Comparative Genome Analysis of '*Candidatus* Phytoplasma luffae' Reveals the Influential Roles of Potential Mobile Units in Phytoplasma Evolution

Running Title: PMU and Phytoplasma Genome Evolution

Ching-Ting Huang[1], Shu-Ting Cho[1], Yu-Chen Lin[1], Choon-Meng Tan[2], Yi-Ching Chiu[2,3], Jun-Yi Yang[2,3,4]*, Chih-Horng Kuo[1,3,5.6]*

[1] Institute of Plant and Microbial Biology, Academia Sinica, Taipei 115, Taiwan

[2] Graduate Institute of Biochemistry, National Chung Hsing University, Taichung 402, Taiwan

[3] PhD Program in Microbial Genomics, National Chung Hsing University and Academia Sinica, Taichung 402, Taiwan

[4] Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan

[5] Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, National Chung-Hsing University and Academia Sinica, Taipei 115, Taiwan

[6] Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan

* Correspondence

Jun-Yi Yang: jyang@nchu.edu.tw

Chih-Horng Kuo: chk@gate.sinica.edu.tw

Abstract

Phytoplasmas are insect-transmitted plant pathogens that cause substantial losses in agriculture. In addition to economic impact, phytoplasmas induce distinct disease symptoms in infected plants, thus attracting attention for research on molecular plant-microbe interactions and plant developmental processes. Due to the difficulty of establishing an axenic culture of these bacteria, culture-independent genome characterization is a crucial tool for phytoplasma research. However, phytoplasma genomes have strong nucleotide composition biases and are repetitive, which make it challenging to produce complete assemblies. In this study, we utilized Illumina and Oxford Nanopore sequencing technologies to obtain the complete genome sequence of '*Candidatus* Phytoplasma luffae' strain NCHU2019 that is associated with witches' broom disease of loofah (*Luffa aegyptiaca*) in Taiwan. The fully assembled circular chromosome is 769 kb in size and is the first representative genome sequence of group 16SrVIII phytoplasmas. Comparative analysis with other phytoplasmas revealed that NCHU2019 has an exceptionally repetitive genome, possessing a pair of 75 kb repeats and at least 13 potential mobile units (PMUs) that account for ~25% of its chromosome. This level of genome repetitiveness is exceptional for bacteria, particularly among obligate pathogens with reduced genomes. Our genus-level analysis of PMUs demonstrated that these phytoplasma-specific mobile genetic elements can be classified into three major types that differ in gene organization and phylogenetic distribution. Notably, PMU abundance explains nearly 80% of the variance in phytoplasma genome sizes, a finding that provides a quantitative estimate for the importance of PMUs in phytoplasma genome variability. Finally, our investigation found that in addition to horizontal gene transfer, PMUs also contribute to intra-genomic duplications of effector genes, which may provide redundancy for neofunctionalization or subfunctionalization.

Taken together, this work improves the taxon sampling for phytoplasma genome research and provides novel information regarding the roles of mobile genetic elements in phytoplasma evolution.

Keywords

Plant pathogen, phytoplasma, genomics, molecular evolution, mobile genetic element, effector

56 **Introduction**

57 Phytoplasmas are uncultivated bacteria associated with plant diseases in several
58 hundred species (Lee et al., 2000; Hogenhout et al., 2008; Bertaccini et al., 2014; Namba,
59 2019). In infected plants, phytoplasma cells are restricted to phloem tissues and can secrete
60 effector proteins that cause developmental abnormalities of the hosts (Sugio et al., 2011b).
61 Typical symptoms of phytoplasma infections include stunting, dwarfism, virescence (i.e.,
62 greening of flowers), phyllody (i.e., abnormal development of floral parts into leaf-like
63 tissues), and witches' broom (i.e., proliferation of stems and leaves), which result in
64 substantial agricultural losses.

65 For classification of these uncultivated bacteria, a system based on restriction
66 fragment length polymorphism (RFLP) analysis of their 16S rRNA genes was developed in
67 the 1990s (Lee et al., 1993, 1998; Gundersen and Lee, 1996) and at least 33 16S rRNA gene
68 RFLP (16Sr) groups have been described (Zhao et al., 2009; Zhao and Davis, 2016). Later, a
69 provisional genus-level taxon '*Candidatus* Phytoplasma' was proposed to accommodate
70 these bacteria (The IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma
71 taxonomy group, 2004) and at least 41 '*Ca*. P.' species have been described or proposed
72 (Bertaccini and Lee, 2018). Based on analysis of 16S rRNA and other conserved genes,
73 phytoplasmas are divided into three major phylogenetic clusters (Hogenhout and Seruga
74 Music, 2009; Chung et al., 2013; Seruga Music et al., 2019). Early genomics studies were
75 mainly conducted for clusters I (i.e., '*Ca*. P. asteris' of group 16SrI and '*Ca*. P. australiense' of
76 16SrXII) (Oshima et al., 2004; Bai et al., 2006; Tran-Nguyen et al., 2008) and II (i.e., '*Ca*. P.
77 mali' of 16SrX) (Kube et al., 2008). In comparison, cluster III harbors the highest level of
78 diversity, yet has received limited attention for comparative genomics studies (Chung et al.,
79 2013; Wang et al., 2018a).

80 To improve our understanding of phytoplasma genome diversity, we conducted
81 whole genome sequencing of a '*Ca*. P. luffae' strain collected in Taiwan. The species-level
82 taxon '*Ca*. P luffae' belongs to group 16SrVIII in cluster III and is associated with witches'
83 broom disease of loofah (*Luffa aegyptiaca*) (Davis et al., 2017). The availability of a complete
84 genome sequence from this taxon provides a complete view of its gene content, which can
85 facilitate the study of its pathogenesis mechanisms and other aspects of its biology. More
86 importantly, with the increased availability of genome sequences from diverse
87 phytoplasmas (**Table 1**), we performed genus-level comparative analysis to obtain a more
88 comprehensive picture of their genomic diversity. This improves upon previous works that
89 are limited to comparisons of closely related taxa or have sparse sampling (Bai et al., 2006;
90 Tran-Nguyen et al., 2008; Kube et al., 2012; Saccardo et al., 2012; Andersen et al., 2013;
91 Chung et al., 2013; Orlovskis et al., 2017; Wang et al., 2018a; Cho et al., 2019, 2020a; Seruga
92 Music et al., 2019; Kirdat et al., 2021; Zhao et al., 2021). Furthermore, our focused analysis
93 of the potential mobile units (PMUs) (Bai et al., 2006) revealed the influential roles of these
94 mobile genetic elements in the evolution of phytoplasma genome organization and effector
95 gene content.
96
97 **Materials and Methods**
98 Biological Materials
99 The strain NCHU2019 was collected from a naturally infected loofah plant found in
100 Dacheng Township (Changhua County, Taiwan; 23.861860 N, 120.291322 E) on July 4th,
101 2019 (**Figure 1A**). After initial collection, the bacterium was transferred to lab-grown loofah
102 plants (cultivar A-Jun, Known-You Seed Co., Kaohsiung, Taiwan) through grafting and
103 maintained in a plant growth facility in the National Chung Hsing University (Taichung,

104 Taiwan) (**Figure 1B**). To confirm the presence and identity of this phytoplasma strain, a
105 partial sequence of the rRNA operon was PCR amplified using the phytoplasma-specific
106 primer set P1/P7 for Sanger sequencing as described (Chung et al., 2013).
107
108 Genome Sequencing, Assembly, and Annotation
109 The procedures for genome sequencing and analysis were based on those described
110 in our previous work on phytoplasma genomes (Chung et al., 2013; Cho et al., 2019, 2020a,
111 2020b). All kits were used according to the manufacturer's protocols and all bioinformatics
112 tools were used with the default settings unless stated otherwise.
113 For whole genome shotgun sequencing, leaves from one artificially infected plant
114 exhibiting typical symptoms (i.e., small leaves and witches' broom) (**Figure 1B**) were
115 collected for total genomic DNA extraction using the Wizard Genomic DNA Purification Kit
116 (A1120; Promega, USA). For Illumina sequencing, the DNA sample was processed using the
117 KAPA Library Preparation Kit (KK8234) and the Invitrogen SizeSelect Gels (G6610-02) to
118 obtain ~550 bp fragments, followed by MiSeq 2x300 bp paired-end sequencing (v3
119 chemistry). For Oxford Nanopore Technologies (ONT) sequencing, the library was prepared
120 using the ONT Ligation Kit (SQK-LSK109) without shearing or size selection, followed by
121 MinION sequencing (R9.4.1 chemistry) and Guppy v3.3.3 base-calling.
122 The *de novo* genome assembly involved two stages. In the first stage, only the
123 Illumina reads were used for running Velvet v1.2.10 (Zerbino and Birney, 2008) with k-mer
124 length set to 151 and minimum contig length set to 2,000 bp. To identify putative
125 phytoplasma contigs, all contigs were was queried against the NCBI non-redundant protein
126 database (Benson et al., 2018) using BLASTX v2.10.0 (Camacho et al., 2009). Those have a hit
127 with an e-value of < 1e-15 to phytoplasma proteins were selected for manual inspection.

