

1 **Cryptic haplotypes of “*Candidatus Liberibacter africanus*”**

2

3 Warrick R Nelson, The New Zealand Institute for Plant & Food Research Limited, Private Bag 4704,
4 Christchurch 8140, New Zealand

5 Sandrine Eveillard, Marie-Pierre Dubrana and Joseph M Bové, Université Bordeaux and INRA, UMR
6 13 32 Biologie du fruit et Pathologie, 71, Ave. Edouard Bourlaux, 33883 Villenave d’Ornon, France

7 Dedicated to the memory of Colette Bové-Dumeau (1927-2014).

8

9 **Abstract**

10 “*Candidatus Liberibacter africanus*” (Laf) has long been recognised as a causal agent of the
11 devastating citrus disease huanglongbing (HLB) or citrus greening. This species is currently restricted
12 to Africa, the Arabian Peninsula and some Indian Ocean islands and vectored by the African citrus
13 psyllid, *Trioza erytreae*. Blotchy mottle on citrus leaves is characteristic of the disease.

14 Somewhat similar symptoms in the Rutaceous tree *Calodendrum capensis* (Cape Chestnut) resulted
15 in the discovery of Laf outside commercial citrus crops in South Africa. This was classed as a
16 subspecies of Laf (*capensis*, hence LafC). In subsequent surveys of both commercial citrus crops and
17 *Calodendrum*, both natural and ornamental specimens, LafC was not found in the citrus crop, nor
18 has Laf been found in *C. capensis*. HLB was reported from Madagascar in 1968 but no sequences
19 from this source have so far been published.

20 Until fairly recently, only the reference 16S rRNA gene sequences of Laf (L22533) and LafC
21 (AF137368) had been deposited in GenBank. Both of these reference sequences contain a number
22 of unresolved nucleotides. Resolving these nucleotide positions by aligning against more recently
23 available sequences, it becomes evident that these unresolved positions represent one percentage
24 point difference in similarity between Laf and LafC. The originally reported 97.4% similarity is
25 therefore incorrect based on this new information. Recalculating the similarity on the full length 16S
26 rDNA sequence results in 99.54% similarity, a value too high to justify a subspecies status. LafC
27 should therefore be reduced to that of a haplotype of Laf.

28 Further, the six 16S rRNA gene sequences currently available in GenBank identified as the species Laf
29 separate into 2 haplotype groups. The 3 haplotypes of Laf are therefore LafA designated as the first
30 accession sequenced (L22533), LafC for the former *capensis* subspecies and to recognise the prior
31 use of this term, and LafB for the third haplotype not previously recognised. Thus the cryptic
32 presence of 3 haplotypes is revealed by this review of the Laf 16S rDNA sequences.

33 **Keywords** “*Candidatus Liberibacter africanus*”; haplotype, citrus

34

35 **Introduction**

36 Huanglongbing (HLB), also known as citrus greening, is a serious disease threatening the economic
37 existence of many commercial citrus orchards. It has recently spread to the major production
38 regions of Brazil, Florida and California (Wang & Trivedi, 2013). Currently, 3 species of “*Candidatus*
39 *Liberibacter*”, named for the continent of first discovery, are associated with symptomatic
40 commercial citrus plants (Bové, 2006), “*Candidatus Liberibacter asiaticus*” (Las) being the most
41 geographically widespread (Bové, 2014).

42 “*Candidatus Liberibacter africanus*” (Laf) is associated with HLB throughout East Africa, the Arabian
43 Peninsula and Indian Ocean islands (Bové, 2006; Bové, 2014), including Madagascar (Bové & Cassin,
44 1968). The Rutaceous tree, *Calodendrum capense* (Cape Chestnut) has also been discovered to be
45 infected by “*Ca. Liberibacter africanus subsp. capensis*” (LafC)(Garnier *et al.*, 2000). This has
46 subsequently been found to be widespread in South Africa where this tree is both native and grown
47 ornamentally (Phahladira *et al.*, 2012). In spite of the subspecies being widespread, only Laf is
48 present in symptomatic commercial citrus orchards in South Africa (Pietersen *et al.*, 2010). Other
49 native Rutaceae species in South Africa also tested positive for Laf and LafC and possibly another
50 variant of the species, but interpreted as a series of subspecies of Laf (Roberts *et al.*, 2014).

