

1 **Revision of the ‘*Candidatus Phytoplasma*’ species description guidelines**

2

3 Assunta Bertaccini¹, Yaima Arocha-Rosete², Nicoletta Contaldo¹, Bojan Duduk³, Nicola Fiore⁴,
4 Helena Guglielmi Montano⁵, Michael Kube⁶, Chih-Horng Kuo⁷, Marta Martini⁸, Kenro Oshima⁹,
5 Fabio Quaglino¹⁰, Bernd Schneider¹¹, Wei Wei¹², Alan Zamorano⁴

6

7

8 ¹Department of Agricultural and Food Sciences, *Alma Mater Studiorum* – University of Bologna, Italy

9 ²Sporometrics Inc., Toronto, Ontario, Canada

10 ³Institute of Pesticides and Environmental Protection, Belgrade, Serbia

11 ⁴Faculty of Agricultural Sciences, Department of Plant Protection, University of Chile, Santiago, Chile

12 ⁵Department of Entomology and Plant Pathology, Federal Rural University of Rio de Janeiro, Brazil

13 ⁶Department of Integrative Infection Biology Crops-Livestock, University of Hohenheim, Stuttgart, Germany

14 ⁷Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

15 ⁸Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Italy

16 ⁹ Faculty of Bioscience and Applied Chemistry, Department of Clinical Plant Science, Hosei University, Japan

17 ¹⁰Department of Agricultural and Environmental Sciences – Production, Landscape, Agroenergy, University
18 of Milan, Milan, Italy

19 ¹¹Julius Kuehn-Institute, Federal Research Centre for Cultivated Plants, Institute for Plant Protection in Fruit
20 Crops and Viticulture, Dossenheim, Germany

21 ¹²Molecular Plant Pathology Laboratory, USDA/ARS, Beltsville, MD, USA

22

23 **Correspondence: Assunta Bertaccini, assunta.bertaccini@unibo.it**

24

25

26 **Abstract**

27 The ‘*Candidatus Phytoplasma*’ genus was proposed to accommodate cell wall-less bacteria which
28 are molecularly and biochemically incompletely characterized, and colonize plant phloem, and
29 insect vector tissues. This provisional classification is highly relevant due to its application in
30 epidemiological and ecological studies, mainly aimed at keeping the severe phytoplasma plant
31 diseases under control worldwide. Given the increasing discovery of molecular diversity within the
32 ‘*Ca. Phytoplasma*’ genus, the proposed guidelines were revised and clarified to accommodate those
33 ‘*Ca. Phytoplasma*’ species strains sharing >98.65% sequence identity of their full or nearly full 16S
34 rRNA gene sequences, obtained with at least 2-fold coverage of the sequence, compared with those

35 of the reference strain of such species. Strains sharing <98.65% of sequence identity with the
36 reference strain but >98.65% with other strain(s) within the same '*Ca. Phytoplasma*' species should
37 be considered related strains to that '*Ca. Phytoplasma*' species. The guidelines herein keep the
38 original published reference strains. However, to improve '*Ca. Phytoplasma*' species assignment,
39 complementary strains are suggested as an alternative to the reference strain. This will be
40 implemented when only a partial 16S rRNA gene and/or a few other genes have been sequenced,
41 or the strain is no longer available for further molecular characterization. Lists of '*Ca. Phytoplasma*'
42 species and alternative reference strains described are reported. For new '*Ca. Phytoplasma*' species
43 that will be assigned with identity $\geq 98.65\%$ of their 16S rRNA gene sequences, a threshold of 95%
44 genome-wide average nucleotide identity is suggested. When the whole genome sequences are
45 unavailable, two among conserved housekeeping genes could be used. The officially published
46 '*Candidatus Phytoplasma*' species are 49 enclosing '*Ca. P. cocostanzaniae*' and '*Ca. P. palmae*'
47 described in this manuscript.

48

49

50 **Key words**

51 Phytoplasmas, provisional classification, alternative reference strains, gene sequences

52

53

54 **Conflicts of interest**

55 The author(s) declare that there are no conflicts of interest.

56

57 **Funding information**

58 This work received no specific grant from any funding agency

59

60

61

62

63

64

65

66

67 **Introduction**

68 The genus '*Candidatus Phytoplasma*' [1] was introduced to classify non-helical, cell wall-less bacteria
69 inhabiting plant phloem and insect vector tissues. The taxon '*Ca. Phytoplasma*' is part of the class
70 *Mollicutes* and its members are associated with over a thousand plant diseases worldwide [2,3].
71 Phytoplasmas are not available as pure colonies [4] and several of them are not yet cultured, thus
72 the Koch's postulates to confirm their role as pathogens are yet not fulfilled. The limited knowledge
73 of their biological properties hindered their classification; therefore, the provisional designation of
74 this genus allowed studying their epidemiology and genetics, revealing in some cases, new
75 molecular features of these bacteria [5]. Considering the increasing number of taxa discovered and,
76 in some cases, their possible overlapping molecular traits, the definition of a '*Ca. Phytoplasma*'
77 species requires revision and clarification.

78

79 **'*Candidatus Phytoplasma*' species threshold for the 16S ribosomal RNA gene**

80 The 16S rRNA gene identity threshold of 97.5% for a new '*Candidatus Phytoplasma*' species
81 designation defined in 2004 [1] was revised by adding the sequence length required for such
82 assignment [6]. To simplify the '*Ca. Phytoplasma*' new species description three thresholds 97.50%,
83 98.00% and 98.65% were evaluated for the deposited 16 rRNA gene sequences in accordance with
84 what was reported for the walled bacteria [7]. From this comparison (Tables 1 and S1) it was
85 concluded that the higher threshold value (98.65%) would increase the number of '*Ca. Phytoplasma*'
86 species, leading to splitting some of the existing ones (*i.e.*, '*Ca. P. phoenicium*', '*Ca. P. pruni*'). In
87 addition, it will reduce the number of misclassified phytoplasma strains. This evaluation performed
88 on the available sequences, indicates that 13 '*Ca. Phytoplasma*' species show modifications in the
89 assigned number of strains, which should be further evaluated in order to reassign them to the
90 pertinent taxon following the revised guidelines. For the taxonomy of bacteria, the average
91 nucleotide identity (ANI) is a very robust tool to support genome comparisons [8,9]. For
92 differentiating two species, an ANI threshold range of 95–96% was proposed. This is consistent with
93 the threshold value of 98.65% based on the 16S rRNA gene sequence identity [6,7]. Therefore,
94 differentiation should be based on a fragment of about 1.5 kb of the 16S rRNA gene sequence
95 (approximately 95% of the entire gene without the primer sequences). This sequence should be
96 based on both strands obtained with Sanger sequencing method with at least a 2-fold coverage for
97 three independent biological samples or from three different locations where a single phytoplasma
98 infection was determined. Primer pairs that amplify the entire 16S rRNA gene include among others

99 P1/P7, P1/16S-SR or P1A/16S-SR [10]. Strains sharing >98.65% of sequence identity when compared
100 with the reference strain are considered members of the respective '*Ca. Phytoplasma*' species.
101 Strains showing identity <98.65% to the reference strain, but >98.65% with other strains of the same
102 '*Ca. Phytoplasma*' species should be considered as related to this '*Ca. Phytoplasma*' species. Since
103 for some '*Ca. Phytoplasma*' species only the incomplete 16S rRNA gene sequence of the originally
104 designated reference strains is available, appropriate complementary additional strains selected
105 among those having the longer 16S rRNA sequence and the larger number of housekeeping gene
106 sequences available [11] are suggested (Tables 2 and 3).

107

108 '***Candidatus Phytoplasma***' species updates

109 In the case of those 16S rDNA sequences having >98.65% of identity, the assignment of a new '*Ca.*
110 *Phytoplasma*' species will be based on a threshold of <95% genome-wide average nucleotide
111 identity (ANI) as suggested for bacteria [12,13]. When whole genome sequences are not available,
112 other conserved/housekeeping genes, showing low mutation rates should be used for genomic
113 comparison. The reference strain of a given '*Ca. Phytoplasma*' species originally published, has been
114 retained, and a set of conserved available genes have been selected from the literature with the
115 capacity of amplifying the largest number of '*Ca. Phytoplasma*' species described so far [11].
116 Reference and alternative reference strains suggested for each '*Ca. Phytoplasma*' species are
117 described in Tables 2 and 3. When the 16S rRNA gene sequence is not supporting the '*Ca.*
118 *Phytoplasma*' species differentiation, other conserved/housekeeping genes or whole genome
119 sequences are suggested to confirm or support the '*Ca. Phytoplasma*' species designation. The
120 conserved/housekeeping genes include *tufB*, *secY*, *secA*, *rplV-rpsC* and *groEL*. For such genes, a
121 threshold of 97.6% for *groEL*, 97.5% for *tuf* and *rp*, 95.7% for *secA* and 95.0% for *secY* genes should
122 be used to allow effective distinction among them (Tables 2, 3 and S2). To consolidate the validity
123 of '*Ca. Phytoplasma*' species having the 16S rRNA gene threshold >98.65% in comparison with other
124 species, at least two conserved genes selected among those having neutral selection should be
125 provided. The biological properties defined as reported [1] could assist in the definition of the '*Ca.*
126 *Phytoplasma*' species when two phytoplasmas share >98.65% of the 16S rRNA gene sequence
127 identity and can be differentiated based on other conserved gene sequences. When the specific
128 insect vectors are added to support the '*Ca. Phytoplasma*' species definition, these should be
129 biologically confirmed as vectors, and the '*Ca. Phytoplasma*' species identity must be proved by
130 comparison of their 16S rRNA or conserved/housekeeping gene sequences.

131

132 **Overall comments on published '*Candidatus* Phytoplasma' species**

133 The published '*Ca. Phytoplasma*' species were grouped according to the availability of their full
134 length 16S rRNA gene sequence and of other conserved genes and listed in alphabetical order. Table
135 2 summarizes 14 '*Ca. Phytoplasma*' species and alternative reference strains with available full
136 length of their 16S rRNA gene sequences and genes *tufB*, *secA*, *secY*, *rplV-rpsC*, *groEL* gene
137 sequences. Table 3 includes '*Ca. Phytoplasma*' species and alternative reference strains with
138 available full length of their 16S rRNA gene sequences, and selected sequences of other
139 conserved/housekeeping genes.