128 False positive contigs derived from plant chloroplast and mitochondrial genomes were
129 identified based on the difference in sequencing depth and verified based on BLASTN
130 v2.10.0 (Camacho et al., 2009) searches against the NCBI standard nucleotide collection
131 database (Benson et al., 2018).

132 To start the second stage of assembly, the putative phytoplasma contigs identified
133 from the first stage were used as the reference for extracting phytoplasma reads from the
134 Illumina data set using BWA v0.7.17 (Li and Durbin, 2009) with an alignment score cutoff of
135 30 and from the ONT data set using Minimap2 v2.15 (Li, 2018) with an alignment score
136 cutoff of 1,000. The extracted reads were processed together for *de novo* hybrid assembly
137 using Unicycler v0.4.9b. The contigs were validated as belonging to the phytoplasma
138 genome by BLAST searches according to the process utilized in the first stage and used as
139 the starting point for an iterative process of assembly improvement. In each iteration, all
140 Illumina and ONT raw reads were mapped to the draft assembly as described. The mapping
141 results were programmatically checked using the “mpileup” function of SAMTOOLS v1.9 (Li
142 et al., 2009) and manually inspected using IGV v2.5.0 (Robinson et al., 2011) to identify
143 possible assembly errors. Regions with raw reads mapping results exhibiting abnormalities
144 were cut and re-arranged manually based on the continuity of ONT long reads, then
145 validated using the mapping results in the next iteration. During the early iterations, the
146 mapping results of ONT reads were used to provide scaffolding information and validate the
147 overall organization of the circular chromosome, particularly the junctions between
148 repetitive and unique regions. Additionally, the reads mapped to contig ends were visually
149 inspected using IGV for manual selection of representative reads that can extend contigs
150 and fill gaps. During the later iterations, the mapping results of Illumina reads were used to
151 validate bp-scale indels and possible sequencing errors introduced by the ONT reads. The

152 process was repeated until the complete genome assembly was obtained and all regions are
153 supported by the raw reads mapping results. Additionally, the "depth" function of
154 SAMTOOLS is used to calculate the sequencing coverage.

155 To provide a genome size estimate based on k-mer distribution, all Illumina reads
156 mapped to the finalized assembly with an alignment score above 200 were extracted. Based
157 on these reads, occurrences of k-mers in the size range between 17 and 63 were calculated
158 by using jellyfish v2.2.8 (Marçais and Kingsford, 2011). The genome size was estimated by
159 dividing the total k-mer count with the peak depth as suggested previously (Lu et al., 2016).

160 The procedure of gene prediction was performed using RNAmmer v1.2 (Lagesen et
161 al., 2007), tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997), and Prodigal v2.6.3 (Hyatt et al.,
162 2010). The annotation was based on the homologous gene clusters present in other
163 phytoplasma genomes (**Table 1**) as identified by BLASTP v 2.10.0 (Camacho et al., 2009) and
164 OrthoMCL v1.3 (Li et al., 2003), followed by manual curation based on information obtained
165 from GenBank (Benson et al., 2018), KEGG (Kanehisa et al., 2010), and COG (Tatusov et al.,
166 2003) databases. Additionally, putative secreted proteins were predicted using SignalP v5.0
167 (Armenteros et al., 2019) based on the Gram-positive bacteria model. Those candidates
168 with transmembrane domains as identified by TMHMM v2.0 (Krogh et al., 2001) were
169 removed and the remaining ones were required to have a signal peptide length in the range
170 of 21 to 52 amino acids. For visualization, the Circos v0.69-6 (Krzywinski et al., 2009) was
171 used to draw the genome map.

172

173 Comparative Analysis

174 For comparative analysis with other representative phytoplasma genomes (**Table 1**),
175 homologous gene clusters were identified using OrthoMCL (Li et al., 2003). Multiple

176 sequence alignments of homologous genes were prepared using MUSCLE v3.8.31 (Edgar,
177 2004) and visualized using JalView v2.11 (Waterhouse et al., 2009). Maximum likelihood
178 phylogenies were inferred using PhyML v3.3 (Guindon and Gascuel, 2003) with the default
179 LG substitution model and visualized using FigTree v1.4.4. PHYLIP v3.697 (Felsenstein, 1989)
180 was used for bootstrap analysis. Classification of the 16S RFLP group was performed using
181 *i*PhyClassifier (Zhao et al., 2009).

182 For whole-genome comparison, FastANI v1.1 (Jain et al., 2018) was used to calculate
183 the proportion of genomic segments mapped and the average nucleotide identity (ANI). For
184 pairwise genome alignments, Mummer v3.23 (Kurtz et al., 2004) was used with the options
185 "--maxmatch --mincluster 30" and the results were visualized using genoPlotR v0.8.9 (Guy et
186 al., 2010).

187 The PMU analysis was based on the eight core genes (i.e., *tra5*, *dnaB*, *dnaG*, *tmk*,
188 *hflB*, *himA*, *ssb*, and *rpoD*) defined previously (Bai et al., 2006). To ensure uniform
189 annotation and to include possible pseudogenes, all genome assemblies were examined by
190 using available PMU genes sequences as queries to run TBLASTN searches. All statistical
191 tests were performed in the R statistical environment (R Core Team, 2019); correlation
192 coefficients were calculated using the 'cor.test' function, linear regression was performed
193 using the 'lm' function and visualized using the 'plot' and 'abline' functions.

194

195 **Results and Discussion**

196 Genomic Characterization of NCHU2019

197 The shotgun sequencing generated ~4.7 Gb of Illumina raw reads and ~3.5 Gb of
198 ONT raw reads. In the first stage of *de novo* assembly with only the Illumina reads, a draft
199 assembly with 49 contigs totaling 450,754 bp was obtained (longest contig: 31,507 bp;

200 shortest contig: 2,001 bp; N50: 12,506 bp). Based on the mapping results to this first draft
201 assembly, we extracted 0.9% (136,116 out of 15,655,220) of all Illumina raw reads and 1.7%
202 (9,416 out of 539,354) of all ONT raw reads. These reads correspond to a sequencing depth
203 of 71- and 110-fold, respectively. In the second stage of *de novo* hybrid assembly using
204 these extracted reads from both sequencing platforms, a second draft assembly with 10
205 contigs totaling 668,430 bp was obtained. These 10 contigs include two long ones that are
206 552,199 and 103,217 bp in size, as well as eight short ones that are shorter than 3,000 bp.
207 Examination of the mapping results to this second draft assembly found 186,979 Illumina
208 reads with 73-fold coverage and 13,242 ONT reads with 95-fold coverage.
209 During the iterative assembly improvement process, a pair of duplicated
210 chromosomal segments that are each ~75 kb in size and multiple smaller repeats were
211 identified. These regions were manually corrected to generate the finalized assembly. All
212 junctions involving repetitive regions were verified based on visual inspection of the ONT
213 long read mapping results. For the finalized assembly, we obtained a circular contig that
214 corresponds to the phytoplasma chromosome; no plasmid was found. This circular contig is
215 769,143 bp in size with 23.3% G+C content (**Figure 2**). For the final round of raw reads
216 mapping examination, 149,857 Illumina reads and 13,649 ONT reads were mapped to this
217 circular contig, corresponding to 51- and 80-fold sequencing depth, respectively. The
218 decrease of sequencing depth compared to the draft assemblies is likely explained by the
219 repetitive regions being resolved. As a support for this inference, the sequencing depth
220 exhibits a nearly uniform distribution across the entire finalized circular contig, with
221 repetitive regions having similar depth as non-repetitive regions (**Figure S1**).
222 For genome size estimation based on k-mer analysis, we utilized k = 21 as a
223 representative data set for visualization (**Figure S2A**). We found that while the peak depth is

224 51, the frequency distribution is nearly flat in the range between ~40-55. Based on this
225 distribution, the genome size is estimated to be in the range between 660 and 842 kb
226 (**Figure S2B**). The assembled genome size of 769 kb is near the middle point (i.e., 751 kb) of
227 these estimates.