51 The original description of LafC was based on three gene fragments (Garnier *et al.*, 2000), of which
52 the 16S rRNA gene is taxonomically the most important (Weisburg *et al.*, 1991; Woese & Fox, 1977).
53 This 16S sequence (GenBank accession AF137368) was noted as having 97.4% similarity with Laf
54 (L22533). Three percent divergence over the full gene sequence is a commonly recognised
55 requirement for bacterial species designation (Stackebrandt & Goebel, 1994), thus the subspecies
56 status for LafC. Closer inspection of these sequences revealed 6 unresolved nucleotides on LafC and
57 5 in the Laf reference sequences. There is thus potentially one percentage point of difference
58 derived simply from these unresolved nucleotide positions. A review of this subspecies status is now
59 feasible with the availability of more recent accession sequences.

60

61 **Methods and Materials**

62 A literature review revealed further examples of both Laf and LafC 16S rDNA sequences, as well as
63 some sequences for 16S-23S ISR, 23S and 50S rRNA gene regions. The megablast protocol in
64 GenBank was used to find further closely related sequences.

65 These sequences were downloaded from the NCBI database (GenBank) and aligned in ClustalX 2.1
66 (Larkin *et al.*, 2007). Sequences for 16S rDNA from the genomes of two other *Liberibacter* species
67 (“*Ca. L. asiaticus*” and “*Ca. L. solanacearum*”, Las and Lso respectively) were also downloaded for
68 additional comparison. These were chosen as phylogenies generally place them closest to Laf
69 (Nelson *et al.*, 2013).

70 DNA material stored from symptomatic leaves of sweet orange collected in Madagascar during 2011,
71 previously reported to be Laf by PCR (Bové, 2014), was retested using the primer pair OA1/Oi2c
72 (Jagouiex *et al.*, 1996) and the resultant 16S rDNA samples sequenced. The rplJ component of the
73 50S rRNA gene was also sequenced.

74

75 **Results**

76 On the 16S rRNA gene, 15 sequences identified as Laf or LafC were found. Three of these proved to
77 contain only a very short component of the 16S gene and were discarded. Aligning the remaining 12
78 sequences showed a number of apparently random SNPs scattered across the accessions,
79 presumably arising as sequencing errors, but the LafC sequences separated clearly.

80 In the sequence L22533 (Laf), there are 5 nucleotide positions labelled “N”. From within the aligned
81 set of sequences, these can be unambiguously resolved to the nucleotides T, C, G, A and G.
82 Similarly, in the sequence AF137368 (LafC), these also resolve unambiguously to A, A, A, G, C, C.

83 Laf sequences also aligned into 2 groups over 7 SNPs (Table 1), here designated as haplotypes LafA
84 and LafB. Excluding the occasional single sequence SNPs, the number of SNPs separating the 2
85 haplotypes and subspecies is outlined in Table 2. Although the sequence of each accession varied in
86 length, these SNPs occur over an approximately 1100 bp length. Aligning the Laf sequences against
87 the full 16S gene sequences for Las and Lso indicated that all the SNPs in Table 1 occur within the
88 overlapping regions. *Liberibacter* 16S genes are approximately 1500 bp (Nelson *et al.*, 2015). Until
89 full Laf sequences are available, the similarities reported here might prove to be high if further SNPs
90 exist in the currently unknown regions. There are 8 SNPs between Las and Lso in these end regions
91 of the gene, representing 0.53% difference. Assuming a similar number of SNPs between the Laf
92 sequences in these end regions, we can calculate a likely worst-case scenario for the number of SNPs
93 between the proposed haplotypes and the percentage similarity, assuming a 16S rRNA gene length
94 of 1500 bp.

95 Only 4 sequences across the 16S-23S ISR and 23S partial gene regions were found in GenBank
96 (LAU61360, EU754741, FJ914622, JF819884, JF819885) for Laf and none for LafC. The LafC sequence
97 has been mentioned as being closer to Las than Laf, because only one tRNA (alanine) is present in
98 the ISR of Laf, while two tRNAs (alanine and isoleucine) occur in the ISR of Las and LafC (Garnier *et*
99 *al.*, 2000).