140 There are 13 '*Ca. Phytoplasma*' species whose only available sequence is the full length 16S rRNA
141 gene sequence. Six '*Ca. Phytoplasma*' species start their 16S rRNA gene sequences at different
142 nucleotide positions when compared to full length sequence of the '*Ca. P. asteris*' species.
143 Sequences that start at nucleotide 6 include '*Ca. P. brasiliense*' strain HibWB26 [14] (GenBank
144 accession number AF147708); '*Ca. P. lycopersici*' strain Santa Cruz [15] (GenBank accession number
145 EF199549); '*Ca. P. oryzae*' strain RYD [16] (GenBank accession number D12581); '*Ca. P. palmicola*'
146 strain LYDM-178 [17] (GenBank accession number KF751387); '*Ca. P. sudamericanum*' strain
147 PassWB-Br3 [18] (GenBank accession number GU292081); '*Ca. P. tamaricis*' strain SCWB1 [19]
148 (GenBank accession number FJ432664). Sequences of '*Ca. P. balanitae*' strain BltWB [20] (GenBank
149 accession number AB689678) and '*Ca. P. spartii*' strain SpaWB [21] (GenBank accession number
150 X92869) start at nucleotide 7. Sequence start for '*Ca. P. castaneae*' strain CnWB [22] (GenBank
151 accession number AB054986) is at nucleotide 10, for '*Ca. P. dypsidis*' strain RID7941 [23] (GenBank
152 accession number MT293886) at nucleotide 25; for '*Ca. P. caricae*' strain PAY [24] (GenBank
153 accession number AY725234) and '*Ca. P. graminis*' strain SCYLP [24] (GenBank accession number
154 AY725228) at nucleotide 28 and for '*Ca. P. costaricanum*' strain SoyST1c1 [25] (GenBank accession
155 number HQ225630) at nucleotide 31.

156 The four following '*Ca. Phytoplasma*' species have their 16S rRNA gene sequences starting after
157 nucleotide 100 and are considered too short according to the newly proposed guidelines. '*Ca. P.*
158 *stylosanthis*' strain VPRI 43683 [26] (GenBank accession number MT431550) sequence starts at
159 nucleotide 169. However, its *tufB*, *secA* and *rplV-rpsC* gene sequences are available under GenBank
160 accession numbers MT432813 (364 nt); MT432821 (291 nt); MT461153 (1257 nt), respectively. The
161 '*Ca. P. omanense*', strain IM-1 [27] (GenBank accession number EF666051) sequence starts at
162 nucleotide 116, the '*Ca. P. wodyetiae*', strain FPYD Bangi-2 [28] (GenBank accession number

163 KC844879) sequence spans between nucleotides 149 and 1,399, and the '*Ca. P. allocasuarinae*',
164 strain AlloY [21] (GenBank accession number AY135523) sequence spans between nucleotides 370
165 and 1,527. These '*Ca. Phytoplasma*' species sequences must be completed for the same strain or for
166 an alternative reference strain.

167

168 '***Candidatus Phytoplasma palmae***' and '***Candidatus Phytoplasma cocostanzaniae***' description

169 Following the previous [1] and the revised guidelines described in this publication, the following two
170 '*Ca. Phytoplasma*' species are described, including some of their epidemiological and
171 phytopathological traits (Table 4).

172 '***Ca. P. cocostanzaniae***'. It includes 17 phytoplasma strains associated with coconut lethal yellowing
173 disease in Africa, mainly in Tanzania. The proposed reference strain is LD, associated with Tanzanian
174 coconut lethal disease [29,30]. The 16S rRNA sequences of 17 strains are deposited, and that of
175 strain Tanz08-05 (GenBank accession number GU952106) also comprises the spacer region (1,718
176 nucleotides). Compared to the reference strains of other known or newly designated '*Ca.*
177 *Phytoplasma*' species, the LD strain shares the highest 16S rDNA sequence identity (96.30%) with
178 that of the newly proposed reference strain '*Ca. P. palmae*' (Table 4). Unique signature sequences
179 (position related to the 16S rDNA sequence of the reference strain) were identified as follows: 5'-
180 GATAAGTCTCTAGTTTAATTCAGC-3' (nt 578-602); 5'-GTGTCGGGGCAACTCGGTAC-3' (nt 815-834);
181 5'-ATCGTTAGTTACCAGCATGTTATGA-3' (nt 1091-1115). All strains of '*Ca. P. cocostanzaniae*' share
182 16S rDNA sequence identities ranging from 99.33% to 100% when compared to the reference strain
183 and share the same unique signature sequences.

184 '***Ca. P. palmae***'. It includes 66 phytoplasma strains associated with coconut lethal yellowing and
185 other diseases affecting palms in the Americas. The strain "coconut lethal yellowing MLO" from
186 *Veitchia merrillii* [31] is proposed as the reference strain. This strain shares the highest 16S rDNA
187 sequence identity (96.30%) when compared to the reference strains of other '*Ca. Phytoplasma*'
188 species, including the newly proposed reference strain '*Ca. P. cocostanzaniae*' (Table 4). Unique
189 signature sequences (position related to the 16S rDNA sequence of the reference strain) were
190 identified as follows: 5'-GGCCTACCAAGACGATGATGTGT-3' (nt 255-277); 5'-
191 GTAGGCGGCTTACTGGGTCTTTACTG-3' (nt 710-735); 5'-GTCGTTAATTGCCAGCACGTTATGGTGGG-3'
192 (nt 1091-1119). A total of 52 '*Ca. P. palmae*' strains share higher 16S rDNA sequence identities
193 ranging from 98.65% to 100% and share the same unique signature sequences when compared to
194 the reference strain. Other 14 strains show sequence identities ranging from 98.05% to 98.50%

195 when compared to the reference strain. Nucleotide sequences of the genes *tufB*, *secA*, *secY*, *rpIV-*
196 *rpsC*, and *groEL* are available for the Texas Phoenix palm phytoplasma strain ACPD, with the draft
197 genome available (GenBank accession number VBRA02000000). The only identified insect vector for
198 '*Ca. P. palmae*' is *Haplaxius (Myndus) crudus* [32].

199

200 **Comments on '*Ca. Phytoplasma*' species listed in Table 2**

201 '***Ca. P. asteris***'. The strain MIAY, reference strain described in 2004, is retained [33]. The whole
202 genome sequences of three strains (OY-M, AY-WB, and M3) and multiple draft assemblies are
203 available. The strain OY-M was added as a complementary additional strain for other genes. A total
204 of 366 strains has 16S rDNA sequences covering >95% and sharing sequence identities ranging from
205 98.67% to 100% when compared to the reference strain MIAY (GenBank accession number
206 M30790). Among strains having 100% of sequence coverage, two share a 99.74% of identity, three
207 99.67%, two 99.61%, one 99.54%, and one 99.48% when compared to the reference strain.
208 Moreover, six strains show sequence identity ranges from 98.06% to 98.63%, while two strains show
209 97.77% sequence identity when compared to the reference strain. A large number of sequences are
210 deposited for either the 16S rRNA gene, and several other conserved/housekeeping genes.

211 '***Ca. P. aurantifolia***'. It was the first '*Ca. Phytoplasma*' species formally described in 1995 [34]; the
212 reference strain WBDL is retained. A total of 97 strains has 16S rDNA sequences covering >95% of
213 the gene and share sequence identities ranging from 98.69% to 99.73% when compared to the
214 reference strain WBDL (GenBank accession number U15442); 19 '*Ca. P. australasia*' strains share
215 from 98.32% to 98.53% of sequence identity with respect to '*Ca. P. aurantifolia*' reference strain.
216 Among the strains with a full sequence coverage, 25 have identities >99% when compared with the
217 reference strain WBDL, but there are no strains with fully matching sequences.

218 '***Ca. P. australiense***'. It was described in association with the Australian grapevine yellows [35]. The
219 reference strain AUSGY has a partial 16S rDNA sequence (1,375 nt; GenBank accession number
220 L76865), and a partial *tufB* gene sequence available. The strain CaPaus whole genome sequence is
221 available (GenBank accession number AM422018), and it is proposed as complementary additional
222 strain. A total of 12 strains had 16S rDNA sequences covering >95% of the gene and sharing
223 sequence identities ranging from 98.68% to 99.93% compared to the CaPaus strain sequence.

224 '***Ca. P. mali***'. The apple proliferation agent, strain AP15, was described in 2004 [36]. The whole
225 genome sequence of the severe strain AT is available [37]. This strain encodes two identical 16S
226 rRNA gene sequences (GenBank accession number CU469464). '*Ca. P. mali*' strains differs up to

227 0.20% in their 16S rRNA sequences based on alignments with a coverage of at least 99%. Sequence
228 identities >97.50% are available when compared with '*Ca. P. pyri*' and '*Ca. P. prunorum*'. '*Ca. P. pyri*'
229 strain PD1 and '*Ca. P. mali*' strain AT show 98.60% identity on their 16S rRNA gene sequence and
230 can be separated. A total of 20 '*Ca. Phytoplasma*' strains are deposited with sequences enclosing
231 >95% coverage and showing identities ranging from 99.74% to 100% to the reference strain.
232 However, 20 strains showing 98.22% to 99.01% are classified within diverse '*Ca. Phytoplasma*'
233 species ('*Ca. P. pyri*' and '*Ca. P. prunorum*'). On the aligned gene sequences strain PD1 shows lower
234 identities, compared to the AT strain for the genes *tufB* (95.00%), *secA* (93.00%), *secY* (94.00%), *rplV*-
235 *rpsC* (87.00%) and *groEL* (96.00%). '*Ca. P. mali*' strains differentiation is also possible based on
236 diverse insect vectors.

237 '***Ca. P. meliae***'. It was described in Argentina in 2016 with four strains identified with the prefix ChTY
238 [38]. The comparison of their 16S rRNA gene sequences described in that manuscript showed a
239 sequence identity of 99.82% to '*Ca. P. hispanicum*', the complete 16S rRNA gene sequence lowered
240 the identity to 98.95%. A total of 5 strains are deposited with sequences covering >95% and showing
241 identities ranging from 99.45% to 99.86% to the reference strain; however, 6 strains showing 98.82
242 to 99.46% of identity are classified within '*Ca. P. hispanicum*'. The *secA* and *rplV-rpsC* gene
243 sequences show identities of 93.75% and 95.14% respectively, indicating a clear taxon separation
244 compared with '*Ca. P. hispanicum*'. Having the draft genome sequence of the strain ChTY-XIII-Mo
245 (GenBank accession number NZ_JACAOD020000000) available, comparisons of its *secY* and *groEL*
246 gene sequences to those of the MPV strain (GenBank accession numbers GU004336 and KT444668)
247 indicate 90.08% and 93.84% of sequence identity respectively, supporting it as the retained
248 reference strain.

249 '***Ca. P. pini***'. It was described in 2005 [39] with the Spanish strain Pin127S (GenBank accession
250 number AJ632155) as the reference strain. Over 35 ribosomal 16S rRNA genes of various lengths
251 (484 bp to 1,250 bp) are available. The three longer sequences share an identities over 99.00%
252 (99.74% to 99.93%), except for the North American '*Ca. P. pini*' strains (GenBank accession numbers
253 KU242428 and VIAE01000001) that share 98.50% of sequence identity. '*Ca. P. pini*' shares the
254 highest 16S rRNA sequence identity compared to '*Ca. P. cynodontis*' and '*Ca. P. palmae*' with values
255 of about 93.00% to 94.00%. For some strains *secA* and *tufB* gene sequences are deposited. A draft
256 genome sequence of the North American '*Ca. P. pini*' strain MDPP is also available [40].