228 The annotation of this phytoplasma genome contains two complete sets of 16S-23S-
229 5S rRNA gene clusters, 31 tRNA genes, 725 coding sequences (CDSs), and 13 pseudogenes.
230 Both copies of the 16S rRNA gene are 100% identical to the reference sequence of '*Ca*. P.
231 luffae' LfWB (GenBank accession AF248956). Among the CDSs, 317 (44%) lacked any COG
232 functional category assignments. Among the CDSs that were assigned to specific functional
233 categories, those assigned to information storage and processing (e.g., replication,
234 transcription, and translation) represent the largest group that account for 32% of all CDSs.
235 In comparison, those assigned to cell process and metabolism account for 9% and 13%,
236 respectively. These observations are consistent with findings from characterization of other
237 phytoplasmas and more distantly related Mollicutes (e.g., *Spiroplasma*, *Entomoplasma*, and
238 *Mycoplasma*) (Chen et al., 2012; Kube et al., 2012; Lo et al., 2018). The observation that a
239 large fraction of genes lack specific functional annotation may be attributed to the elevated
240 evolutionary rates of Mollicutes and their distant phylogenetic relationships from model
241 organisms (Wu and Eisen, 2008; Wu et al., 2009; McCutcheon and Moran, 2012). The
242 observation that genes related to information storage and processing genes are relatively
243 abundant compared to those involved in metabolism is common among symbiotic bacteria
244 with small genomes (McCutcheon and Moran, 2012; Lo et al., 2016).

245 Other than the low G+C content and reduced gene content, phytoplasma genomes
246 are generally known to be repetitive, partly due to the presence of PMUs (Sugio and
247 Hogenhout, 2012). Interestingly, the genome of this strain is far more repetitive than other

248 phytoplasmas that have been characterized. On average, PMU core genes account for ~4.7%
249 (Std. Dev. = 2.7%) of the genome size among those 19 representative phytoplasma strains
250 analyzed (**Table 1**). For strain NCHU2019, there are 117 PMU core genes that are organized
251 into at least 13 distinct PMU regions (**Figure 2**) and account for 11% of the chromosome
252 length. Additionally, a pair of chromosomal segments, each is ~75 kb in size, were found to
253 be duplicated (positions: 315,975-391,140 and 391,446-466,612). Together, these two
254 duplicated segments and the 13 PMU regions account for 25% of the chromosome length.
255 Explanation for the high genome repetitiveness of strain NCHU2019 compared to other
256 phytoplasmas is unclear.
257
258 Comparisons with Other Phytoplasmas
259 Based on the established classification scheme and a previous study of 16S rRNA
260 gene phylogeny (Davis et al., 2017), '*Ca*. P. luffae' belongs to group 16SrVIII and is most
261 closely related to '*Ca*. P. malaysianum' (group 16SrXXXII; GenBank accession EU371934)
262 (Nejat et al., 2013). These two species-level taxa have 95.9% sequence identity (i.e.,
263 1,463/1,526 aligned nucleotides) in their 16S rRNA genes. However, no genome sequence is
264 available for '*Ca*. P. malaysianum' for comparative analysis.
265 Other than '*Ca*. P. malaysianum', the next closest relatives of '*Ca*. P. luffae' include
266 phytoplasmas belonging to groups 16SrV ('*Ca*. P. vitis' and '*Ca*. P. ziziphi'), 16SrVI ('*Ca*. P.
267 sudamericanum' and '*Ca*. P. trifolii'), and 16SrVII ('*Ca*. P. fraxini'). Among these, one
268 complete genome sequence (GenBank accession CP025121) is available for '*Ca*. P. ziziphi'
269 strain Jwb-nky (Wang et al., 2018a), which represents the most closely related lineage for
270 comparative analysis (**Figure 3A**). Comparison based on the 16S rRNA gene sequences
271 indicated that '*Ca*. P. luffae' NCHU2019 and '*Ca*. P. ziziphi' Jwb-nky have 94.9% sequence

272 identity (i.e., 1,447/1,524 aligned nucleotides). For whole-genome comparison, only 49% of
273 the chromosomal segments can be mapped between these two strains and these segments
274 have < 80% ANI. Pairwise genome alignment indicated that the most of the conserved
275 regions correspond to PMUs (**Figure 4**), which further supports that the sequence
276 divergence between these two genomes is too high for nucleotide-level comparisons. The
277 lack of chromosome-level synteny conservation was commonly reported in previous
278 comparisons between complete genome sequences of phytoplasmas (Bai et al., 2006; Tran-
279 Nguyen et al., 2008; Andersen et al., 2013; Orlovskis et al., 2017; Wang et al., 2018a), even
280 for '*Ca*. P. asteris' strains sharing > 99.9% 16S rRNA gene sequence identity and > 98.1%
281 genome-wide ANI (Cho et al., 2020a). These observations may be attributed to the strong
282 nucleotide composition biases, the high mutation accumulation rates, and the influence of
283 PMUs (Cho et al., 2020a).
284 Considering the high level of nucleotide sequence divergence between '*Ca*. P. luffae'
285 NCHU2019 and '*Ca*. P. ziziphi' Jwb-nky, we also examined gene content based on protein
286 sequence comparisons. In the pairwise comparison, 344 homologous gene clusters are
287 shared between these two strains, while 175 and 241 are specific to NCHU2019 and Jwb-
288 nky, respectively. These counts of shared and strain-specific genes are comparable to those
289 found in previous studies of phytoplasma genome comparisons at between-species level
290 (Chung et al., 2013; Tan et al., 2021). Although relatively large numbers of strain-specific
291 genes were identified, > 80% of these genes lack functional annotation, making the
292 inference of their roles difficult. As such, the gene content regarding metabolic capacity and
293 transporters of '*Ca*. P. luffae' NCHU2019 is expected to be highly similar to that of '*Ca*. P.
294 ziziphi' Jwb-nky, which was described in detail previously (Wang et al., 2018a).

295 For comparisons with those more divergent phytoplasmas with genomic information
296 available (**Table 1**), only 134 homologous gene clusters are conserved among the 19
297 representatives analyzed. This estimate of phytoplasma core genome is much lower than
298 the ~200 genes consistently reported in previous studies (Chen et al., 2012; Chung et al.,
299 2013; Tan et al., 2021), an observation that is likely explained by the inclusion of several
300 draft assemblies in this study. Detailed description regarding the functions of these ~200
301 core genes were reported previously (Chen et al., 2012; Kube et al., 2012; Chung et al.,
302 2013).
303
304 Detailed Characterization of PMUs
305 To better understand the roles of PMUs in phytoplasma genome evolution, we
306 conducted detailed characterization of these mobile genetic elements. Among the 13 PMUs
307 identified in '*Ca*. P. luffae' NCHU2019 (**Figure 5A**), 11 are considered as complete ones and
308 range from 14 to 18 kb in size. The remaining two (i.e., #6 and #7) are both 12 kb in size and
309 appear to be truncated based in the lack of multiple PMU core genes (i.e., *tmk*, *dnaB*, *dnaG*,
310 and *tra5*). These 13 PMUs are variable in the gene content in between *rad50* and *tmk*, while
311 the sequences of shared core genes are highly conserved. For example, the *tmk* and *dnaB*
312 homologs among these PMUs have identical sequences (**Figure 6**). It is unclear if the
313 sequence conservation of these core genes is explained by lack of mutation accumulation,
314 purifying selection, or assembly artifacts. Regarding chromosomal locations, these PMUs are
315 interspersed across the entire chromosome and there is no obvious pattern of clustering
316 (**Figures 2 and 4**). Two sets of PMUs (i.e., #4-6 and #7-9) are associated with the two 75-kb
317 repeat regions. Notably, one set of junctions for these large repeats (chromosomal positions
318 315,975 and 466,612) are located inside PMU #4 (positions 308,032..323,983) and PMU #9

319 (positions 458,604..475,923). This finding suggests that homologous recombination
320 facilitated by PMUs may have facilitated the segmental duplication of this chromosome,
321 which in turn further increased the PMU copy number.