100 Over a common segment of 479 bp on the 50S gene region (β -operon) there are 68 SNPs between
101 Laf and LafC (14 and 4 accessions respectively)(Phahladira *et al.*, 2012; Pietersen *et al.*, 2010). The
102 available sequence material in the 50S gene confirms the basic Laf/LafC separation, but we feel that
103 separating LafA and LafB is as yet inconclusive with only 2 SNPs across the 4 accessions.

104 The 16S sequences derived from Madagascar material aligned with LafA and the rplJ sequence
105 (GenBank) aligned with EF122255, a sequence with metadata indicating it is derived from
106 commercial citrus material in South Africa.

107

108 **Table 1 Proposed haplotypes and SNP differences on the 16S rRNA gene within “*Candidatus Liberibacter***
 109 ***africanus*” (Laf) sequences. The nucleotide numbers count from the beginning of the reference sequence**
 110 **L22533 although the alignment was done using a 1057 bp length common to all sequences. Shaded SNPs**
 111 **are positions of ambiguity between the sequences within a series where the SNP is in the same position**
 112 **as the nucleotide in one of the other haplotypes.**

GenBank	Source	Author	Haplotype	SNP description													
				g.149G>	g.329T>C	g.355G>A	g.411G>A	g.413G>A	g.427insA	g.486insG	g.507C>T	g.562A>G	g.565G>A	g.908C>T	g.920T>G	g.921T>A	g.1124insG
L22533	Nelspruit	(Jagoueix <i>et al.</i> , 1994)	Laf A	G	G	G	G	G	-	-	C	A	G	C	T	T	-
EU921620	Mpumalang a-UPCRI-06-0026	(Lin <i>et al.</i> , 2009)	Laf A	G	T	G	G	G	-	-	C	A	G	C	T	T	-
EU921619	Mpumalang a-UPCRI-05-0252	(Lin <i>et al.</i> , 2009)	Laf A	G	T	G	G	G	-	-	C	A	G	C	T	T	-
KJ1152137	UPCRI 12-0004 (Zanthoxylum capensis), Knysna	(Roberts <i>et al.</i> , 2014)	Laf A	-	T	G	G	G	-	-	C	G	G	C	G	A	-
LN795909	Madagascar	This study	Laf A	G	T	G	G	G	-	-	C	A	G	C	T	T	-
EU921621	Mpumalang a-UPCRI-06-0071	(Lin <i>et al.</i> , 2009)	Laf B	G	C	A	A	A	A	G	C	A	G	C	T	T	G
EU754741	Mpumalang a-UPCRI-06-0071	(Doddapane ni <i>et al.</i> , 2008)	Laf B	G	C	A	A	A	A	G	C	A	A	C	A	T	G
FJ914622	Mpumalang a-UPCRI-05-0232	(Lin <i>et al.</i> , 2009)	Laf B	G	C	A	A	A	A	G	C	A	G	C	T	T	G
AF137368	LafC (from <i>Calodendrum</i> , Western Cape province)	(Garnier <i>et al.</i> , 2000)	Laf C	-	T	G	G	G	-	-	T	G	A	T	G	A	-

KJ152127	UPCRI 10-2267 (Calodendrum capense), Good Hope	(Roberts <i>et al.</i> , 2014)	Laf C	-	T	G	G	G	-	-	T	G	A	T	G	A	-
KJ152131	UPCRI 11-4505 (Clauseniana anisata), Enseleni	(Roberts <i>et al.</i> , 2014)	Laf C	-	T	G	G	G	-	-	T	G	A	T	G	A	-
KJ197226	UPCRI 11-4270 (Zanthoxylum capense), St Francis Bay	(Roberts <i>et al.</i> , 2014)	Laf C	-	T	G	G	G	-	-	T	G	A	T	G	A	-
KJ152134	UPCRI 11-4044 (Vepris lanceolata), Knysna	(Roberts <i>et al.</i> , 2014)	Laf C	G	T	G	G	G	-	-	T	G	A	T	G	A	-