257 '***Ca. P. phoenicium***'. The A4 reference strain is retained [41]. The strain SA213 (draft genome
258 available but with a partial 16S rDNA sequence) was added as a complementary additional strain for

259 the other genes. For the strain SA213, 16S rRNA and *tufB* gene sequences were obtained from PCR
260 products, while nucleotide sequences of the genes *secA*, *secY*, *rplV-rpsC*, and *groEL* were retrieved
261 from the draft genome. Fifty-one strains have 16S rDNA sequences covering >95% of the gene and
262 sharing a sequence identities from 98.71% to 99.93% compared to the reference strain (A4)
263 sequence (GenBank accession number AF515636). A total of 27 strains shares sequence identities
264 ranging from 97.54% to 98.61% compared with the reference strain.

265 '***Ca. P. pruni***'. Proposed in 2013 [42], it is one of the phytoplasmas with the highest number of
266 described strains (203), showing 98.80% to 100% of nucleotide identities compared to the reference
267 strain. Four strains show 98.09%, 98.34%, 97.87% and 97.96% of sequence identity compared to the
268 reference strains. It is widely distributed, being mostly described in the American continent.

269 '***Ca. P. pyri***'. The reference strain PD1 (GenBank accession number AJ542543) was described along
270 with '*Ca. P. mali*' and '*Ca. P. prunorum*' [36] and sharing a 16S rRNA sequence identity >97.50%
271 compared with both. A total of 34 strains having >95% of sequence available showing 99.14% to
272 100% of sequence identity compared to the reference strain. Among strains classified within '*Ca. P.*
273 *mali*' and '*Ca. P. prunorum*', 64 show 98.68% to 99.27% and seven 97.50% to 98.64% of sequence
274 identity compared to the reference strain. The closely related '*Ca. P. pyri*' PD1 and '*Ca. P. mali*' AT
275 strains can be separated by sequences of additional genetic markers as mentioned for '*Ca. P. mali*'.
276 Differentiation is also possible based on differential insect vector transmission.

277 '***Ca. P. sacchari***'. Described from sugarcane in India [43]; the reference strain SCGS is retained. A
278 total of 30 strains has 16S rDNA sequences covering >95% of the gene and sharing sequence
279 identities ranging from 98.69% to 99.93% compared to the reference strain SCGS (GenBank
280 accession number MN889545), however no strains show 100% of sequence identity to it. A total of
281 11 strains having >95% of sequence available show 98.79% to 98.99% of sequence identity
282 compared to the reference strain. Seventeen strains are assigned to '*Ca. P. cynodontis*'. A draft
283 genome assembly is available for the reference strain SCGS^R (GenBank accession number
284 VWXM000000000), which has a genome-wide ANI value of 79.42% compared to the closest relative
285 with a genome assembly available, '*Ca. P. cynodontis*' strain LW01 (GenBank accession number
286 VWOH000000000). For comparison of the *groEL* gene, the partial sequence available from SCGS
287 shares 86.17% sequence identity with that of '*Ca. P. cynodontis*' strain 305/13.

288 '***Ca. P. solani***'. The reference strain STOL was described in 2013 and is retained [44]. For the strain
289 284/09, all gene sequences were available and retrieved from its genome sequence (GenBank
290 accession number FO393427). A total of 72 strains has 16S rDNA sequences covering >95% of the

291 gene and sharing a sequence identities ranging from 99.12% to 99.93%, compared to the reference
292 strain (STOL) sequence (GenBank accession number AF248959); one strain shows 98.17% of
293 sequence identity compared to the reference strain. Some strains [142/09, GenBank accession
294 number JQ730739 (98.05%); 429/19, GenBank accession number MT157232 (98.04%); 204/10,
295 GenBank accession number JQ730744 (98.04%); 198/10, GenBank accession number JQ730743
296 (98.04%); 224/09, GenBank accession number JQ730742 (98.04%); G66, GenBank accession number
297 JN887313 (98.04%); 241/13, GenBank accession number KF907506 (98.04%); Conv2/2010-Bg,
298 GenBank accession number JN561702 (98.03%); 161/16, GenBank accession number KY579338
299 (98.02%) share a sequence identity >98% also with the '*Ca. P. australiense*' strain CaPaus
300 (AM422018). '*Ca. P. solani*' and '*Ca. P. australiense*' strains are, however, clearly distinct based on
301 sequence identity comparisons of their *tufB* (82.00%-87.00%), *rplV-rpsC* (75.00%-82.00%), and *secY*
302 (55.00-75.00%) genes.

303 '***Ca. P. tritici***'. It was recently described from wheat in China [45]. The reference strain WBD has
304 98.68% to 99.93% sequence identity compared with 434 '*Ca. P. asteris*' strains, therefore, it does
305 not have the required threshold to be described as species based on the 16S rRNA gene. The
306 proposal of this new taxon was based on its unique vectorship, a distinctive symptomatology in its
307 predominant plant host, and <95% genome-wide ANI identity compared to several '*Ca. P. asteris*'
308 strains. Two other genes used for comparisons with '*Ca. P. asteris*' strains are *amp* and *secY*, which
309 have the highest amino acid sequence identities of 61.60% and 95.40%, respectively to '*Ca. P.*
310 *asteris*'.

311 '***Ca. P. ulmi***'. It was first described as a '*Ca. Phytoplasma*' species with appropriate threshold values
312 in compliance with the revised rules [46], however it is now impossible to distinguish it only based
313 on its 16S rRNA gene sequence since '*Ca. P. ziziphi*', '*Ca. P. rubi*' and "flavescence dorée"
314 phytoplasmas have identity percentages above the old and new thresholds. The reference strain
315 EY1 described in 2004, is retained. The strain ULW is added as a complementary additional strain.
316 Fourteen strains have 16S rDNA sequences covering >95% of the gene and sharing sequence
317 identities of 99.53% to 99.93% compared with strain EY1 (GenBank accession number AY197655).
318 A total of 69 strains showing sequence identities ranging from 98.82% to 99.80% were assigned to
319 other '*Ca. Phytoplasma*' species. Strains of '*Ca. P. ulmi*' with *rplV-rpsC* gene sequences covering
320 >89% share a sequence identity higher than 99.25% compared to the strain EY1, which shares a *rplV-*
321 *rpsC* gene sequence identity between 97.33% and 97.50%, and from 96.0% to 96.25% with strains
322 of '*Ca. P. rubi*' and '*Ca. P. ziziphi*' respectively, with a sequence coverage of 100%. Strains of '*Ca. P.*

323 ulmi' with *secY* gene sequences covering >77% share a sequence identity higher than 97.21%
324 compared to the reference strain EY1; whereas '*Ca. P. ulmi*' strain EY1 shares a *secY* gene sequence
325 identity of 92.37% and between 88.57-88.84% with strains of '*Ca. P. rubi*' and '*Ca. P. ziziphi*',
326 respectively, with a sequence coverage of 100%.

327 '***Ca. P. ziziphi***'. Among the strains deposited with 100% of sequence coverage, seven show 99.93%
328 of sequence identity compared to the reference strain JWB-G1 (GenBank accession number
329 AB052876) [47]. A total of 33 strains with >95% of sequence coverage shows identity percentages
330 ranging from 99.54% to 100% compared to the reference strain. Other 47 strains having a threshold
331 ranging between 98.69% and 99.35% were included within '*Ca. P. ulmi*', '*Ca. P. rubi*' and
332 "flavescence dorée" but should be reclassified due to substantial differences in other gene
333 sequences.

334

335 **Comments on '*Ca. Phytoplasma*' species listed in Table 3**

336 '***Ca. P. americanum***'. Four strains including the reference strain [48] are available, all from the USA,
337 showing 16S rRNA gene sequence identities ranging from 99.67% to 99.87% (GenBank accession
338 numbers DQ174118, DQ174120, MN227133 and DQ174121).

339 '***Ca. P. australasia***'. It was originally described in New Zealand [49], but it is also distributed in the
340 Asian and North African continents. Presently 249 strains show sequence identities ranging from
341 98.65% to 100%. A total of 97, 94 and 74 strains classified as '*Ca. P. aurantifolia*' share sequence
342 identities >97.5%, 98.00%, and 98.65%, respectively, compared with '*Ca. P. australasia*' reference
343 strain PpYC (GenBank accession number Y10097). GenBank shows 53 strains with a full length 16S
344 rRNA gene sequence identity ranging between 99.22% and 99.94%, compared to the reference
345 strain. Six strains were erroneously attributed to '*Ca. P. aurantifolia*' and one to '*Ca. P. australiense*',
346 while many other strains were attributed to the '*Ca. P. australiense*' due to a shorter sequence
347 coverage resulting in identity values >98%. However, a distinction can be achieved by comparing
348 additional genes (Tables 2 and S2). For the *tufB* gene, the majority of GenBank sequences fully match
349 that of '*Ca. P. australasia*', but were misclassified as '*Ca. P. aurantifolia*', while the actual identity
350 threshold for '*Ca. P. aurantifolia*' starts from 94.03%. Similarly, for *secA* and ribosomal protein (*rp*)
351 genes, hundreds of strains in GenBank described as '*Ca. P. aurantifolia*' are misclassified showing
352 sequence identities above 99.00%, compared to the corresponding genes of '*Ca. P. australasia*'.

353 '***Ca. P. cirsii***'. It was described in the Czech Republic in association with yellowing, stunting, and
354 proliferation of creeping thistle and dahlia [50]. Two strains, enclosing the reference strain CirYS

355 (GenBank accession number KR869146), are available. Compared to the reference strain CirYS, the
356 strains CirYS1 and DahIP have a 16S rDNA sequence identity percentage of 100.00% and 99.93%,
357 respectively. The nucleotide sequence of the *secA* gene is available for the reference strain.

358 '***Ca. P. convolvuli***'. It was described associated with bindweed yellows in several European countries
359 [51]. The reference strain BY-S57/11 (GenBank accession number JN833705) is retained. A total of
360 9 strains had 16S rDNA sequences covering >95% of the gene sharing sequence identity \geq 99.93%-
361 100% compared to the reference strain. In GenBank, three phytoplasma strains associated with
362 *Carica papaya* bunchy top in Nigeria showed a 98.73% sequence identity compared to '*Ca. P.*
363 *convolvuli*' reference strain.

364 '***Ca. P. cynodontis***'. It was described in association with the Bermuda grass white leaf (BGWL)
365 disease and includes strains from Asian and European countries [52]. The reference strain BGWL-C1
366 (GenBank accession number AJ550984) is retained, and the strain 305/13 (GenBank accession
367 number KP019340) is proposed as a complementary additional strain for the availability of its *groEL*
368 gene. A total of 35 strains with a 16S rDNA sequence coverage of >95% shares a sequence identities
369 ranging from 98.66% to 100% compared to the reference strain. Among them, four strains from
370 Italy, Albania and Iran show a 100% of sequence identity, compared to each other and to the
371 reference strain. One strain shows 98.37% and 15 strains show sequence identity percentages
372 ranging from 98.33% to 98.84% compared to the reference strain. Other strains had a 16S rDNA
373 sequence identity ranging from 98.00% to 99.87% compared to the reference strain, including four
374 strains assigned to '*Ca. P. sacchari*' and one strain assigned to '*Ca. P. oryzae*'. Comparing the *groEL*
375 gene, '*Ca. P. cynodontis*' and '*Ca. P. sacchari*' can be distinguished (sequence identity 83.80%).