322 Based on initial characterization of the PMUs in '*Ca*. P. asteris' AYWB genome, eight
323 core genes were identified (Bai et al., 2006). As more genome sequences become available
324 from more diverse phytoplasmas, we are able to include representatives of 13 '*Ca*. P.'
325 species from all three phylogenetic clusters in this analysis (**Table 1**). Compared to other
326 phytoplasmas, '*Ca*. P. luffae' NCHU2019 has the highest number of PMUs and is distinctive
327 in that all of its PMUs are similar (**Figure 5A**). Among the 11 complete PMUs in this genome,
328 the eight PMU core genes and *rad50* are all organized in the same order. The minor
329 variation in gene organization is mainly in between *rad50* and *tmk*, where genes encode
330 different putative secreted proteins and hypothetical proteins are found. For comparison, in
331 the closely related '*Ca*. P. ziziphi' Jwb-nky of cluster III (i.e., the yellow clade in **Figure 3A**),
332 the four PMUs exhibit much higher levels of diversity in gene organization between *hflB* and
333 *tmk* (**Figure 5A**). For the distantly related '*Ca*. P. asteris' AYWB and '*Ca*. P. australiense' PAa
334 of cluster I (i.e., the blue clade in **Figure 3A**), high levels of intra-genomic PMU diversity are
335 also observed (**Figure 5B**).

336 Based on the presence/absence pattern and order of eight PMU core genes defined
337 previously (Bai et al., 2006), we found that the PMUs in these representative phytoplasmas
338 with complete genome sequences available can be classified into three major types (**Figure**
339 **5B**). The strains omitted in the visualization all have close relatives belonging to the same
340 16Sr group and share similar PMUs. Among the three major types, type A PMUs, in which
341 *tmk* is upstream of *dnaB*, are the most common ones that include PMUs found in
342 phylogenetic clusters I (e.g., '*Ca*. P. asteris' and '*Ca*. P. australiense') and III (e.g., '*Ca*. P.

343 luffae', '*Ca*. P. ziziphi', and '*Ca*. P. aurantifolia'). Type B PMUs are shorter, have *tmk*
344 downstream of *dnaB*, and are found only in '*Ca*. P. australiense' that belongs to cluster I.
345 Type C is the rarest type, with only one representative found in '*Ca*. P. mali' that belongs to
346 cluster II, and has *hflB* and *dnaG* located in between *tmk* and *dnaB*. In addition to the gene
347 organization, molecular phylogenies of *dnaB* and *tmk* also revealed divergence among
348 homologs from different PMU types (**Figure 6**). These patterns provide further support for
349 our PMU classification scheme. It is interesting to note that regardless of PMU copy
350 numbers, most of the phytoplasmas with genome sequences available harbor only one
351 single type of PMUs. '*Ca*. P. australiense' is the only exception that harbors both type A and
352 B PMUs. Future improvements in sampling more diverse phytoplasma genomes, particularly
353 cluster II lineages, are necessary to provide a more comprehensive understanding of PMU
354 diversity.
355
356 PMU and Phytoplasma Genome Size Variation
357 One notable observation regarding phytoplasma genomes is the extensive size
358 variation at both between- and within-species levels (**Table 1**). Phylogenetic relatedness, as
359 quantified by core genome sequence divergence, does not appear to provide a reliable
360 predictor for genome sizes (**Figure 3A**). Previous within-species comparisons suggested that
361 PMU abundance is an important factor (Bai et al., 2006; Andersen et al., 2013). To further
362 test if this pattern holds true for genus-level analysis, we performed regression analysis to
363 examine the correlation between the combined length of PMU core genes and genome size.
364 Strikingly, when all 19 representative phytoplasma genome assemblies were examined
365 together, the combined length of PMU core genes explains 79% of the variance in genome
366 sizes (r = 0.89, p = 3.7e-07) (**Figure 7**). Due to the concern that draft assemblies cannot

367 provide accurate information regarding these two metrics, we also performed regression
368 analysis with only the 10 complete assemblies and obtained similar results (r = 0.87, p =
369 9.7e-04).
370 The evolution of bacterial genome sizes is a topic that received much research
371 attention and previous studies have identified multiple relevant factors, such as
372 composition of gene content, prevalence of mobile genetic elements, effective population
373 sizes, level of mutational biases toward deletion, and ecological niches (Mira et al., 2001;
374 Konstantinidis and Tiedje, 2004; Ochman and Davalos, 2006; Kuo and Ochman, 2009; Kuo et
375 al., 2009; Novichkov et al., 2009; McCutcheon and Moran, 2012; Lo et al., 2016; Sabater-
376 Muñoz et al., 2017; Weinert and Welch, 2017). Due to the complexity of bacterial genome
377 size evolution and the multitude of factors involved in the process, one single factor (i.e.,
378 PMU abundance) can have such strong correlation with phytoplasma genome size is
379 surprising. Notably, as genome reduction appears to be a common and recurring theme of
380 symbiont evolution, the roles of PMU in genome expansion of some phytoplasmas require
381 further investigation. Because PMUs are known to be associated with effector genes and
382 likely can transfer horizontally between closely- (Cho et al., 2019; Seruga Music et al., 2019)
383 or distantly-related phytoplasmas (Chung et al., 2013; Ku et al., 2013), the involvement of
384 PMUs in phytoplasma effector gene content evolution is particularly important. However, it
385 is worth noting that even though horizontal transfers of PMUs may provide novel
386 combination of effector gene content, PMU evolution may have been dominated by vertical
387 inheritance or at least transfers between closely related lineages based on the observation
388 that in most cases related lineages have similar PMUs.
389
390 <u>Effector Genes</u>

391 An important feature of phytoplasmas is their ability to modulate host plant
392 development through effectors, which are small secreted proteins (Sugio et al., 2011b;
393 Sugio and Hogenhout, 2012). To date, four phytoplasma effectors have been experimentally
394 characterized, including SAP05 (Gamboa et al., 2019; Huang and Hogenhout, 2019; Huang et
395 al., 2021), SAP11/SWP1 (Bai et al., 2009; Sugio et al., 2011a; Lu et al., 2014; Chang et al.,
396 2018; Wang et al., 2018c, 2018b), SAP54/PHYL1 (MacLean et al., 2011; Maejima et al., 2014;
397 Orlovskis and Hogenhout, 2016), and TENGU (Hoshi et al., 2009; Sugawara et al., 2013;
398 Minato et al., 2014). The expanded genome sequence availability allowed us to investigate
399 the phylogenetic distribution of homologous effector genes among diverse phytoplasmas.
400 The highly variable distribution patterns suggest that the effector gene content may have
401 rapid turnover during the diversification of phytoplasmas (**Figure 3**). For example, at within-
402 species level comparison among '*Ca*. P. asteris' strains, the patterns of presence/absence
403 and gene copy number are variable for three of these effectors (**Figure 3B**) although the
404 level of core genome sequence divergence is very low as indicated by the short branch
405 lengths (**Figure 3A**). This finding is consistent with our previous in-depth characterization of
406 16SrI phytoplasmas (Cho et al., 2020a). At genus-level, TENGU appears to be restricted to
407 and conserved among 16SrI lineages in cluster I (i.e., '*Ca*. P. asteris' and '*Ca*. P. tritici'), while
408 the other three effectors are variable. Based on these patterns, it is likely that TENGU
409 originated in the most recent common ancestor (MRCA) of 16SrI phytoplasmas. However,
410 for the other three effectors, it is unclear if the MRCA of all extant phytoplasmas harbor
411 these genes or not. If yes, then multiple independent gene losses are required to explain the
412 distribution of these genes among extant phytoplasmas. Alternatively, multiple independent
413 origins, likely mediated by PMU-mediated horizontal gene transfer (Chung et al., 2013; Cho
414 et al., 2019; Seruga Music et al., 2019), are required to explain the gene distribution pattern.