113

114 **Table 2 Percent identity similarity between “*Candidatus Liberibacter africanus*” (Laf) haplotypes based**
 115 **on 16S rRNA gene sequences, % similarity calculated on a 1500 bp length. An alternate % similarity**
 116 **figure is indicated in parentheses assuming a further 8 SNPs exists in the currently unknown segments**
 117 **of the gene.**

Comparison identity	SNPs	% similarity
LafA:LafB	7 (15)	99.53 (99.0)
LafA:LafC	7 (15)	99.53 (99.0)
LafB:LafC	14 (22)	99.07 (98.54)

118

119 Discussion

120 The unresolved nucleotides in the sequences for Laf and LafC (11 positions in total) are readily
 121 resolved to their correct nucleotide by comparison against the other 10 sequences now available in
 122 GenBank. These newly resolved positions represent one percentage point of similarity between Laf
 123 and LafC, thus the originally described 97.4% (Garnier *et al.*, 2000) similarity becomes 98.4%.
 124 However, this similarity calculation was conducted on the known sequence proportion of the gene
 125 since at the time the full length of the gene was not known. Although the full 16S gene sequence for
 126 Laf is still not available, it is likely to be approximately 1500 bp in length as found in 4 other species
 127 of *Liberibacter* (Nelson *et al.*, 2015). Across the sequences currently available, only approximately 40
 128 bases at the beginning and 30 at the end of the gene sequence remain unknown, thus allowing a
 129 closer estimate of similarity between the species and the subspecies (Table 2).

130 Variation across the 16S rRNA gene appears to be quite large with 14 SNPs apparent across the
 131 Laf/LafC accessions available, 7 of them between LafA/LafB and a different 7 between LafA/LafC,
 132 with all 14 being different between LafB/LafC. For comparison, Las appears to be homogenous
 133 across accessions on the 16S rRNA gene (Moreno-Enriquez *et al.*, 2014; Nelson, 2012), while for Lso
 134 5 haplotypes have been described, with only 7 SNPs defining them on the 16S rRNA gene (Nelson *et al.*,
 135 2011; Nelson *et al.*, 2013; Teresani *et al.*, 2014).

136 The unexpected separation into 2 groups on the 16S rRNA gene within the currently known Laf
137 species suggests 2 haplotypes, here designated LafA and LafB (Table 2). No prior study has indicated
138 that there were genetic differences within Laf on commercial citrus. This suggests 3 haplotypes
139 within the Laf species, 2 primarily indicated by sequences derived from commercial citrus crops, and
140 the previous subspecies designation “capensis” can be resolved to a haplotype but retaining the
141 short form of LafC. Caution is required in this conclusion since the LafB designation is based on only
142 3 available sequences and 2 of these, although sequenced by different laboratories, are clearly
143 derived from the same field sample. The metadata associated with these sequences give little
144 indication of geographical or plant/insect host separation. For LafA/LafB they are primarily from
145 commercial citrus crops and LafC from native Rutaceous species. This provides a stark contrast to
146 the analogous situation in Lso where the haplotypes also express partial plant/insect host and
147 geographic separation (Nelson *et al.*, 2011; Nelson *et al.*, 2013; Teresani *et al.*, 2014). If the A/B
148 haplotypes can be confirmed, they indicate 2 separate pathosystem events from the respective
149 currently unknown African native plant hosts into commercial citrus, since *Citrus* is not native to
150 Africa (Beattie *et al.*, 2008), and LafC has yet to make this jump as it is not yet known from citrus
151 (Phahladira *et al.*, 2012).

152 The 50S rRNA gene sequences show clear separation between Laf and LafC as expected, with only a
153 suggestion on 4 of 15 Laf sequences showing SNPs. Caution is suggested in considering this indicative
154 of LafA/B haplotypes as these SNPs occur together and only 4 bases from the end of the sequence.
155 Further, there is no indication from the metadata to tie these 4 sequences together with the 3
156 suggested as LafB from the 16S gene.

157 Although previously known that Laf was the species responsible for HLB on Madagascar by both
158 epidemiological and PCR studies (Bové & Cassin, 1968; Bové, 2014), this study confirms not only the
159 species as Laf via both 16S and rplJ genes, but that it is also haplotype LafA. This very strongly
160 suggests an incursion event of LafA from Africa to Madagascar, rather than a Gondwanan origin, in
161 spite of the position of Madagascar in Gondwana between East Africa and India.

