB*. Now, after completion of the *β* operon sequence of '*Ca.* L. africanus' Nelspruit strain, the comparison could be carried out over ~3000 bp, from the 3′ half of *nusG* to the 5′ half of *rpoB*, and the percentage of identity for '*Ca.* L. asiaticus' Poona strain and '*Ca.* L. africanus' Nelspruit strain was found to be 81.2 % (Fig. 4). When the corresponding sequence from '*Ca.* L. americanus' São Paulo strain was compared with those of '*Ca.* L. asiaticus' Poona strain and '*Ca.* L. africanus' Nelspruit strain, the percentages of identity were respectively 71.3 and 72.1 %. A very similar value, 71.2 %, was found when the comparison with '*Ca.* L. americanus' São Paulo strain involved the Brazil strain of '*Ca.* L. asiaticus'. This is not surprising, since the identity between the Indian strain (Poona) and the Brazilian strain (Brazil) of '*Ca.* L. asiaticus' is 99.8 % (Fig. 4). Very similar identity values were found when the comparisons involved the whole gene cluster (*nusG–rplKAJL–rpoB*) or only *rplKAJL*.



	nusG–rpoB	*rplKAJL*
'*Ca.* L. americanus' (São Paulo) vs '*Ca.* L. asiaticus' (Brazil)	71.2 %	70.7 %
'*Ca.* L. americanus' (São Paulo) vs '*Ca.* L. asiaticus' (Poona)	71.3 %	71.4 %
'*Ca.* L. americanus' (São Paulo) vs '*Ca.* L. africanus' (Nelspruit)	72.1 %	71.8 %
'*Ca.* L. asiaticus' (Brazil) vs '*Ca.* L. asiaticus' (Poona)	99.8 %	99.8 %
'*Ca.* L. asiaticus' (Brazil) vs '*Ca.* L. africanus' (Nelspruit)	81.1 %	80.6 %
'*Ca.* L. asiaticus' (Poona) vs '*Ca.* L. asiaticus' (Nelspruit)	81.2 %	81.0 %

Fig. 4. Identity (%) between the *nusG–rplKAJL–rpoB* sequences of the three liberibacters.

It is generally assumed that two members of the same bacterial genus belong to different species when the percentage of identity of their total DNA is below 70 % (Johnson, 1984). In the comparisons of Fig. 4, values somewhat above 70 % (~72 %) were obtained when the American liberibacter was compared with either the Asian or African liberibacters. Since, however, the *rplKAJL* genes code for highly conserved proteins (ribosomal proteins), it is likely that the percentage of identity of total DNA of '*Ca.* L. americanus' to total DNA of the African or Asian liberibacters will be below 70 %, indicating that the American liberibacter strain belongs to a species different from '*Ca.* L. africanus' and '*Ca.* L. asiaticus'. When the latter two are compared, the identity is ~80 %, higher than the 70 % identity found for '*Ca.* L. americanus' and '*Ca.* L. asiaticus' or '*Ca.* L. africanus', indicating that '*Ca.* L. americanus' is more distantly related to '*Ca.* L. asiaticus' or '*Ca.* L. africanus' than '*Ca.* L. asiaticus' is to '*Ca.* L. africanus'.

Additional differences concern the G+C content of the *β* operon sequences of the three liberibacters. At 35.4 mol%, '*Ca.* L. americanus' has the lowest G+C content. For '*Ca.* L. asiaticus' strain Poona and '*Ca.* L. africanus' strain Nelspruit, the G+C contents are respectively 39.7 and 37.7 mol%. In comparison, *Bartonella henselae* and *Agrobacterium tumefaciens*, alphaproteobacteria like the liberibacters, have higher G+C contents, respectively 40.40 and 59.45 mol%.