415 Another interesting observation is that several phytoplasmas do not possess any of these
416 four effector genes. It is likely that these phytoplasmas harbor novel effector genes that are
417 yet to be characterized and further investigation is required to obtain a more complete
418 picture of phytoplasma effector diversity.

419 For closer inspection of these four known effectors, we performed multiple
420 sequence alignments to examine the protein sequence divergence among homologs (**Figure**
421 **S3**). Consistent with the expectation derived from their phylogenetic distribution patterns,
422 the three effector genes found among diverse phytoplasmas have higher levels of sequence
423 divergence compared to phylogenetically restricted TENGU homologs.

424 For copy number variation, '*Ca*. P. luffae' NCHU2019 stands out as having the highest
425 copy numbers for SAP11 and SAP54 (**Figure 3B**). These homologous genes are all located
426 within PMUs (**Figure 5**) and have nearly identical sequences (**Figure S3**), which suggest that
427 recent intra-genomic PMU duplications are responsible for expansions in effector gene copy
428 numbers. Similar patterns were observed for SAP11 and SAP54 homologs in '*Ca*. P. ziziphi'
429 Jwb-nky. Experimental characterization of the two SAP11 homologs in '*Ca*. P. ziziphi'
430 demonstrated that both can stimulate lateral bud outgrowth for witches' broom symptoms
431 when expressed in *Nicotiana benthamiana* (Zhou et al., 2021). Intriguingly, in phytoplasma-
432 infected jujube plants, these two SAP11 homologs have different expression patterns
433 among tissue types. These findings suggest that such gene duplication events may lead to
434 neofunctionalization or subfunctionalization, thus promoting the genetic diversity of
435 effectors.

436

437 **Conclusion**

438 In this work, we determined the complete genome sequence of an uncultivated '*Ca*.
439 P. luffae' strain associated with the witches' broom disease of loofah. This assembly
440 provides the first representative genome sequence for the 16SrVIII group of phytoplasmas
441 and improves the taxon sampling of these diverse plant-pathogenic bacteria. For
442 comparative genomics analysis conducted at genus level, we provided a global view of the
443 PMUs (i.e., phytoplasma-specific mobile genetic elements) and identified three major PMU
444 types that differ in gene organization and phylogenetic distribution. Importantly, statistical
445 analysis revealed that PMU abundance explains nearly 80% of the variance in phytoplasma
446 genome sizes, thus providing a quantitative estimate on the importance of these elements.
447 Finally, our investigation of effector genes highlighted the genetic diversity associated with
448 phytoplasma virulence and established the roles of PMUs in shaping such diversity.
449
450 **Data Availability**
451 This genome sequencing project was deposited in the NCBI under BioProject
452 PRJNA636624. The raw reads were deposited in the Sequence Read Archive (SRA) under the
453 accessions SRR11921288 (Illumina MiSeq paired-end reads) and SRR11921289 (ONT MinION
454 reads). The complete and annotated genome sequence of '*Ca*. P. luffae' NCHU2019 was
455 deposited in GenBank under the accession CP054393.
456
457 **Author Contribution**
458 JYY and CHK conceptualized the study, acquired funding, and supervised the project.
459 JYY, YCC, and CMT provided the biological materials. CMT and STC coordinated the
460 sequencing. CMT and STC conducted the initial genome assembly. STC completed the
461 assembly. YCL and CTH validate the assembly. STC and CTH conducted the comparative

analysis and prepared the figures. CTH and CHK wrote the draft. All authors approved the submitted version.

Funding

Work in the Yang lab was supported by grants-in-aid from the Ministry of Science and Technology (110-2628-B-005-002) and the Advanced Plant Biotechnology Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan. Work in the Kuo lab was supported by Academia Sinica and the Ministry of Science and Technology (106-2311-B-001-028-MY3) of Taiwan. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments

We thank Ai-Ping Chen and Shu-Jen Chou for technical assistance. The Illumina sequencing library preparation service was provided by the Genomic Technology Core Facility (Institute of Plant and Microbial Biology, Academia Sinica, Taipei. Taiwan). The Illumina paired-end sequencing service was provided by the Genomics Core Facility (Institute of Molecular Biology, Academia Sinica, Taipei. Taiwan).

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

486 **References**

487 Andersen, M. T., Liefting, L. W., Havukkala, I., and Beever, R. E. (2013). Comparison of the
488 complete genome sequence of two closely related isolates of '*Candidatus*
489 Phytoplasma australiense' reveals genome plasticity. *BMC Genomics* 14, 529.
490 doi:10.1186/1471-2164-14-529.

491 Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S.,
492 et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural
493 networks. *Nat. Biotechnol.* 37, 420–423. doi:10.1038/s41587-019-0036-z.

494 Bai, X., Correa, V. R., Toruño, T. Y., Ammar, E.-D., Kamoun, S., and Hogenhout, S. A. (2009).
495 AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant.*
496 *Microbe Interact.* 22, 18–30. doi:10.1094/MPMI-22-1-0018.

497 Bai, X., Zhang, J., Ewing, A., Miller, S. A., Jancso Radek, A., Shevchenko, D. V., et al. (2006).
498 Living with genome instability: the adaptation of phytoplasmas to diverse
499 environments of their insect and plant hosts. *J. Bacteriol.* 188, 3682–3696.
500 doi:10.1128/JB.188.10.3682-3696.2006.

501 Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., et al.
502 (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47. doi:10.1093/nar/gkx1094.

503 Bertaccini, A., Duduk, B., Paltrinieri, S., and Contaldo, N. (2014). Phytoplasmas and
504 phytoplasma diseases: a severe threat to agriculture. *Am. J. Plant Sci.* 05, 1763–1788.
505 doi:10.4236/ajps.2014.512191.

506 Bertaccini, A., and Lee, I.-M. (2018). "Phytoplasmas: An update," in *Phytoplasmas: Plant*
507 *Pathogenic Bacteria - I: Characterisation and Epidemiology of Phytoplasma -*
508 *Associated Diseases*, eds. G. P. Rao, A. Bertaccini, N. Fiore, and L. W. Liefting
509 (Singapore: Springer Singapore), 1–29. doi:10.1007/978-981-13-0119-3_1.

510 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009).
511 BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
512 doi:10.1186/1471-2105-10-421.

513 Chang, S. H., Tan, C. M., Wu, C.-T., Lin, T.-H., Jiang, S.-Y., Liu, R.-C., et al. (2018). Alterations
514 of plant architecture and phase transition by the phytoplasma virulence factor
515 SAP11. *J. Exp. Bot.* 69, 5389–5401. doi:10.1093/jxb/ery318.

516 Chen, L.-L., Chung, W.-C., Lin, C.-P., and Kuo, C.-H. (2012). Comparative analysis of gene
517 content evolution in phytoplasmas and mycoplasmas. *PLOS ONE* 7, e34407.
518 doi:10.1371/journal.pone.0034407.

519 Cho, S.-T., Kung, H.-J., Huang, W., Hogenhout, S. A., and Kuo, C.-H. (2020a). Species
520 boundaries and molecular markers for the classification of 16SrI phytoplasmas
521 inferred by genome analysis. *Front. Microbiol.* 11, 1531.
522 doi:10.3389/fmicb.2020.01531.

523 Cho, S.-T., Lin, C.-P., and Kuo, C.-H. (2019). Genomic characterization of the periwinkle leaf
524 yellowing (PLY) phytoplasmas in Taiwan. *Front. Microbiol.* 10, 2194.
525 doi:10.3389/fmicb.2019.02194.