162 The very recent suggestion of a further three subspecies of Laf (Roberts *et al.*, 2014) can be resolved
163 within this proposal of haplotypes rather than subspecies within Laf by giving them a biotype
164 designation, recognising the current host plant differences. Laf subspecies *vepridis* is a biotype of
165 LafA, while Laf subspecies *zanthoxyli* and Laf subspecies *clausenae* are biotypes of LafC.

166

167 **Conclusion**

168 Instead of the species Laf and subspecies *capensis*, re-analysis of the phylogenetically important 16S
169 rRNA gene suggests the subspecies is not that far removed from the species and should be revised to
170 haplotype status. Further, existing sequences of Laf indicate that this species also comprises two
171 haplotypes. Therefore two haplotypes (LafA and LafB) are known symptomatically in commercial
172 citrus orchards while the third (LafC) is known only from very mild symptoms in native Rutaceous
173 plants but not (yet) in citrus. Thus the cryptic presence of three haplotypes is revealed by this review
174 of the Laf 16S rDNA sequences.

175

176 **References**

- 177 **Beattie, G., Holford, P., Mabberley, D., Haigh, A. & Broadbent, P. (2008).** On the origins of *Citrus*,
178 Huanglongbing, *Diaphorina citri* and *Trioza erytreae*. *International Research Conference on*
179 *Huanglongbing, Orlando, Florida* , 25-57.
- 180 **Bové, J. M. (2006).** Huanglongbing: a destructive, newly-emerging, century-old disease of citrus.
181 *Journal of Plant Pathology* **88**, 7-37.
- 182 **Bové, J. M. (2014).** Heat-tolerant Asian HLB meets Heat-sensitive African HLB on the Arabian
183 Peninsula. Why?. *Journal of Citrus Pathology* **1**, 1-78.
- 184 **Bové, J. M. & Cassin, J. P. (1968).** Problèmes de l'agrumiculture malgache; comperendu de mission.
185 *Document IRFA*: 52 p.
- 186 **Doddapaneni, H., Liao, H., Lin, H., Bai, X., Zhao, X., Civerolo, E., Irely, M., Coletta-Filho, H. &**
187 **Pietersen, G. (2008).** Comparative phylogenomics and multi-gene cluster analyses of the Citrus
188 Huanglongbing (HLB)-associated bacterium *Candidatus Liberibacter*. *BMC Res Notes* **1**, 72.
- 189 **Garnier, M., Jagoueix-Eveillard, S., Cronje, P., Le Roux, H. & Bové, J. (2000).** Genomic
190 characterization of a liberibacter present in an ornamental rutaceous tree, *Calodendrum capense*, in
191 the Western Cape province of South Africa. Proposal of "*Candidatus Liberibacter africanus* subsp.
192 *capensis*". *Int J Syst Evol Microbiol* **50**, 2119-2125.
- 193 **Jagoueix, S., Bové, J. M. & Garnier, M. (1994).** The phloem-limited bacterium of greening disease of
194 citrus is a member of the α -subdivision of the Proteobacteria. *Int J Syst Bacteriol* **44**, 379-386.
- 195 **Jagoueix, S., Bové, J. M. & Garnier, M. (1996).** PCR detection of the two "*Candidatus*" liberobacter
196 species associated with greening disease of citrus. *Mol Cell Probes* **10**, 43-50.
- 197 **Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H.,**
198 **Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007).** Clustal W and Clustal X version 2.0.
199 *Bioinformatics* **23**, 2947-2948.
- 200 **Lin, H., Doddapaneni, H., Munyaneza, J. E., Civerolo, E. L., Sengoda, V. G., Buchman, J. L. & Stenger,**
201 **D. C. (2009).** Molecular characterization and phylogenetic analysis of 16S rRNA from a new species
202 of "*Candidatus Liberibacter*" associated with Zebra chip disease of potato (*Solanum tuberosum* L.)
203 and the potato psyllid (*Bactericera cockerelli* Sulc). *Journal of Plant Pathology* **91**, 215-219.
- 204 **Moreno-Enríquez, A., Minero-García, Y., Ramírez-Prado, J. H., Loeza-Kuk, E., Uc-Varguez, A. &**
205 **Moreno-Valenzuela, O. A. (2014).** Comparative analysis of 16S ribosomal RNA of '*Candidatus*
206 *Liberibacter asiaticus*' associated with Huanglongbing disease of Persian lime and Mexican lime
207 reveals a major haplotype with worldwide distribution. *African Journal of Microbiology Research* **8**,
208 2861-2873.
- 209 **Nelson, W. (2012).** First record of *Trioza vitreoradiata* (Maskell)(Hemiptera: Triozidae) in citrus.
210 *Citrus Research & Technology* **33**, 35-38.