376 '***Ca. P. fragariae***'. It was described as associated with strawberry yellows in Lithuania [53]. Strain
377 StrawY is the reference strain (GenBank accession number DQ086423) and it is retained. The strain
378 GBFC_SY_01 was added as a complementary additional strain. A total of 10 strains had a 16S rDNA
379 sequences coverage >95% and share sequence identities from 99.46% to 99.93% compared to the
380 reference strain. Four strains show 98.23% to 98.63% and one show a 97.67% of sequence identity
381 to the reference strain. The species '*Ca. P. fragariae*' is most closely related to '*Ca. P. japonicum*'
382 and it was also detected in potato plants in China.

383 '***Ca. P. fraxini***' [54]. The reference strain AshY1 is retained, with *secY* and *groEL* gene sequences
384 available. There are 9 strains having their 16S rDNA sequences covered for over >95% and sequence
385 identities from 99.53% to 99.93% compared to AshY1 reference strain. There are further 8 strains
386 showing identities ranging from 98.01% to 98.61% to the reference strain and 5 strains with

387 identities ranging between 97.61% and 97.81%. Two complete sequences and four partial
388 sequences of *rplV-rpsC* genes are available showing nucleotide identities over 99.45% among them.
389 'Ca. P. fraxini' strains were mainly identified in the American continent. Further studies are
390 necessary to clarify the identity of a phytoplasma detected in *Crotalaria juncea* in Brazil [55] showing
391 97.6% of the 16S rRNA sequence identity (GenBank accession number KF941132) and 92.03% for
392 the *rplV-rpsC* genes (GenBank accession number KJ806620) to the reference strain.

393 '**Ca. P. japonicum**'. It was identified in Japan [56] (reference strain JHP, GenBank accession number
394 AB010425); one strain was detected in China in *Sophora japonica* (GenBank accession number
395 FJ685751). Nucleotide sequences of the genes *secY* and *groEL* are available for the reference strain.

396 '**Ca. P. hispanicum**'. Identified in periwinkle in Mexico [57] it has >98% 16S rRNA gene identity to
397 'Ca. P. meliae'. The reference strain MPV, for which the sequences of the genes *rplV-rpsC*, *secY*,
398 *secA*, and *groEL* are available, is retained. A total of six strains had 16S rDNA sequences covering
399 >95% of the gene sharing sequence identities of 98.74% to 99.47%. Moreover a 98.53% of sequence
400 identity is present for one 'Ca. P. hispanicum' strain, while 98.82% to 98.89% sequence identity
401 correspond to four strains of 'Ca. P. meliae'.

402 '**Ca. P. luffae**'. First identified in Taiwan [58], there are now 13 16S rRNA gene sequences longer
403 than 1,200 bp deposited in GenBank showing 99.87% to 99.93% of sequence identity compared with
404 the reference strain LfWB^R. However, LfWB^R was lost, and another strain collected in Taiwan,
405 NCHU2019, with a complete genome sequence available (GenBank accession number CP054393) is
406 proposed as an additional complementary strain. NCHU2019 and LfWB^R share a 100% of sequence
407 identity of their 16S rRNA and *secY* gene sequences, and a 97.58% sequence identity of their *tufB*
408 gene sequences.

409 '**Ca. P. malaysianum**'. Described from Malaysia [59], one more strain is described from South Korea
410 having 99.54% of sequence identity to its 16S rRNA sequence. One partial sequence (482 bp) of the
411 *secA* gene is also available.

412 '**Ca. P. noviguineense**'. It was described in Papua New Guinea in association with Bogia coconut
413 syndrome (BCS) and banana wilt (BW) in coconut and banana plants, respectively [60]. The
414 reference strain is BCS-Bo^R from coconut (GenBank accession number LC228755). A total of 26
415 strains with a 16S rDNA sequence coverage >95% shares sequence identities from 99.66% to 100%
416 compared to the reference strain. Sequence identities of both BCS and BW strains to other
417 phytoplasma taxa are about 96% with maximum values of 96.08%, 95.91%, 95.20% for 'Ca. P.
418 palmae', 'Ca. P. cocostanzaniae', and 'Ca. P. palmicola', respectively. Another phytoplasma on the

419 same island associated with arecanut yellow leaf (ALL) disease showed high 16S rRNA gene
420 sequence identities and is closely related to BCS-Bo^R. Additional sequence comparison of the *secY*
421 (GenBank accession number LC228769) and *rpIV-rpsC* genes (GenBank accession numbers
422 LC228762) for the strain BCS-Bo^R, (GenBank accession number LC228763) for strain BCS-G, and
423 (GenBank accession number LC228764) for strain BCS-S showed 100% sequence identity compared
424 to the reference strain BCS-Bo^R.

425 '**Ca. P. prunorum**'. The reference strain EFSY-G1 (GenBank accession number AJ542544) was
426 described with '*Ca. P. mali*' and '*Ca. P. pyri*' [36]. There are 32 strains having 16S rDNA sequences
427 covering >95% with identity percentages from 99.50% to 100% compared to the reference strain.
428 Moreover 63 '*Ca. P. pyri*' and '*Ca. P. mali*' strains show sequence identities from 98.67% to 99.14%,
429 and 13 show identities between 98.31% and 98.62% to the reference strain. However,
430 differentiation of these '*Ca. Phytoplasma*' species is possible based on insect vector differential
431 transmission.

432 '**Ca. P. rhamnii**'. It was identified in 1994 in buckthorn plants in south-west Germany [21]. Nine
433 additional '*Ca. P. rhamnii*' 16S rRNA sequences from Austria (GenBank accession number KF498655),
434 Germany (GenBank accession numbers KF498651-52, JQ868449), Serbia (GenBank accession
435 numbers KF498656-58) and Switzerland (GenBank accession numbers KF498653-54) are available
436 which are fully identical to each other's but differ from the reference strain BAWB. The complete
437 *rpIV* and a partial sequence of the *rpsC* gene are available from eight strains (GenBank accession
438 numbers KF498659–66) all showing a high sequence identity. *SecA* and *tufB* gene sequences are also
439 available (Table 2).

440 '**Ca. P. rubi**' [61]. It was described based on biological properties and 7 additional strains have been
441 identified with a 16S rRNA sequence identity between 99.83% and 99.88% compared to the
442 reference strain. There are 86 strains with sequence identity ranging from 98.66% to 99.52% which
443 are classified within different '*Ca. Phytoplasma*' species, or taxa not yet named as '*Ca. Phytoplasma*'
444 species. A threshold for its differentiation and support can be settled for *secA* (98.37%) and *secY*
445 (97.03%) genes.

446 '**Ca. P. trifolii**'. It was associated with clover proliferation in *Trifolium hybridum* plants [62]. The
447 reference strain CP (GenBank accession number KJ462045) is retained. For this strain the sequences
448 of *secY*, *secA* and *rpIV-rpsC* genes are available. Seventy-six strains have 16S rDNA sequences
449 covering >95% of the gene sharing sequence identities between 98.77 and 100%, three strains range
450 from 98.15 to 98.63% and one has 97.98% sequence identity compared to the reference strain.

451 There are 13 *secA* gene sequences available showing sequence identities between 99.10% and
452 99.28% to the reference strain CP.

453

454 **Comments on ‘*Ca. Phytoplasma*’ species described only based on the 16S rRNA gene**

455 ‘*Ca. P. allocasuarinae*’. Only the reference strain sequence is available from Australia [21]. A strain
456 from *Empoasca* sp. in Cuba (GenBank accession number AY725236) shows 98.52% sequence
457 identity to it (with a 99% coverage of the 16S rRNA sequence).

458 ‘*Ca. P. balanitae*’. The reference strain BltWB (GenBank accession number AB689678) was
459 described in Myanmar, infecting the unique species *Balanites triflora* [20], endemic in that state. A
460 total of 13 strains with >95% of 16S rRNA gene sequence coverage is available and share 99.41% to
461 99.80% sequence identity to the reference strain. It was attributed to several other phytoplasma
462 strains detected in other plant species in India. This attribution is not correct since, following the
463 previous rules it must only be used for this phytoplasma when it is infecting *B. triflora* not present
464 in India.

465 ‘*Ca. P. brasiliense*’ [14]. The reference strain HibWB26 (GenBank accession number AF147708) has
466 16S rRNA gene sequence identities ranging from 88.80% to 96.40% when compared with other ‘*Ca.*
467 *Phytoplasma*’ species. A total of 10 strains with >95% coverage of their 16S rRNA gene sequences is
468 available and share 98.73% to 99.93% sequence identity to the reference strain. Further 16S rRNA
469 gene sequences are available for the Peruvian papaya phytoplasma (GenBank accession number
470 KX810334-36) and grapevine (GenBank accession numbers KX670807-9) phytoplasma strains that
471 show 98.64% sequence identity with the reference strain. A *groEL* gene sequence of 555 nt of the
472 strain identified in papaya in Peru is available (GenBank accession number MH279494).
473 Phytoplasmas were reported with 16S rDNA sequence identities of 98.84% from *Guazuma ulmifolia*
474 (GenBank accession number HQ258882) in Costa Rica, 99.68% from *Sida rhombifolia* (GenBank
475 accession number HQ230579) and 99.94% from *Crotalaria juncea* (GenBank accession number
476 KF878382) in Brazil. A ‘*Ca. P. brasiliense*’ strain found in hibiscus in Egypt (GenBank accession
477 number KF716175) showed 99.60% sequence identity with the reference strain. Further strains
478 were identified in peach in Azerbaijan (GenBank accession number FR717540) and in *Catharanthus*
479 *roseus* in Costa Rica (GenBank accession number MH428963).

480 ‘*Ca. P. caricae*’. It was identified from papaya plants in Cuba in 2005 and the reference strain is PAY
481 (GenBank accession number AY725234) [24]. It shares 95.80% sequence identity of the 16S rRNA
482 gene with ‘*Ca. P. graminis*’.

483 '**Ca. P. castaneae**'. It was described from infected chestnut in South Korea [23] and two more strains
484 (GenBank accession numbers MW264918 and EU599362) were reported from China.

485 '**Ca. P. costaricanum**'. It was described as associated with phytoplasma diseases in soybean, sweet
486 pepper, and passionfruit in Costa Rica [25]. The retained reference strain is SoyST1c1 (GenBank
487 accession number HQ225630). A total of 17 strains has 16S rDNA sequence coverage >95% and
488 share sequence identities between 99.15% and 99.61% compared to the reference strain.

489 '**Ca. P. dypsidis**'. Recently described in Australia [23] from dying ornamental palms belonging to
490 several species. Six strains with coverage >95% show sequence identities between 99.83% and
491 99.88% compared to the reference strain RID7692 (GenBank accession number MT536195). The
492 closest phytoplasma is '*Ca. P. cocostanzaniae*' with a sequence identity <96%.

493 '**Ca. P. graminis**'. It was identified in sugarcane in Cuba [24]. The reference strain is SCYLP (GenBank
494 accession number AY725228). Other strains were identified in Cuba in *Saccharosydne saccharivora*,
495 *Cedusa* spp., *Cynodon dactylon*, *Conyza canadensis*, *Macroptilium lathyroides*, and *Sorghum*
496 *halepense*. Four strains with 16S rDNA coverage >95% show identities between 99.61% and 99.74%
497 and one strain has identity of 98.43% compared to the reference strain.