526 Cho, S.-T., Zwolińska, A., Huang, W., Wouters, R. H. M., Mugford, S. T., Hogenhout, S. A., et
527 al. (2020b). Complete genome sequence of "*Candidatus* Phytoplasma asteris" RP166,
528 a plant pathogen associated with rapeseed phyllody disease in Poland. *Microbiol.*
529 *Resour. Announc.* 9. doi:10.1128/MRA.00760-20.

530 Chung, W.-C., Chen, L.-L., Lo, W.-S., Lin, C.-P., and Kuo, C.-H. (2013). Comparative analysis of
531 the peanut witches'-broom phytoplasma genome reveals horizontal transfer of
532 potential mobile units and effectors. *PLOS ONE* 8, e62770.
533 doi:10.1371/journal.pone.0062770.

534 Davis, R. E., Zhao, Y., Wei, W., Dally, E. L., and Lee, I.-M. (2017). '*Candidatus* Phytoplasma
535 luffae', a novel taxon associated with witches' broom disease of loofah, Luffa
536 aegyptica Mill. *Int. J. Syst. Evol. Microbiol.* 67, 3127–3133.
537 doi:10.1099/ijsem.0.001980.

538 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
539 throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.

540 Felsenstein, J. (1989). PHYLIP - phylogeny inference package (version 3.2). *Cladistics* 5, 164–
541 166.

542 Gamboa, C., Cui, W., Quiroga, N., Fernández, C., Fiore, N., and Zamorano, A. (2019).
543 Identification of 16SrIII-J phytoplasma effectors using a viral vector. *Phytopathogenic*
544 *Mollicutes* 9, 229. doi:10.5958/2249-4677.2019.00115.4.

545 Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large
546 phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
547 doi:10.1080/10635150390235520.

548 Gundersen, D. E., and Lee, I.-M. (1996). Ultrasensitive detection of phytoplasmas by nested-
549 PCR assays using two universal primer pairs. *Phytopathol. Mediterr.* 35, 144–151.

550 Guy, L., Roat Kultima, J., and Andersson, S. G. E. (2010). genoPlotR: comparative gene and
551 genome visualization in R. *Bioinformatics* 26, 2334–2335.
552 doi:10.1093/bioinformatics/btq413.

553 Hogenhout, S. A., Oshima, K., Ammar, E.-D., Kakizawa, S., Kingdom, H. N., and Namba, S.
554 (2008). Phytoplasmas: bacteria that manipulate plants and insects. *Mol. Plant Pathol.*
555 9, 403–423. doi:10.1111/j.1364-3703.2008.00472.x.

556 Hogenhout, S. A., and Seruga Music, M. (2009). "Phytoplasma genomics, from sequencing to
557 comparative and functional genomics - what have we learnt?," in *Phytoplasmas:*
558 *genomes, plant hosts and vectors* (CABI), 19–36.

559 Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., et al. (2009). A unique
560 virulence factor for proliferation and dwarfism in plants identified from a
561 phytopathogenic bacterium. *Proc. Natl. Acad. Sci.* 106, 6416–6421.
562 doi:10.1073/pnas.0813038106.

563 Huang, W., and Hogenhout, S. A. (2019). Phytoplasma effectors have converged onto
564 degrading plant transcription factors with fundamental roles in plant development
565 and defense to insects. in *IS-MPMI XVIII Congress Abstracts of Poster Presentations*
566 (Glasgow, UK: IS-MPMI), 158. doi:https://doi.org/10.1094/MPMI-32-10-S1.1.

567 Huang, W., MacLean, A. M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., et al. (2021).
568 Parasitic modulation of host development by ubiquitin-independent protein
569 degradation. *Cell* 184, 5201–5214. doi:10.1016/j.cell.2021.08.029.

570 Hyatt, D., Chen, G.-L., LoCascio, P., Land, M., Larimer, F., and Hauser, L. (2010). Prodigal:
571 prokaryotic gene recognition and translation initiation site identification. *BMC*
572 *Bioinformatics* 11, 119. doi:10.1186/1471-2105-11-119.

573 Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High
574 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
575 boundaries. *Nat. Commun.* 9, 5114. doi:10.1038/s41467-018-07641-9.

576 Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for
577 representation and analysis of molecular networks involving diseases and drugs.
578 *Nucleic Acids Res.* 38, D355–D360. doi:10.1093/nar/gkp896.

579 Kirdat, K., Tiwarekar, B., Thorat, V., Sathe, S., Shouche, Y., and Yadav, A. (2021). "*Candidatus*
580 Phytoplasma sacchari", a novel taxon - associated with Sugarcane Grassy Shoot
581 (SCGS) disease. *Int. J. Syst. Evol. Microbiol.* 71, 004591. doi:10.1099/ijsem.0.004591.

582 Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome
583 size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci.* 101, 3160–
584 3165. doi:10.1073/pnas.0308653100.

585 Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting
586 transmembrane protein topology with a hidden Markov model: application to
587 complete genomes. *J. Mol. Biol.* 305, 567–580. doi:10.1006/jmbi.2000.4315.

588 Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., et al. (2009).
589 Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–
590 1645. doi:10.1101/gr.092759.109.

591 Ku, C., Lo, W.-S., and Kuo, C.-H. (2013). Horizontal transfer of potential mobile units in
592 phytoplasmas. *Mob. Genet. Elem.* 3, e26145. doi:10.4161/mge.26145.

593 Kube, M., Mitrovic, J., Duduk, B., Rabus, R., and Seemüller, E. (2012). Current view on
594 phytoplasma genomes and encoded metabolism. *Sci. World J.* 2012, 1–25.
595 doi:10.1100/2012/185942.

596 Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A., et al. (2008). The
597 linear chromosome of the plant-pathogenic mycoplasma "*Candidatus* Phytoplasma
598 mali." *BMC Genomics* 9, 306. doi:10.1186/1471-2164-9-306.

599 Kuo, C.-H., Moran, N. A., and Ochman, H. (2009). The consequences of genetic drift for
600 bacterial genome complexity. *Genome Res.* 19, 1450–1454.
601 doi:10.1101/gr.091785.109.

602 Kuo, C.-H., and Ochman, H. (2009). Deletional bias across the three domains of life. *Genome*
603 *Biol. Evol.* 1, 145–152. doi:10.1093/gbe/evp016.

604 Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004).
605 Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12.
606 doi:10.1186/gb-2004-5-2-r12.

607 Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007).
608 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids*
609 *Res.* 35, 3100–3108. doi:10.1093/nar/gkm160.

610 Lee, I.-M., Davis, R. E., and Gundersen-Rindal, D. E. (2000). Phytoplasma: phytopathogenic
611 mollicutes. *Annu. Rev. Microbiol.* 54, 221–255. doi:10.1146/annurev.micro.54.1.221.

612 Lee, I.-M., Gundersen-Rindal, D. E., Davis, R. E., and Bartoszyk, I. M. (1998). Revised
613 classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and
614 ribosomal protein gene sequences. *Int. J. Syst. Evol. Microbiol.* 48, 1153–1169.
615 doi:10.1099/00207713-48-4-1153.

616 Lee, I.-M., Hammond, R., Davis, R., and Gundersen, D. (1993). Universal amplification and
617 analysis of pathogen 16S rDNA for classification and identification of mycoplasmalike
618 organisms. *Phytopathology* 83, 834–842. doi:10.1094/phyto-83-834.

619 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34,
620 3094–3100. doi:10.1093/bioinformatics/bty191.

621 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler
622 transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324.

623 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The
624 Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
625 doi:10.1093/bioinformatics/btp352.

626 Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for
627 eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi:10.1101/gr.1224503.

628 Lo, W.-S., Gasparich, G. E., and Kuo, C.-H. (2018). Convergent evolution among ruminant-
629 pathogenic *Mycoplasma* involved extensive gene content changes. *Genome Biol.*
630 *Evol.* 10, 2130–2139. doi:10.1093/gbe/evy172.