- 211 **Nelson, W. R., Fisher, T. W. & Munyaneza, J. E. (2011).** Haplotypes of "*Candidatus Liberibacter*
212 *solanacearum*" suggest long-standing separation. *European Journal of Plant Pathology* **130**, 5-12.
- 213 **Nelson, W. R., Munyaneza, J. E., McCue, K. F. & Bové, J. M. (2013).** The Pangaeian origin of
214 "*Candidatus Liberibacter*" species. *Journal of Plant Pathology* **95**, 455-461.
- 215 **Nelson, W. R., Wulff, N. A. & Bové, J. M. (2015).** Intragenomic homogeneity on *Liberibacter* 16S
216 rDNA confirms phylogeny and explains ecological strategy. *bioRxiv* doi:
217 <http://dx.doi.org/10.1101/016188>.
- 218 **Nelson, W., Sengoda, V., Alfaro-Fernandez, A., Font, M., Crosslin, J. & Munyaneza, J. (2013).** A new
219 haplotype of "*Candidatus Liberibacter solanacearum*" identified in the Mediterranean region.
220 *European Journal of Plant Pathology* **135**, 633-639.
- 221 **Phahladira, M., Viljoen, R. & Pietersen, G. (2012).** Widespread occurrence of "*Candidatus*
222 *liberibacter africanus* subspecies *capensis*" in *Calodendrum capense* in South Africa. *European*
223 *Journal of Plant Pathology* **134**, 39-47.
- 224 **Pietersen, G., Arrebola, E., Breytenbach, J., Korsten, L., le Roux, H., la Grange, H., Lopes, S., Meyer,**
225 **J., Pretorius, M. & other authors (2010).** A survey for '*Candidatus Liberibacter*' species in South
226 Africa confirms the presence of only '*Ca. L. africanus*' in commercial citrus. *Plant Dis* **94**, 244-249.
- 227 **Roberts, R., Steenkamp, E. T. & Pietersen, G. (2014).** Novel lineages of '*Candidatus Liberibacter*
228 *africanus*' associated with native Rutaceae hosts of *Trioza erythrae* in South Africa. *Int J Syst Evol*
229 *Microbiol* , doi:10.1099/ijs.0.069864-0.
- 230 **Stackebrandt, E. A. & Goebel, B. M. (1994).** Taxonomic note: a place for DNA-DNA reassociation
231 and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol*
232 **44**, 846-849.
- 233 **Teresani, G. R., Bertolini, E., Alfaro-Fernandez, A., Martínez, C., Tanaka, F. A. O., Kitajima, E.,**
234 **Rosello, M., Sanjuan, S., Ferrandiz, J. C. & other authors (2014).** Association of '*Candidatus*
235 *Liberibacter solanacearum*' with a vegetative disorder of celery in Spain and development of a real-
236 time PCR method for its detection. *Phytopathology* **104**, 804-811.
- 237 **Wang, N. & Trivedi, P. (2013).** Citrus Huanglongbing: a newly relevant disease presents
238 unprecedented challenges. *Phytopathology* **103**, 652-665.
- 239 **Weisburg, W., Barns, S., Pelletier, D. & Lane, D. (1991).** 16S ribosomal DNA amplification for
240 phylogenetic study. *J Bacteriol* **173**, 697-703.
- 241 **Woese, C. & Fox, G. (1977).** Phylogenetic structure of the prokaryotic domain: The primary
242 kingdoms. *Proceedings of the National Academy of Sciences* **74**, 5088-5090.
- 243