498 '**Ca. P. lycopersici**'. It was described in Bolivia [15] in tomato and *Morrenia variegata*. The reference
499 strain is Santa Cruz (GenBank accession number EF199549). The 16S rDNA sequence of only one
500 strain is deposited in GenBank sharing 97.51% sequence identity with some '*Ca. P. asteris*' strains.

501 '**Ca. P. omanense**'. Identified in Oman in *Cassia italica* [27], the reference strain IM-4 (GenBank
502 accession number EF666054) has 99.64% sequence identity compared to the other three strains
503 (GenBank accession numbers EF666051-53); a 99.58% of sequence identity is present in the only
504 strain with >95% of sequence available. A strain from Australia from *Vigna* sp. (GenBank accession
505 number AJ289195) shares 98.04% sequence identity to the reference strain.

506 '**Ca. P. oryzae**'. It was reported infecting rice [16]; the reference strain is RYD-J^R collected in Japan.
507 The strain RYD-Th from Thailand (GenBank accession number AB052873), shares 99.20% of 16S
508 rRNA gene identity with RYD-J^R. Two other strains reported as '*Ca. P. oryzae*' have draft genome
509 sequences available, both were collected from infected Napier grass, including Mbita1 (GenBank
510 accession number LTBM00000000) from Kenya and NGS-S10 (GenBank accession number
511 JHUK00000000) from Ethiopia. However, Mbita1 lacks a 16S rRNA gene in the draft genome
512 sequence while the 16S rDNA of NGS-S10 shares 97.38% sequence identity with RYD-J^R. Three more
513 strains are available with a 16S rRNA gene sequence identity of 99.93%, 99.78% and 99.56%,

514 respectively to that of the NGS-S10 strain. More strains have a sequence identity of <98.20% and
515 are deposited under the '*Ca. P. cynodontis*' taxon.

516 '***Ca. P. palmicola***'. It was described from coconut palms with symptoms of lethal yellowing (LY) in
517 Mozambique [17], the reference strain is LYDM-178 (GenBank accession number KF751387). The
518 16S rDNA sequence alignments confirmed the identity with those of Awka wilt disease (LDN) in
519 Nigeria (GenBank accession number Y14175), and 99.00% to 99.60% of sequence identity with those
520 of the Cape St Paul Wilt disease (CSPWD) in Ghana (GenBank accession numbers Y13912 and
521 JQ868442) and Côte d'Ivoire, CILY, (GenBank accession numbers KC999037, KF364359, KF387570
522 and KF419286). A total of 24 sequences with identities ranging from 99.28% to 100% compared to
523 the reference strain is available. Strains from Côte d'Ivoire and Ghana can be differentiated by single
524 nucleotide polymorphisms (SNPs) in their 16S rRNA gene sequences. There are 32 *secA* sequences
525 of 627 bp (GenBank accession numbers LR029104 to LR029135) corresponding to Mozambican
526 strains, but none for the LYDM reference strain. Other *secA* sequences of strains from Nigeria
527 include GenBank accession numbers LR029136 to LR029139. Ghana/Côte d'Ivoire strain sequences
528 for the genes *tufB* (GenBank accession number JQ824292, 391 bp) and *rplV-rpsC* (GenBank
529 accession numbers KU925788 to KU925794, 825 nt; GenBank accession numbers LR028744 to
530 LR028839, 321 nt) are available.

531 '***Ca. P. spartii***'. Only identified in *Spartium junceum* in Europe. The reference strain is SpaWB
532 (GenBank accession number X92869) [21]. There are six strains deposited with sequence identities
533 ranging from 98.97% to 99.84% to the reference strain.

534 '***Ca. P. stylosanthis***'. Described in Australia in diverse plant species [26]. Among the four sequences
535 available, only one with >95% 16S rRNA gene sequence coverage shares 99.94% of sequence
536 identity with the reference strain. The following GenBank accession numbers correspond to the
537 reference strain: MT431550, 16S rRNA gene; MT432821, partial *secA* gene; MT432813, partial *tufB*
538 gene; MT461153, partial *rps19-rpl22-rps3* gene.

539 '***Ca. P. tamaricis***'. Identified in China [19], one more strain (GenBank accession number MW447513)
540 is available sharing a 99.67% of sequence identity to it but with only 67% of coverage.

541 '***Ca. P. sudamericanum***'. It was identified in passion fruit plants in Brazil [18]: the reference strain is
542 PassWB-Br3 (GenBank accession number GU292081) and has <97.50% sequence identity on 16S
543 rRNA gene compared with other '*Ca. Phytoplasma*' species.

544 '***Ca. P. wodyetiae***'. Described from *Wodyetia bifurcata* [28] in Malaysia following cloning mixed
545 phytoplasma infected samples, one more cloned sequence (GenBank accession number KY069029)

546 is available with a 98.48% of 16S rDNA sequence identity, compared to the original clone provided
547 as reference strain.

548

549 **Concluding remarks**

550 A correct designation and naming of '*Ca. Phytoplasma*' species is needed to support epidemiological
551 studies in order to effectively manage the phytoplasma-associated diseases, most of them, known
552 as devastating for agricultural crops worldwide. The appropriate name attribution is also important
553 for quarantine purposes to restrict the spreading of such diseases. The previous guidelines
554 recognized a new '*Ca. Phytoplasma*' species if the phytoplasma shared < 97.5% of its 16S rRNA gene
555 sequence identity, when compared with previously published '*Ca. Phytoplasma*' species. In addition,
556 a '*Ca. Phytoplasma*' species was also considered new, when it shared > 97.5% of its 16S rRNA gene
557 sequence identity with existing species, and it was clearly proven that it represented an ecologically
558 separated population. For such cases, the description of two different '*Ca. Phytoplasma*' species
559 was recommended only when all three of the following conditions apply (i) the two phytoplasmas
560 are transmitted by different vectors; (ii) the two phytoplasmas have a different natural plant host
561 (or, at least, their behaviour is significantly different in the same plant host); (iii) there is evidence
562 of significant molecular diversity, achieved by either hybridization to cloned DNA probes, serological
563 reaction or a PCR-based assay [1].

564 To date, more than 30 phytoplasma genomes (completed and drafted) have been published. Two
565 species have been named based on their average nucleotide identity (ANI) value of <95% [43,45].
566 The revised guidelines do not support the '*Ca. P. stylosanthis*', '*Ca. P. omanense*', '*Ca. P. wodyetiae*',
567 and '*Ca. P. allocasuarinae*' species since they only have short or not long enough 16Sr RNA gene
568 sequences available in GenBank. Sequences in compliance with the revised guidelines should be
569 provided for these '*Ca. Phytoplasma*' species to be retained. The list of signature sequences
570 deposited for all the officially published '*Ca. Phytoplasma*' species (Table S3) is provided in order to
571 assist in the process of '*Ca. Phytoplasma*' species designation.

572 For assigning '*Ca. Phytoplasma*' species it is necessary that the phytoplasma has the whole 16S rRNA
573 gene sequence with identity <98.65% or <95% ANI value, available. For those '*Ca. Phytoplasma*'
574 species with 16S rRNA gene sequence identities >98.65%, the following threshold values based on
575 housekeeping genes should be used to support their effective distinction: 97.6% for *groEL* gene,
576 97.5% for *tuf* and *rp* genes, 95.7% for *secA* gene and 95.0% for *secY* gene. The new thresholds
577 include: a 16S rRNA gene sequence identity of 98.65%; a genome ANI of 95%; and two among five

578 suggested housekeeping genes. For example, if the 16S rRNA gene sequence identity for a given
579 phytoplasma is < 98.65% there is no need to check other sequences. If the 16S rRNA gene sequence
580 identity is > 98.65% and the genome ANI is < 95%, checking other genes is not required. The
581 housekeeping genes are used when 16S rRNA gene sequence identities are > 98.65% and the whole
582 genome sequence is unavailable. There are not guidelines to select or validate the name of a
583 particular 'Ca. Phytoplasma' species. This should follow the required grammatical rules, the specific
584 geographic distribution, the major or first plant host where the phytoplasma was identified. The
585 revised guidelines support all the previously assigned 'Ca. Phytoplasma' species, except four that
586 must be adjusted to fit the revised guidelines. Previous IRPCM guidelines [1] should be followed if
587 relevant for a complete description of the new 'Ca. Phytoplasma' species when not conflicting with
588 the present revised guidelines.

589

590 **Acknowledgements:** Y. Arocha-Rosete, N. Contaldo, B. Duduk, N. Fiore, H.G. Montano, M. Kube,
591 M. Martini, K. Oshima, F. Quaglino, B. Schneider and A. Zamorano belong to the informal taxonomy
592 committee of the International Phytoplasmologist Working Group (IPWG) to which belongs also A.
593 Bertaccini and C-H. Kuo. A. Bertaccini, C-H. Kuo and W. Wei are members of the International
594 Committee on Systematics of Prokaryotes (ICSP) Subcommittee on the taxonomy of Mollicutes. The
595 update guidelines are approved by the entire Subcommittee.

596

597 Literature

- 598 **1. IRPCM.** 'Candidatus Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that
599 colonise plant phloem and insects. *Int J Syst Evolut Microbiol* 2004;54:1243-1255.
- 600 **2. Bertaccini A, Duduk B, Paltrinieri S, Contaldo N.** Phytoplasmas and phytoplasma diseases: a
601 severe threat to agriculture. *Am J Plant Sci* 2014;5:1763-1788.
- 602 **3. Kumari S, Nagendran K, Rai AB, Singh B, Rao GP, et al.** Global status of phytoplasma diseases
603 in vegetable crops. *Front Microbiol* 2019;10:1349.
- 604 **4. Contaldo N, Satta E, Zambon Y, Paltrinieri S, Bertaccini A.** Development and evaluation of
605 different complex media for phytoplasma isolation and growth. *J Microbiol Methods*
606 2016;127:105-110.
- 607 **5. Namba S.** Molecular and biological properties of phytoplasmas. *Proc Jpn Acad Ser B Phys Biol*
608 *Sci.* 2019;95:401-418.