Proteins encoded by the liberibacter *β* operon genes: amino-acid sequence comparisons

Protein L10, encoded by *rplJ*, was found to start with a leucine (TTG), instead of a methionine (ATG), in all three liberibacters '*Ca.* L. africanus' (strain Nelspruit; GenBank accession no. EF122255), '*Ca.* L. asiaticus' [strains India (M94319), Japan (AY342001) and Brazil (EU078703)] and '*Ca.* L. americanus' (strain São Paulo; EF122254). In this cluster, a second gene of '*Ca.* L. americanus', encoded by *secE*, also starts with a leucine. In Fig. 5, the amino acid sequences of the *β* operon proteins of the three liberibacters were compared. The major differences concerned proteins L10, encoded by *rplJ*, and protein L12, encoded by *rplL*. For instance, protein L10 from '*Ca.* L. africanus' shared 80.1 % identity and 95.3 % similarity with protein L10 of '*Ca.* L. asiaticus'. The percentages were

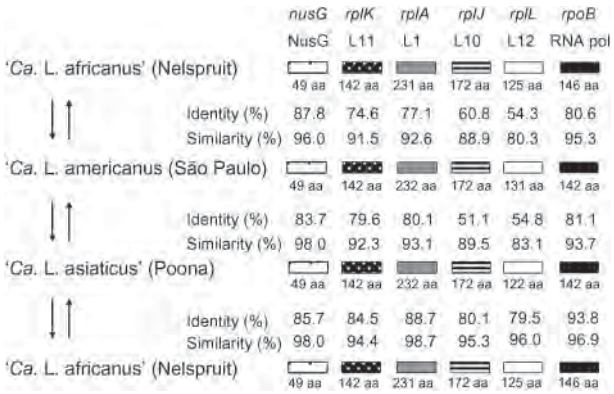


Fig. 5. Identity and similarity (%) among corresponding genes of the three liberibacters.

much smaller when the comparison involved protein L10 from either '*Ca.* L. americanus' and '*Ca.* L. asiaticus' or '*Ca.* L. americanus' and '*Ca.* L. africanus'. In the first case, the identity and similarity were 51.1 and 89.5 %, respectively; in the second case, they were 60.8 and 88.9 %. About the same identity and similarity were found when the comparisons involved protein L12. These results show that proteins L10 and L12 of '*Ca.* L. asiaticus' share less identity and similarity with the corresponding proteins of '*Ca.* L. americanus' than with those of '*Ca.* L. africanus'.

Phylogenetic analysis based on the *nusG–rplKAJL–rpoB* gene cluster

The topology of the phylogenetic tree inferred from the nucleotide sequence of the whole *nusG–rplKAJL–rpoB* gene cluster was found to be identical to the one based on the 16S rRNA gene (Teixeira *et al.*, 2005a) (Fig. 6). While the three liberibacters formed separate groups, '*Ca.* L. africanus' and '*Ca.* L. asiaticus' shared a common branch, but '*Ca.* L. americanus' did not, and clustered on a separate branch. Phylogenetic trees obtained with individual genes of the *nusG–rplKAJL–rpoB* cluster were identical (not shown) to the tree of Fig. 6. A phylogenetic tree based on

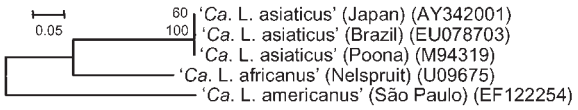


Fig. 6. Phylogenetic tree based on the nucleotide sequence of the *nusG–rplKAJL–rpoB* fragment.

the deduced amino acid sequence of *rplJ* was also constructed (Fig. 7). As observed with 16S rRNA gene sequences, *rplJ* sequences from liberibacters clustered together, well separated from the other alphaproteobacteria. Moreover, the topology of the tree within the liberibacter cluster was the same as observed with the 16S rRNA gene and the nucleotide sequence of the *nusG–rplKAJL–rpoB* cluster.

Liberibacter speciation dating: an estimation

With the aim of estimating the time of splitting between the three liberibacter lineages, evolutionary divergence between 16S rRNA gene sequences was calculated using the maximum composite likelihood method. Evolutionary divergence between '*Ca.* L. africanus' and '*Ca.* L. asiaticus' was estimated at 0.021 substitutions per position. In contrast, '*Ca.* L. americanus' showed calculated divergence of 0.042 and 0.045 with '*Ca.* L. africanus' and '*Ca.* L. asiaticus', respectively. As a way of comparing these estimations with data from well-known bacteria, the evolutionary divergence between the *Salmonella* and *Escherichia* lineages was calculated by the same method. A value of 0.014 substitutions per position was obtained, indicating that the evolutionary distances between the three liberibacters were significantly greater than that between the *Salmonella* and *Escherichia* lineages.