631 Lo, W.-S., Huang, Y.-Y., and Kuo, C.-H. (2016). Winding paths to simplicity: genome evolution
632 in facultative insect symbionts. *FEMS Microbiol. Rev.* 40, 855–874.
633 doi:10.1093/femsre/fuw028.

634 Lowe, T., and Eddy, S. (1997). tRNAscan-SE: a program for improved detection of transfer
635 RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
636 doi:10.1093/nar/25.5.955.

637 Lu, M., An, H., and Li, L. (2016). Genome survey sequencing for the characterization of the
638 genetic background of *Rosa roxburghii* Tratt and leaf ascorbate metabolism genes.
639 *PLOS ONE* 11, e0147530. doi:10.1371/journal.pone.0147530.

640 Lu, Y.-T., Li, M.-Y., Cheng, K.-T., Tan, C. M., Su, L.-W., Lin, W.-Y., et al. (2014). Transgenic
641 plants that express the phytoplasma effector SAP11 show altered phosphate
642 starvation and defense responses. *Plant Physiol.* 164, 1456–1469.
643 doi:10.1104/pp.113.229740.

644 MacLean, A. M., Sugio, A., Makarova, O. V., Findlay, K. C., Grieve, V. M., Tóth, R., et al.
645 (2011). Phytoplasma effector SAP54 induces indeterminate leaf-like flower
646 development in *Arabidopsis* plants. *Plant Physiol.* 157, 831–841.
647 doi:10.1104/pp.111.181586.

648 Maejima, K., Iwai, R., Himeno, M., Komatsu, K., Kitazawa, Y., Fujita, N., et al. (2014).
649 Recognition of floral homeotic MADS domain transcription factors by a phytoplasmal
650 effector, phyllogen, induces phyllody. *Plant J.* 78, 541–554. doi:10.1111/tpj.12495.

651 Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel
652 counting of occurrences of k-mers. *Bioinformatics* 27, 764–770.
653 doi:10.1093/bioinformatics/btr011.

654 McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic
655 bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi:10.1038/nrmicro2670.

656 Minato, N., Himeno, M., Hoshi, A., Maejima, K., Komatsu, K., Takebayashi, Y., et al. (2014).
657 The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of
658 the jasmonic acid and auxin pathways. *Sci. Rep.* 4, 7399. doi:10.1038/srep07399.

659 Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial
660 genomes. *Trends Genet.* 17, 589–596. doi:10.1016/S0168-9525(01)02447-7.

661 Namba, S. (2019). Molecular and biological properties of phytoplasmas. *Proc. Jpn. Acad. Ser.*
662 *B* 95, 401–418. doi:10.2183/pjab.95.028.

663 Nejat, N., Vadamalai, G., Davis, R. E., Harrison, N. A., Sijam, K., Dickinson, M., et al. (2013).
664 '*Candidatus* Phytoplasma malaysianum', a novel taxon associated with virescence
665 and phyllody of Madagascar periwinkle (*Catharanthus roseus*). *Int. J. Syst. Evol.*
666 *Microbiol.* 63, 540–548. doi:10.1099/ijs.0.041467-0.

667 Novichkov, P. S., Wolf, Y. I., Dubchak, I., and Koonin, E. V. (2009). Trends in prokaryotic
668 evolution revealed by comparison of closely related bacterial and archaeal genomes.
669 *J. Bacteriol.* 191, 65–73. doi:10.1128/JB.01237-08.

670 Ochman, H., and Davalos, L. M. (2006). The nature and dynamics of bacterial genomes.
671 *Science* 311, 1730–1733. doi:10.1126/science.1119966.

672 Orlovskis, Z., Canale, M. C., Haryono, M., Lopes, J. R. S., Kuo, C.-H., and Hogenhout, S. A.
673 (2017). A few sequence polymorphisms among isolates of maize bushy stunt
674 phytoplasma associate with organ proliferation symptoms of infected maize plants.
675 *Ann. Bot.* 119, 869–884. doi:10.1093/aob/mcw213.

676 Orlovskis, Z., and Hogenhout, S. A. (2016). A bacterial parasite effector mediates insect
677 vector attraction in host plants independently of developmental changes. *Front.*
678 *Plant Sci.* 7, 885. doi:10.3389/fpls.2016.00885.

679 Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.-Y., Wei, W., Suzuki, S., et al. (2004).
680 Reductive evolution suggested from the complete genome sequence of a plant-
681 pathogenic phytoplasma. *Nat. Genet.* 36, 27–29. doi:10.1038/ng1277.

682 R Core Team (2019). *R: A language and environment for statistical computing*. Vienna,
683 Austria: R Foundation for Statistical Computing Available at: https://www.r-
684 project.org/ [Accessed August 1, 2019].

685 Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al.
686 (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
687 doi:10.1038/nbt.1754.

688 Sabater-Muñoz, B., Toft, C., Alvarez-Ponce, D., and Fares, M. A. (2017). Chance and
689 necessity in the genome evolution of endosymbiotic bacteria of insects. *ISME J.* 11,
690 1291–1304. doi:10.1038/ismej.2017.18.

691 Saccardo, F., Martini, M., Palmano, S., Ermacora, P., Scortichini, M., Loi, N., et al. (2012).
692 Genome drafts of four phytoplasma strains of the ribosomal group 16SrIII.
693 *Microbiology* 158, 2805–2814. doi:10.1099/mic.0.061432-0.

694 Seruga Music, M., Samarzija, I., Hogenhout, S. A., Haryono, M., Cho, S.-T., and Kuo, C.-H.
695 (2019). The genome of '*Candidatus* Phytoplasma solani' strain SA-1 is highly dynamic
696 and prone to adopting foreign sequences. *Syst. Appl. Microbiol.* 42, 117–127.
697 doi:10.1016/j.syapm.2018.10.008.

698 Sugawara, K., Honma, Y., Komatsu, K., Himeno, M., Oshima, K., and Namba, S. (2013). The
699 alteration of plant morphology by small peptides released from the proteolytic
700 processing of the bacterial peptide TENGU. *Plant Physiol.* 162, 2005–2014.
701 doi:10.1104/pp.113.218586.

702 Sugio, A., and Hogenhout, S. A. (2012). The genome biology of phytoplasma: modulators of
703 plants and insects. *Curr. Opin. Microbiol.* 15, 247–254.
704 doi:10.1016/j.mib.2012.04.002.

705 Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M., and Hogenhout, S. A. (2011a).
706 Phytoplasma protein effector SAP11 enhances insect vector reproduction by
707 manipulating plant development and defense hormone biosynthesis. *Proc. Natl.*
708 *Acad. Sci.* 108, E1254–E1263. doi:10.1073/pnas.1105664108.

709 Sugio, A., MacLean, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R., and Hogenhout,
710 S. A. (2011b). Diverse targets of phytoplasma effectors: from plant development to
711 defense against insects. *Annu. Rev. Phytopathol.* 49, 175–195. doi:10.1146/annurev-
712 phyto-072910-095323.

713 Tan, C. M., Lin, Y.-C., Li, J.-R., Chien, Y.-Y., Wang, C.-J., Chou, L., et al. (2021). Accelerating
714 complete phytoplasma genome assembly by immunoprecipitation-based enrichment
715 and MinION-based DNA sequencing for comparative analyses. *Front. Microbiol.* 12,
716 766221. doi:10.3389/fmicb.2021.766221.

717 Tatusov, R., Fedorova, N., Jackson, J., Jacobs, A., Kiryutin, B., Koonin, E., et al. (2003). The
718 COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
719 doi:10.1186/1471-2105-4-41.

720 The IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma taxonomy group (2004).
721 '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that
722 colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* 54, 1243–1255.
723 doi:10.1099/ijs.0.02854-0.

724 Tran-Nguyen, L. T. T., Kube, M., Schneider, B., Reinhardt, R., and Gibb, K. S. (2008).
725 Comparative genome analysis of "*Candidatus* Phytoplasma australiense" (subgroup
726 tuf-Australia I; rp-A) and "*Ca*. Phytoplasma asteris" strains OY-M and AY-WB. *J.*
727 *Bacteriol.* 190, 3979–3991. doi:10.1128/JB.01301-07.