- 609 **6. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, et al.** Uniting the classification of cultured
610 and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*
611 2014;12(9):635-645.
- 612 **7. Kim M, Oh H-S, Park S-C, Chun J.** Towards a taxonomic coherence between average nucleotide
613 identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J*
614 *Syst Evol Microbiol* 2014;64:346-351.
- 615 **8. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P et al.** DNA–DNA
616 hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst*
617 *Evol Microbiol* 2007;57:81–91.
- 618 **9. Richter M, Rosselló-Móra R.** Shifting the genomic gold standard for the prokaryotic species
619 definition. *Proc Natl Acad Sci USA* 2009;106:19126–19131.
- 620 **10. Bertaccini A, Paltrinieri S, Contaldo N.** Standard detection protocol: PCR and RFLP analyses
621 based on 16S rRNA gene. *Methods Mol Biol* 2019;1875:83-95.
- 622 **11. Martini M, Quaglino F, Bertaccini A.** Multilocus genetic characterization of phytoplasmas. In:
623 Phytoplasmas: Plant Pathogenic Bacteria - III Genomics, host pathogen interactions, and
624 diagnosis. Eds. Bertaccini A., K. Oshima, M. Kube, G.P. Rao. Chapter 9. Pag. 161-200. 2019.
625 Springer, Singapore.
- 626 **12. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S.** High throughput ANI analysis
627 of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
- 628 **13. Cho S-T, Kung H-J, Huang W, Hogenhout SA, Kuo C-H.** Species boundaries and molecular
629 markers for the classification of 16Srl phytoplasmas inferred by genome analysis. *Front*
630 *Microbiol* 2020;11:1531.
- 631 **14. Montano HG, Davis RE, Dally EL, Hogenhout S, Pimentel JP, et al.** 'Candidatus Phytoplasma
632 brasiliense', a new phytoplasma taxon associated with hibiscus witches' broom disease. *Int J*
633 *Syst Evol Microbiol* 2001;51:1109-1118.
- 634 **15. Arocha Y, Antesana O, Montellano E, Franco P, Plata G, et al.** 'Candidatus Phytoplasma
635 lycopersici', a phytoplasma associated with "hoja de perejil" disease in Bolivia. *Int J Syst Evol*
636 *Microbiol* 2007;57:1704-1710.
- 637 **16. Jung H-Y, Sawayanagi T, Wongkaew P, Kakizawa S, Nishigawa H, et al.** 'Candidatus
638 Phytoplasma oryzae', a novel phytoplasma taxon associated with rice yellow dwarf disease. *Int*
639 *J Syst Evol Microbiol* 2003;53, 1925-1929.

- 640 **17. Harrison NA, Davis RE, Oropeza C, Helmick EE, Narváez M, et al.** 'Candidatus Phytoplasma
641 palmicola', associated with a lethal yellowing-type disease of coconut (*Cocos nucifera* L.) in
642 Mozambique. *Int J Syst Evol Microbiol* 2014;64:1890-1899.
- 643 **18. Davis RE, Zhao Y, Dally EL, Jomantiene R, Lee IM, et al.** 'Candidatus Phytoplasma
644 sudamericanum', a novel taxon, and strain PassWB-Br4, a new subgroup 16SrIII-V phytoplasma,
645 from diseased passion fruit (*Passiflora edulis* f. *flavicarpa* Deg.). *Int J Syst Evol Microbiol*
646 2012;62:984-989.
- 647 **19. Zhao Y, Sun Q, Wei W, Davis RE, Wu W, et al.** 'Candidatus Phytoplasma tamaricis', a novel
648 taxon discovered in witches'-broom-diseased salt cedar (*Tamarix chinensis* Lour.). *Int J Syst Evol*
649 *Microbiol* 2009;59:2496-2504.
- 650 **20. Win NKK, Lee S-Y, Bertaccini A, Namba S, Jung H-Y.** 'Candidatus Phytoplasma balanitae'
651 associated with witches' broom disease of *Balanites triflora*. *Int J Syst Evol Microbiol*
652 2013;63:636-640.
- 653 **21. Marcone C, Gibb KS Streten C, Schneider B.** 'Candidatus Phytoplasma spartii', 'Candidatus
654 Phytoplasma rhamni' and 'Candidatus Phytoplasma allocasuarinae', respectively associated
655 with spartium witches' broom, buckthorn witches' broom and allocasuarina yellows diseases.
656 *Int J Syst Evol Microbiol* 2004;54:1025-1029.
- 657 **22. Jung H-Y, Sawayanagi T, Kakizawa S, Nishigawa H, Miyata S, et al.** 'Candidatus Phytoplasma
658 castaneae', a novel phytoplasma taxon associated with chestnut witches' broom disease. *Int J*
659 *Syst Evol Microbiol* 2002;52:1543-1549.
- 660 **23. Jones LM, Pease B, Perkins SL, Constable FE, Kinoti WM et al.** 'Candidatus Phytoplasma
661 dyspidsis', a novel taxon associated with a lethal wilt disease of palms in Australia. *Int J Syst Evol*
662 *Microbiol* 2021;71:doi.org/10.1099/ijsem.0.004818.
- 663 **24. Arocha Y, López M, Piñol B, Fernández M, Picornell B, et al.** 'Candidatus Phytoplasma graminis'
664 and 'Candidatus Phytoplasma caricae', two novel phytoplasmas associated with diseases of
665 sugarcane, weeds and papaya in Cuba. *Int J Syst Evol Microbiol* 2005;55:2451-2463.
- 666 **25. Lee I-M, Bottner-Parker KD, Zhao Y, Villalobos W, Moreira L.** 'Candidatus Phytoplasma
667 costaricanum' a novel phytoplasma associated with an emerging disease in soybean (*Glycine*
668 *max*). *Int J Syst Evol Microbiol* 2011;61:2822-2826.
- 669 **26. Jardim BR, Kinoti WM, Tran-Nguyen LTT, Gambley C, Rodoni B, et al.** 'Candidatus Phytoplasma
670 stylosanthis', a novel taxon with a diverse host range in Australia, characterised using multilocus
671 sequence analysis of 16S rRNA, *secA*, *tuf*, and *rp* genes. *Int J Syst Evol Microbiol* 2021;71:004589.

- 672 **27. Al-Saady NA, Khan AJ, Calari A, Al-Subhi AM, Bertaccini A.** 'Candidatus Phytoplasma
673 omanense', associated with witches' broom of *Cassia italica* (Mill.) Spreng in Oman. *Int J Syst*
674 *Evol Microbiol* 2008;58:461-466.
- 675 **28. Naderali N, Nejat N, Vadamalai G, Davis RE, Wei W, et al.** 'Candidatus Phytoplasma wodyetiae',
676 a new taxon associated with yellow decline disease of foxtail palm (*Wodyetia bifurcata*) in
677 Malaysia. *Int J Syst Evol Microbiol* 2017;67:3765-3772.
- 678 **29. Schuiling M, Nienhaus F, Kaiza DA.** The syndrome in coconut palms affected by a lethal disease
679 in Tanzania. *J Plant Dis Prot* 1981;88:665-677.
- 680 **30. Tymon AM, Jones P, Harrison NA.** Phylogenetic relationships of coconut phytoplasmas and the
681 development of specific oligonucleotide PCR primers. *Ann Appl Biol* 1998;132:437-452.
- 682 **31. Harrison NA, Richardson PA, Kramer JB, Tsai JH.** Detection of the mycoplasma-like organism
683 associated with lethal yellowing disease of palms in Florida by polymerase chain reaction. *Plant*
684 *Pathol* 1994;43:998-1008.
- 685 **32. Ogle L, Harries H.** Introducing the vector: how coconut lethal yellowing disease may have
686 reached the Caribbean. *Ethnobot Res Appl* 2005;3:139-142.
- 687 **33. Lee I-M, Gundersen-Rindal DE, Davis RE, Bottner KD, Marcone C, et al.** 'Candidatus
688 Phytoplasma asteris', a novel taxon associated with aster yellows and related diseases. *Int J Syst*
689 *Evol Microbiol* 2004;54:1037-1048.
- 690 **34. Zreik L, Carle P, Bové J-M, Garnier M.** Characterization of the mycoplasma-like organism
691 associated with witches' broom disease of lime and proposition of a 'Candidatus' taxon for the
692 organism, 'Candidatus Phytoplasma aurantifolia'. *Int J Syst Bacteriol* 1995;45, 449-453.
- 693 **35. Davis RE, Dally EL, Gundersen DE, Lee I-M, Habili N.** 'Candidatus Phytoplasma australiense', a
694 new phytoplasma taxon associated with Australian grapevine yellows. *Int J Syst Bacteriol*
695 1997;47:262-269.
- 696 **36. Seemüller E, Schneider B.** 'Candidatus Phytoplasma mali', 'Candidatus Phytoplasma pyri' and
697 'Candidatus Phytoplasma prunorum', the causal agents of apple proliferation, pear decline and
698 European stone fruit yellows, respectively. *Int J Syst Evol Microbiol* 2004;54:1217-1226.
- 699 **37. Kube M, Schneider B, Kuhl H, Dandekar T, Heitmann K, et al.** The linear chromosome of the
700 plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. *BMC Genomics* 2008;26:306.
- 701 **38. Fernández FD, Galdeano E, Kornowski MV, Arneodo JD, Conci LR.** Description of 'Candidatus
702 Phytoplasma meliae', a phytoplasma associated with Chinaberry (*Melia azedarach* L.) yellowing
703 in South America. *Int J Syst Evol Microbiol* 2016;66:5244-5251.

- 704 **39. Schneider B, Torres E, Martín MP, Schröder M, Behnke HD, et al.** 'Candidatus Phytoplasma
705 pini', a novel taxon from *Pinus silvestris* and *Pinus halepensis*. *Int J Syst Evol Microbiol*
706 2005;55:303-307.
- 707 **40. Cai W, Shao J, Zhao Y, Davis RE, Costanzo S.** Draft genome sequence of 'Candidatus
708 Phytoplasma pini'-related strain MDPP: a resource for comparative genomics of gymnosperm-
709 infecting phytoplasmas. *Plant Dis* 2020;104:1009-1010.
- 710 **41. Verdin E, Salar P, Danet J-L, Choueiri E, Jreijiri F, et al.** 'Candidatus Phytoplasma phoenicium'
711 sp. nov., a novel phytoplasma associated with an emerging lethal disease of almond trees in
712 Lebanon and Iran. *Int J Syst Evol Microbiol* 2003;53:833-838.
- 713 **42. Davis RE, Zhao Y, Dally EL, Lee I-M, Jomantiene R, et al.** 'Candidatus Phytoplasma pruni', a
714 novel taxon associated with X-disease of stone fruits, *Prunus* spp.: multilocus characterization
715 based on 16S rRNA, *secY*, and ribosomal protein genes. *Int J Syst Evol Microbiol* 2013;63:766-
716 776.
- 717 **43. Kirdat K, Tiwarekar B, Thorat V, Sathe S, Shouche Y, et al.** 'Candidatus Phytoplasma sacchari',
718 a novel taxon - associated with Sugarcane Grassy Shoot (SCGS) disease. *Int J Syst Evol Microbiol*
719 2020;71: 004591.
- 720 **44. Quaglino F, Zhao Y, Casati P, Bulgari D, Bianco PA, et al.** 'Candidatus Phytoplasma solani', a
721 novel taxon associated with "stolbur"- and "bois noir"-related diseases of plants. *Int J Syst Evol*
722 *Microbiol* 2013;63:2879-2894.
- 723 **45. Zhao Y, Wei W, Davis RE, Lee I-M, Bottner-Parker KD.** The agent associated with blue dwarf
724 disease in wheat represents a new phytoplasma taxon, 'Candidatus Phytoplasma tritici'. *Int J*
725 *Syst Evol Microbiol* 2021;71:doi: 10.1099/ijsem.0.004604.
- 726 **46. Lee I-M, Martini M, Marccone C, Zhu SF.** Classification of phytoplasma strains in the elm yellows
727 group (16SrV) and proposal of 'Candidatus Phytoplasma ulmi' for the phytoplasma associated
728 with elm yellows. *Int J Syst Evol Microbiol* 2004;54:337-347.
- 729 **47. Jung H-Y, Sawayanagi T, Kakizawa S, Nishigawa H, Wei W, et al.** 'Candidatus Phytoplasma
730 ziziphi', a novel phytoplasma taxon associated with jujube witches' broom disease. *Int J Syst*
731 *Evol Microbiol* 2003;53:1037-1041.
- 732 **48. Lee I-M, Bottner KD, Secor G, Rivera-Varas V.** 'Candidatus Phytoplasma americanum', a
733 phytoplasma associated with a potato purple top wilt disease complex. *Int J Syst Evol Microbiol*
734 2006;56:1593-1597.