A phylogenetic tree was then inferred from the 16S rRNA gene sequences of the above-mentioned bacteria and others, using the neighbour-joining method. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages, and the clock to convert distance to time was calibrated using the time of divergence between

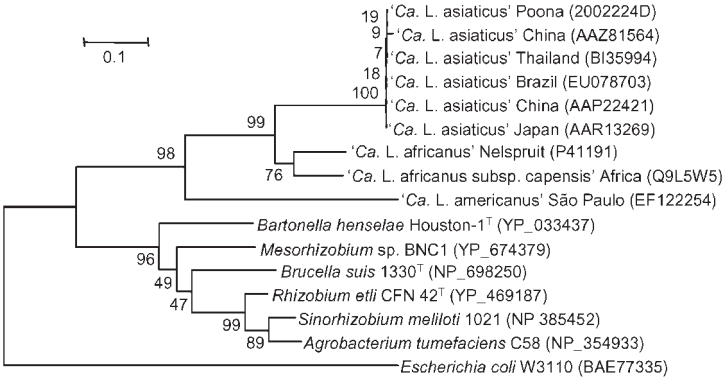


Fig. 7. Phylogenetic tree based on deduced amino acid sequences of *rplJ*.

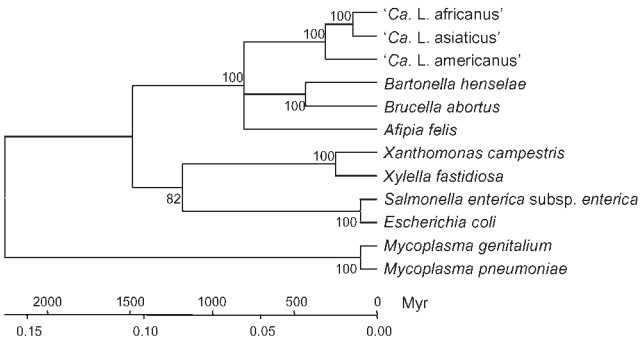


Fig. 8. Speciation dating of liberibacters. The phylogenetic tree was inferred from 16S rRNA gene sequences using the neighbour-joining method. Percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to branches. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages and the clock was calibrated using the estimated time of divergence between the *Salmonella* and *Escherichia* lineages (102 Myr). Evolutionary distances were computed using the maximum composite likelihood method and are expressed as the number of base substitutions per site. See text for strain details and accession numbers.

the *Salmonella* and *Escherichia* lineages as a reference. The value of 102 Myr was chosen according to Battistuzzi *et al.* (2004). This estimation was in accordance with other evaluations proposing that the two lineages had diverged 100–160 Myr ago (Ochman & Wilson, 1987; Doolittle *et al.*, 1996; Fig. 8). Using this parameter, the time of divergence between '*Ca.* L. africanus' and '*Ca.* L. asiaticus' was estimated at 147 Myr and the splitting between '*Ca.* L. americanus' and the asiaticus/africanus branch would have occurred 309 Myr ago.

It has been suggested that the liberibacters had Gondwanan origin (Beattie *et al.*, 2005). Speciation of the putative Gondwanan liberibacter ancestor could have occurred after dislocation and fractionation of the Gondwana supercontinent. Dislocation occurred about 160 Myr ago, when Africa split away from India, isolating the African liberibacter lineage within Africa and the Asian lineage within India, with subsequent eastward spread to China. As seen above, the African and Asian liberibacters are estimated to have diverged some 150 Myr ago. This time is in agreement with the period at which dislocation of Gondwana took place, some 160 Myr ago. The Gondwanan origin of the American liberibacter is, at the moment, more difficult to explain.

REFERENCES

Battistuzzi, F. U., Feijao, A. & Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* **9**, 44.

Beattie, G. A. C., Mabberley, D. J., Holford, P., Broadbent, P. & De Barro, P. (2005). Huanglongbing: its possible origins, collaborative research in Southeast Asia, and developing incursion management plans for Australia. In *Proceedings of the 2nd International Citrus Canker and Huanglongbing Research Workshop*, H-5. Orlando, FL: Florida Citrus Mutual.

Bové, J. M. (2006). Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *J Plant Pathol* **88**, 7–37.

Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470–477.

Dower, W. J., Miller, J. F. & Ragdsdale, C. W. (1988). High efficiency transformation of *Escherichia coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127–6145.

Garnier, M. & Bové, J. M. (1983). Transmission of the organism associated with citrus greening disease from sweet orange to periwinkle by dodder. *Phytopathology* **73**, 1358–1363.

Hocquellet, A., Bové, J. M. & Garnier, M. (1997). Production and evaluation of non-radioactive probes for the detection of the two ''*Candidatus* Liberobacter'' species associated with citrus Huanglongbing (greening). *Mol Cell Probes* **11**, 433–438.

Hocquellet, A., Bové, J. M. & Garnier, M. (1999). Isolation of DNA from uncultured ''*Candidatus* Liberobacter'' species associated with citrus Huanglongbing by RAPD. *Curr Microbiol* **38**, 176–182.

Johnson, J. L. (1984). Nucleic acids in bacterial classification. In *Bergey's Manual of Systematic Bacteriology*, vol. 1, pp. 8–11. Edited by N. R. Krieg & J. G. Holt. Baltimore: Williams & Wilkins.

Murray, M. G. & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**, 4321–4325.

Ochman, H. & Wilson, A. C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**, 74–86.

Okuda, M., Matsumoto, M., Tanaka, Y., Subandiyah, S. & Iwanami, T. (2005). Characterization of the *tufB-secE-nusG-rplKAJL-rpoB* gene cluster of the citrus greening organism and detection by loop-mediated isothermal amplification. *Plant Dis* **89**, 705–711.

Planet, P., Jagoueix, S., Bové, J. M. & Garnier, M. (1995). Detection and characterization of the African citrus greening Liberibacter by amplification, cloning and sequencing of the *rplKAJL-rpoB* operon. *Curr Microbiol* **30**, 137–141.

Schiex, T., Gouzy, J., Moisan, A. & de Oliveira, Y. (2003). FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res* **31**, 3738–3741.

Tamura, K., Nei, M. & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**, 11030–11035.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Teixeira, D. C., Ayres, A. J., Kitajima, E. W., Tanaka, F. A. O., Danet, J. L., Jagoueix-Eveillard, S., Saillard, C. & Bové, J. M. (2005a). First report of a huanglongbing-like disease of citrus in Sao Paulo State, Brazil, and association of a new liberibacter species, '*Candidatus* Liberibacter americanus', with the disease. *Plant Dis* **89**, 107.

Teixeira, D. C., Danet, J. L., Eveillard, S., Martins, E. C., de Jesus, W. C., Jr, Yamamoto, P. T., Lopes, S. A., Bassanezi, R. B., Ayres, A. J. & other authors (2005b). Citrus huanglongbing in São Paulo State, Brazil: PCR detection of the '*Candidatus* Liberibacter' species associated with the disease. *Mol Cell Probes* **19**, 173–179.

Teixeira, D. C., Saillard, C., Couture, C., Martins, E. C., Wulff, N. A., Eveillard-Jagoueix, S., Yamamoto, P. T., Ayres, A. J. & Bové, J. M. (2008). Distribution and quantification of *Candidatus* Liberibacter americanus, agent of huanglongbing disease of citrus in São Paulo

State, Brasil, in leaves of an affected sweet orange tree as determined by PCR. *Mol Cell Probes* **22**, 139–150.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Villechanoux, S., Garnier, M., Renaudin, J. & Bové, J. M. (1992). Detection of several strains of the bacterium-like organism of citrus greening disease by DNA probes. *Curr Microbiol* **24**, 89–95.

Villechanoux, S., Garnier, M., Laigret, F., Renaudin, J. & Bové, J. M. (1993). The genome of the non-cultured, bacterium-like organism associated with citrus greening disease contains the *nusG-rplKAJL-rpoB* gene cluster and the gene for a bacteriophage type DNA polymerase. *Curr Microbiol* **26**, 161–166.

von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T., Krüger, B., Snel, B. & Bork, P. (2007). STRING 7 – recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**, D358–D362.