- 735 **49. White DT, Blackall LL, Scott PT, Walsh KB.** Phylogenetic positions of phytoplasmas associated
736 with dieback, yellow crinkle and mosaic diseases of papaya, and their proposed inclusion in
737 '*Candidatus Phytoplasma australiense*' and a new taxon, '*Candidatus Phytoplasma australasia*'.
738 *Int J Syst Bacteriol* 1998;48:941-951.
- 739 **50. Šafářová D, Zemánek T, Válová P, Navrátil M.** '*Candidatus Phytoplasma cirsii*', a novel taxon
740 from creeping thistle [*Cirsium arvense* (L.) Scop]. *Int J Syst Evol Microbiol* 2016;66:1745-1753.
- 741 **51. Martini M, Marcone C, Mitrović J, Maixner M, Delić D, et al.** '*Candidatus Phytoplasma*
742 *convolvuli*', a new phytoplasma taxon associated with bindweed yellows in four European
743 countries. *Int J Syst Evol Microbiol* 2012;62:2910-2915.
- 744 **52. Marcone C, Schneider B, Seemüller E.** '*Candidatus Phytoplasma cynodontis*', the phytoplasma
745 associated with Bermuda grass white leaf disease. *Int J Syst Evol Microbiol* 2004;54:1077-1082.
- 746 **53. Valiunas D, Staniulis J, Davis RE.** '*Candidatus Phytoplasma fragariae*', a novel phytoplasma
747 taxon discovered in yellows diseased strawberry, *Fragaria x ananassa*. *Int J Syst Evol Microbiol*
748 2006;56:277-281.
- 749 **54. Griffiths HM, Sinclair WA, Smart CD, Davis RE.** The phytoplasma associated with ash yellows
750 and lilac witches' broom: '*Candidatus Phytoplasma fraxini*'. *Int J Syst Bacteriol* 1999;49:1605-
751 1614.
- 752 **55. Wulff NA, Teixeira DC, Martins EC, Toloy RS, Bianco LF, et al.** Sunn hemp, a major source-plant
753 of the phytoplasma associated with huanglongbing symptoms of sweet orange in São Paulo
754 State, Brazil. *J Citrus Pathol* 2015;1-15.iocv_journalcitruspathology_2695.
- 755 **56. Sawayanagi T, Horikoshi N, Kanehira T, Shinohara M, Bertaccini A, et al.** '*Candidatus*
756 *Phytoplasma japonicum*', a new phytoplasma taxon associated with Japanese *Hydrangea*
757 phyllody. *Int J Syst Bacteriol* 1999;49:1275-1285.
- 758 **57. Davis RE, Harrison NA, Zhao Y, Wei W, Dally EL.** '*Candidatus Phytoplasma hispanicum*', a novel
759 taxon associated with Mexican periwinkle virescence disease of *Catharanthus roseus*. *Int J Syst*
760 *Evol Microbiol* 2016;66:3463-3467.
- 761 **58. Davis RE, Zhao Y, Wei W, Dally EL, Lee I-M.** '*Candidatus Phytoplasma luffae*', a novel taxon
762 associated with witches' broom disease of loofah, *Luffa aegyptica* Mill. *Int J Syst Evol Microbiol*
763 2017;67:3127-3133.
- 764 **59. Nejat N, Vadamalai G, Davis RE, Harrison NA, Sijam K., et al.** '*Candidatus Phytoplasma*
765 *malaysianum*', a novel taxon associated with virescence and phyllody of Madagascar periwinkle
766 (*Catharanthus roseus*). *Int J Syst Evol Microbiol* 2013;63:540-548.

- 767 **60. Miyazaki A, Shigaki T, Koinuma H, Iwabuchi N, Rauka GB, et al.** '*Candidatus* Phytoplasma
768 noviguineense', a novel taxon associated with Bogia coconut syndrome and banana wilt disease
769 on the island of New Guinea. *Int J Syst Evol Microbiol* 2018;68:170-175.
- 770 **61. Malembic-Maher S, Salar P, Filippin L, Carle P, Angelini E, et al.** Genetic diversity of European
771 phytoplasmas of the 16SrV taxonomic group and proposal of '*Candidatus* Phytoplasma rubi'. *Int*
772 *J Syst Evol Microbiol* 2011;61:2129-2134.
- 773 **62. Hiruki C, Wang K.** Clover proliferation phytoplasma: '*Candidatus* Phytoplasma trifolii'. *Int J Syst*
774 *Evol Microbiol* 2004;54:1349-1353.
- 775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797

798 **Table 1.** Strain composition of '*Ca. Phytoplasma*' species at diverse 16S rDNA sequence identity thresholds.
 799 Analysis performed with BLASTn (<https://www.ncbi.nlm.nih.gov>) with the following settings: query coverage
 800 95-100%; percent identity 97.5-100%.

' <i>Ca. Phytoplasma</i> ' species	Min/Max sequence id. (%) vs reference strain	No. of member strains			No. of related strains	
		≥97.5%	≥98%	≥98.65%	≥98%	≥98.65%
' <i>Ca. P. allocasuarinae</i> '	98.52	1	1	0	0	1
' <i>Ca. P. americanum</i> '	99.67/99.87	4	4	4	0	0
' <i>Ca. P. asteris</i> '	97.77/100	374	372	366	2	8
' <i>Ca. P. aurantifolia</i> '	97.50/99.73	293	265	97	28	196
' <i>Ca. P. australasia</i> '	97.51/100	236	229	175	7	61
' <i>Ca. P. australiense</i> '	98.68/99.93	12	12	12	0	0
' <i>Ca. P. balanitae</i> '	99.41/99.80	13	13	13	0	0
' <i>Ca. P. brasiliense</i> '	98.73/99.93	10	10	10	0	0
' <i>Ca. P. caricae</i> '*						
' <i>Ca. P. castanae</i> '*						
' <i>Ca. P. cirsii</i> '	99.93/100	2	2	2	0	0
' <i>Ca. P. cocostanzaniae</i> '	99.33/100	17	17	17	0	0
' <i>Ca. P. convolvuli</i> '	99.93/100	9	9	9	0	0
' <i>Ca. P. costaricanum</i> '	99.15/99.61	17	17	17	0	0
' <i>Ca. P. cynodontis</i> '	98.37/100	36	36	35	0	1
' <i>Ca. P. dypsidiis</i> '	99.83/99.88	6	6	6	0	0
' <i>Ca. P. fragariae</i> '	97.67/99.93	15	14	10	1	5
' <i>Ca. P. fraxini</i> '	97.61/99.93	22	17	9	5	13
' <i>Ca. P. graminis</i> '	98.34/99.74	5	5	4	0	1
' <i>Ca. P. hispanicum</i> '	98.53/99.47	7	7	6	0	1
' <i>Ca. P. japonicum</i> '*						
' <i>Ca. P. luffae</i> '	99.87/99.93	13	13	13	0	0
' <i>Ca. P. lycopersici</i> '*						
' <i>Ca. P. malaysianum</i> '	99.54	1	1	1	0	0
' <i>Ca. P. mali</i> '	99.74/100	20	20	20	0	0
' <i>Ca. P. meliae</i> '	99.45/99.86	5	5	5	0	0
' <i>Ca. P. noviguineense</i> '	99.66/100	26	26	26	0	0
' <i>Ca. P. omanense</i> '	99.58	1	1	1	0	0
' <i>Ca. P. oryzae</i> '*						
' <i>Ca. P. palmae</i> '	98.05/100	80	80	66	0	14
' <i>Ca. P. palmicola</i> '	99.28/100	24	24	24	0	0
' <i>Ca. P. phoenicium</i> '	97.54/99.93	78	73	51	5	27
' <i>Ca. P. pini</i> '	99.74/99.93	3	3	3	0	0
' <i>Ca. P. pruni</i> '	97.87/100	207	205	203	2	4
' <i>Ca. P. prunorum</i> '	99.50/100	35	35	35	0	0
' <i>Ca. P. pyri</i> '	99.14/100	34	34	34	0	0
' <i>Ca. P. rhamnii</i> '	100	1	1	1	0	0
' <i>Ca. P. rubi</i> '	99.35/99.77	7	7	7	0	0
' <i>Ca. P. sacchari</i> '	98.68/99.93	30	30	30	0	0
' <i>Ca. P. solani</i> '	98.17/99.93	73	72	72	1	1

' <i>Ca. P. spartii</i> '*						
' <i>Ca. P. stylosantis</i> '	99.94	1	1	1	0	0
' <i>Ca. P. sudamericanum</i> '*						
' <i>Ca. P. tamaricis</i> '*						
' <i>Ca. P. trifolii</i> '	97.98/100	80	79	76	1	4
' <i>Ca. P. tritici</i> '	100	1	1	1	0	0
' <i>Ca. P. ulmi</i> '	99.53/99.93	14	14	14	0	0
' <i>Ca. P. woydetiae</i> '	98.48	1	1	0	0	1
' <i>Ca. P. ziziphi</i> '	99.54/100	33	33	33	0	0

801 *, only one strain available; in bold '*Ca. Phytoplasma*' in which reassignment is needed to follow
802 the revised guidelines.

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827 **Table 2.** Fourteen '*Ca. Phytoplasma*' species with their full 16S rRNA and five other gene sequences available
828 for both reference and alternative reference strains. GenBank numbers, available acronyms of the
829 phytoplasma strains and nucleotide length, span and, for 16S rRNA gene, the starting nucleotide are
830 reported.

'Candidatus Phytoplasma'	16S rRNA*	<i>tufB</i>	<i>secA</i>	<i>secY</i>	<i>rplV-rpsC</i>	<i>groEL</i>
' <i>Ca. P. asteris</i> '	M30790 (1,542 nt; 1) strain OAY (=MIAY)					
	AP006628 (1,534 nt; 279394-280928 & 555984-557518) strain OY-M	AP006628 (1,185 nt; 305131- 306315)	AP006628 (2,507 nt; 530192- 532699)	AP006628 (1,241 nt 252238- 253479)	AP006628 (1,131 nt; 246003- 247134)	AP006628 (1,610 nt; 142480- 144090)
' <i>Ca. P. aurantifolia</i> '	U15442 (1,513 nt; 18) strain WBDL	NZ_MWKN01 000002 (1,202 nt; 5087- 6289)	NZ_MWKN01 000041 (2,495 nt; 8846- 11341)	NZ_MWKN01 000015 (1,262 nt; 11153- 12415)	NZ_MWKN01 000015 (1,085 nt; 4491-5576)	NZ_MWKN010 00043 (1,640 nt; 14982- 16622)
' <i>Ca. P. australiense</i> '	L76865 (1375 nt; 156) strain AUSGY	JQ824254 (391 nt)				
	AM422018 (1,521; 682142-683674) strain CaPaus	AM422018 (1,185 nt; 656879- 658063)	AM422018 (2,499 nt; 557807- 560305)	AM422018 (1,248 nt; 573941- 575188)	AM422018 (1,104 nt; 580287- 581391)	AM422018 (1,611 nt; 775042- 776652)
' <i>Ca. P. mali</i> '	AJ542541 (1,784 nt; 7) strain AP15					
	CU469464 (1,522 nt; 264144-265665 & 450385-451906) strain AT	CU469464 (1,178 nt; 474914- 476092)	CU469464 (2,393 nt; 134124- 131731)	CU469464 (1,244 nt; 433420- 434664)	CU469464 (1,244 nt; 427172- 428288)	CU469464 (1,610 nt; 237119- 238729)
' <i>Ca. P. meliae</i> '	KU850940 (1528 nt; 6) strain ChTY-Mo3		KU850948 (681 nt)		KU850944 (1259 nt)	
	NZ_JAGVRH010000 003 (1,532 nt; 33957-35489) strain StrPh-Cl	NZ_JAGVRH0 10000003 (1,184 nt; 59600-60784)	NZ_JAGVRH0 10000007 (2,504 nt; 3107-5611)	NZ_JAGVRH01 0000003 (1,235 nt; 10335-11570)	NZ_JAGVRH0 10000003 (1,122 nt; 4148-5270)	NZ_JAGVRH01 0000001 (1,613 nt; 104866- 106479)
' <i>Ca. P. pini</i> '	AJ632155 (1,537nt; 6) strain Pin127S					
	VIAE01000001 (1,518 nt; 19320- 20838) strain MDPP	VIAE0100000 3 (1,227 nt; 13731-12505)	VIAE0100000 5 (2,498 nt; 22267-19769)	VIAE0100000 2 (1,326 nt; 25071-23746)	VIAE0100000 2 (1,141 nt; 31440-30300)	VIAE01000001 (1,623 nt; 45686-44064)
' <i>Ca. P. phoenicium</i> '	AF515636 (1,502 nt; 31) strain A4					
	KM275491 (1,250 nt; 149) strain SA213	KM275492 (1,185 nt)	JPSQ0100003 4, (2,415 nt; 2-2416)	JPSQ0100000 2, (1,311 nt; 11880-13190)	JPSQ0100000 2, (1,102 nt; 18377-19478)	JPSQ01000038, (1,632 nt; 3858-5489)
' <i>Ca. P. pruni</i> '	JQ044397 (1,527 nt; 14) strain CX-95	LHCF0100000 2 (1,184 nt; 8092-9276)	LHCF0100000 2 (2,507 nt; 49619-52126)	JQ268249 (1,263 nt)	JQ360955 (1,239 nt)	LHCF01000008 (164 nt; 22632- 22796)

' <i>Ca. P. pyri</i> '	AJ542543 (1,516 nt; 7) strain PD1	MZ507700 (1,179 nt)	MZ507699 (2,394 nt)	GU004363, (269-1,489)	EF193370, (1,117 nt)	MZ507698 (1,614 nt)
' <i>Ca. P. sacchari</i> '	MN889545 (1,516 nt; 7) strain SCGS	VWXM01000002 (1,220 nt; 9425-10645)	VWXM01000004; (2,510 nt; 3173-5683)	VWXM01000002; (1,276 nt; 26714-27990)	VWXM01000002; (1,044 nt; 20387-21431)	VWXM01000012; (1,105 nt; 1-1105)
' <i>Ca. P. solani</i> '	AF248959 (1,527 nt; 6) strain STOL	JQ797670 (946 nt)		JQ797668 (1,224 nt)	JQ797662 (1,093 nt)	
	JQ730740 (1,491 nt; 44) strain 284/09	FO393427 (1,185 nt; 515449-516633)	FO393427 (2,499 nt; 488961-491459)	FO393427 (1,224 nt; 162270-163493)	FO393427 (1,093 nt; 156005-157097)	FO393427 (1,605 nt; 311260-312864)
' <i>Ca. P. tritici</i> '	DQ078304 (1,432 nt; 53) strain WBD	AVAO01000003 (1,184 nt; c230832-232016)	AVAO01000003 (2,501 nt; 112769-115276)	AVAO01000003 (1,241 nt; c282126-283367)	AVAO01000003 (691 nt; 289222-289913)	AVAO01000001 (1,610 nt; 47299-48909)
' <i>Ca. P. ulmi</i> '	AY197655 (1527 nt; 6) strain EY1	FN561879 (925 bp)	KJ462034 (559 bp)	AY197690 (1,433 nt)	AY197675 (1,199 nt)	
	OU413475 (1,527 nt; 65) strain ULW	MZ507705 (1,164 nt)	MZ507703 (2,517 nt)	MZ507704 (1,260 nt)	JN851866 (1,216 nt)	MT418907 (1,617 nt)
' <i>Ca. P. ziziphi</i> '	AB052876 (1,367 nt; 10) strain JWB-G1					
	CP025121 (1,531 nt; 423854-425385 & 597926-599457) strain JWB-nky	CP025121 (1,163 nt; 478755-479918)	CP025121 (2,513 nt; 365998-368511)	CP025121 (1,247 nt; 4888-6135)	CP025121 (1,110 nt; 11450-12645) **	CP025121 (1,616nt; 106623-108239)

831 *, numbers in bold refer to the starting position of the sequence relative to the 16S rRNA sequence
832 deposited under accession number M30790; **sequences inverted, *rpsC* gene is followed by *rplV* gene.

833
834
835
836
837
838
839
840
841
842
843
844
845
846
847

848 **Table 3.** Sixteen '*Ca. Phytoplasma*' species with their full 16S rRNA and available other gene sequences for
 849 reference or alternative reference strains. GenBank numbers, available acronyms of the phytoplasma strains
 850 and nucleotide length, span and, for 16S rRNA gene, the starting nucleotide are reported.

'Candidatus Phytoplasma'	16S rRNA*	<i>tufB</i>	<i>secA</i>	<i>secY</i>	<i>rplV-rpsC</i>	<i>groEL</i>
' <i>Ca. P. americanum</i> '	DQ174122 (1,503 nt; 31) strain PPT12-NE					
	MN227133 (1,477 nt) strain SRL1-PA	MN22713 5 (445 nt)	MN227136 (795 nt)	MN227134 (1,450 nt)		
' <i>Ca. P. australasia</i> '	Y10097 (1,521 nt; 11) strain PpYC					
	JQ868448 (1,505 nt; 48) strain TBB	JQ824250 (385 nt)	EU168729 (482 nt)		EF193373 (1,294 nt)	
' <i>Ca. P. cirsi</i> '	KR869146 (1,498 nt; 31) strain CirYS		KU557489 (462 nt)			
' <i>Ca. P. convolvuli</i> '	JN833705 (1,496; 31) strain 57/11	OK127877 (898 nt)			OK127878 (1,417 nt)	
' <i>Ca. P. cynodontis</i> '	AJ550984 (1,499 nt; 29) strain BGWL-C1					
	KP019340 (1,499 nt; 31) strain 305/13					KP019342 (1,528 nt)
' <i>Ca. P. fragariae</i> '	HM104662 (1,502 nt; 31) strain Straw					
	MK501641 (1,531 nt; 20) strain GBFC_SY_01				MN914137 (1,122 nt)	
' <i>Ca. P. fraxini</i> '	AF092209 (1,462 nt, 36) strain AshY1			GU004329 (1,250 nt)**		KJ939978 (552 nt)
' <i>Ca. P. japonicum</i> '	AB010425 (1,521 nt; 10) strain JHP			AB738739 (1237 bp)		AB746432 (1611 bp)
' <i>Ca. P. hispanicum</i> '	AF248960 (1,527 nt; 6) strain MPV		EU168753 (482 nt)	GU004336 (1,235 nt)	EF193365 (1,126 nt)	KT444668 (552 nt)
' <i>Ca. P. luffae</i> '	AF248956 rRNAa AF353090 rRNAb (1,464 nt; 69) strain LfWB	AF086617 *** (1,188 nt)		GU004319 (1,257 nt)		
' <i>Ca. P. malaysianum</i> '	EU371934 (1,523 nt; 6) strain MaPV		FJ755005 (482 nt)			
' <i>Ca. P. noviguineense</i> '	LC228755 (1,480 nt; 55) strain BCS-Bo ^R			LC228769 (1,247 nt; 337-1584)	LC228762 (1,439 nt)	
' <i>Ca. P. prunorum</i> '	AJ542545 (1,516 nt; 7) strain ESFY-G2					
	JQ868450 (1,494 nt; 28) strain LNp	JQ824235 (385 nt)			x	
' <i>Ca. P. rhamn</i> '	X76431 (1,473 nt; 47) strain BWB				KF498659 (1,073 nt)	
	JQ868449 (1,494 nt; 28) strain RhCa	JQ824207 (391 nt)	KJ462067 (559 nt)			
' <i>Ca. P. rubi</i> '	AY197648 (1,529 nt; 6) strain RuS	FN561887 (925 nt)	KJ462043 (596 nt)	AY197696 (1,412 nt)	FN562164 (797 nt)	
' <i>Ca. P. trifolii</i> '	AY390261 (1,531 nt; 6) strain CP		KJ462045 (559 nt)	GU004315 (1,262 nt)	AY197668, (1,154 nt)	

851 *, numbers in bold refer to the starting position of the sequence relative to the 16S rRNA sequence
 852 deposited under accession number M30790; **Strain designation AshY=AshY1; ***, from Taiwan
 853
 854
 855
 856
 857 **Table 4.** New '*Candidatus* Phytoplasma' species officially proposed (names from IRPCM, 2004). GenBank
 858 numbers, available acronyms of the phytoplasma strains and nucleotide length, span and, for 16S rRNA gene,
 859 the starting nucleotide are reported.

Phytoplasma	Full 16S rDNA*	<i>tufB</i>	<i>secA</i>	<i>secY</i>	<i>rplV-rpsC</i>
' <i>Ca. P. cocostanzaniae</i> '	X80117 (1,524 nt; 5) strain LD				
' <i>Ca. P. palmae</i> '	U18747 (1,524 nt; 9) no name				
	VBRA02000009 (1,544 nt; 3425- 4968) strain ACPD	VBRA02000007 (1,191 nt; 23653-24843)	VBRA02000000 9 (2,418 nt; 171796- 174213)	VBRA02000009, (1,254 nt; 29596- 30849)	VBRA02000009 (1,054 nt; 35998- 37051)

860 *, numbers in bold refer to the starting position of the sequence relative to the 16S rRNA sequence
 861 deposited under accession number M30790.
 862
 863
 864
 865
 866
 867
 868
 869