728 Wang, J., Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., et al. (2018a). Comparative genome
729 analysis of jujube witches'-broom phytoplasma, an obligate pathogen that causes
730 jujube witches'-broom disease. *BMC Genomics* 19, 689. doi:10.1186/s12864-018-
731 5075-1.

732 Wang, N., Li, Y., Chen, W., Yang, H. Z., Zhang, P. H., and Wu, Y. F. (2018b). Identification of
733 wheat blue dwarf phytoplasma effectors targeting plant proliferation and defence
734 responses. *Plant Pathol.* 67, 603–609. doi:10.1111/ppa.12786.

735 Wang, N., Yang, H., Yin, Z., Liu, W., Sun, L., and Wu, Y. (2018c). Phytoplasma effector SWP1
736 induces witches' broom symptom by destabilizing the TCP transcription factor
737 BRANCHED1. *Mol. Plant Pathol.* 19, 2623–2634. doi:10.1111/mpp.12733.

738 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009).
739 Jalview version 2—a multiple sequence alignment editor and analysis workbench.
740 *Bioinformatics* 25, 1189–1191. doi:10.1093/bioinformatics/btp033.

741 Weinert, L. A., and Welch, J. J. (2017). Why might bacterial pathogens have small genomes?
742 *Trends Ecol. Evol.* 32, 936–947. doi:10.1016/j.tree.2017.09.006.

743 Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., et al. (2009). A
744 phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462,
745 1056–1060. doi:10.1038/nature08656.

746 Wu, M., and Eisen, J. (2008). A simple, fast, and accurate method of phylogenomic
747 inference. *Genome Biol.* 9, R151. doi:10.1186/gb-2008-9-10-r151.

748 Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly
749 using de Bruijn graphs. *Genome Res.* 18, 821–829. doi:10.1101/gr.074492.107.

750 Zhao, Y., and Davis, R. E. (2016). Criteria for phytoplasma 16Sr group/subgroup delineation
751 and the need of a platform for proper registration of new groups and subgroups. *Int.*
752 *J. Syst. Evol. Microbiol.* 66, 2121–2123. doi:10.1099/ijsem.0.000999.

753 Zhao, Y., Wei, W., Davis, R. E., Lee, I.-M., and Bottner-Parker, K. D. (2021). The agent
754 associated with blue dwarf disease in wheat represents a new phytoplasma taxon,
755 '*Candidatus* Phytoplasma tritici.' *Int. J. Syst. Evol. Microbiol.*
756 doi:10.1099/ijsem.0.004604.

757 Zhao, Y., Wei, W., Lee, I.-M., Shao, J., Suo, X., and Davis, R. E. (2009). Construction of an
758 interactive online phytoplasma classification tool, *i*PhyClassifier, and its application
759 in analysis of the peach X-disease phytoplasma group (16SrIII). *Int. J. Syst. Evol.*
760 *Microbiol.* 59, 2582–2593. doi:10.1099/ijs.0.010249-0.

761 Zhou, J., Ma, F., Yao, Y., Deng, M., Chen, M., Zhang, S., et al. (2021). JWB phytoplasma
762 effectors SJP1 and SJP2 induce lateral bud outgrowth by repressing the ZjBRC1-
763 controlled auxin efflux channel. *Plant Cell Environ.* n/a. doi:10.1111/pce.14141.

764

765

766

767

768 **Table**

769 **Table 1.** List of the genome assemblies analyzed. For each strain, information regarding the
770 16S rRNA gene (16Sr) group, genome accession number, assembly status, genome size,
771 coding sequence (CDS) count, potential mobile unit (PMU) gene count, and combined length
772 of PMU genes are provided. The PMU gene information is based on homologs of eight core
773 genes (i.e., *tra5*, *dnaB*, *dnaG*, *tmk*, *hflB*, *himA*, *ssb*, and *rpoD*); other genes such as those
774 encode hypothetical proteins or putative secreted proteins are excluded. *Acholeplasma*
775 *laidlawii* is included as an outgroup.

776

777 **Figure Legends**

778 **Figure 1.** Infection symptoms. (A) The loofah plant exhibiting phytoplasma infection
779 symptoms that was collected in Changhua, Taiwan. (B) Loofah plants grown in the lab. Left,
780 healthy control; right, grafted with the infected plant shown in panel A.

781

782 **Figure 2.** Genome map of '*Candidatus* Phytoplasma luffae' NCHU2019. Rings from outside
783 in: (1) Scale marks (kb). (2 and 3) Coding sequences on the forward and reverse strand,
784 respectively. Color-coded by functional categories. (4) Genes associated with potential
785 mobile units (PMUs), color-coded by annotation (see **Figure 5**). Gene clusters that represent
786 individual PMUs are labeled by orange lines. Genes encode putative secreted proteins are in
787 red. (5) GC skew (positive: dark gray; negative: light gray). (6) GC content (above average:
788 dark gray; below average; light gray). One high GC peak located in the 531 to 544 kb region
789 corresponds to two adjacent rRNA operons.

790

791 **Figure 3.** Evolutionary relationships among representative phytoplasmas with genome
792 sequences available. The 16S rRNA gene (16Sr) group assignments are labeled on the right.
793 The three major phylogenetic clusters are indicated by colored backgrounds (I: blue; II;
794 green; III: yellow). *Acholeplasma laidlawii* is included as the outgroup. (A) Maximum
795 likelihood phylogeny inferred based on a concatenated alignment of 80 conserved single-
796 copy genes (27,742 aligned amino acid sites). All internal nodes received > 80% bootstrap
797 support based on 1,000 re-sampling. Strains with complete genome assemblies available are
798 highlighted in bold. The number in parentheses following the strain name indicates the
799 genome size (in kb). (B) Distribution of known effector genes. Gene presence and absence
800 are indicated by filled (with copy numbers) and empty circles, respectively.
801
802 **Figure 4.** Pairwise genome alignments. The genome of '*Ca*. P. luffae' NCHU2019 was used as
803 the reference for comparison with (A) itself and (B) '*Ca*. P. ziziphi' Jwb-nky. Matches on the
804 same strand and the opposite strand are indicated in red and blue, respectively. Potential
805 mobile units (PMUs) are illustrated as orange arrows on both axes.
806
807 **Figure 5.** Gene organization of potential mobile units (PMUs). Each individual PMU is
808 labeled by the phytoplasma strain name and a numerical identifier. Background colors for
809 the PMU identifiers indicate the three phylogenetic clusters of phytoplasmas (I: blue; II;
810 green; III: yellow). Genes are drawn to scale and color-coded according to annotation.
811 Homologs of putative effectors identified in the '*Ca*. P. asteris' AYWB genomes are labeled,
812 those with experimental evidence (i.e., SAP05, SAP11, and SAP54) are highlighted in bold.
813 (A) All PMUs in '*Ca*. P. luffae' NCHU2019 and '*Ca*. P. ziziphi' Jwb-nky. Grouped by genomes.
814 (B) Representative PMUs from selected phytoplasmas. Grouped by PMU types.

815

816 **Figure 6.** Maximum likelihood phylogenies of PMU core genes (A) *tmk* (228 aligned amino

817 acid sites). (B) *dnaB* (531 aligned amino acid sites). In both panels, a non-PMU homolog is

818 included as the outgroup. Numbers next to internal branches indicate the bootstrap support

819 levels based on 1,000 re-sampling. Background colors for the gene identifiers indicate the

820 three phylogenetic clusters of phytoplasmas (I: blue; II; green; III: yellow).

821

822 **Figure 7.** Correlation between the combined length of PMU core genes and genome size.

823 Strains with complete and draft genome assemblies are indicated by filled and open circles,

824 respectively. The linear regression line was based on all available assemblies.

825

826

827 **Supplementary Materials**

828 **Figure S1.** Coverage pattern of raw read mapping results to the genome assembly. Rings

829 from outside in: (1) Scale marks (kb). (2) Sequencing depth calculated from Illumina (dark

830 blue; outward) and ONT (light blue; inward) raw reads mapped to the assembly using a

831 sliding window of 200-bp. An alignment score cutoff of 100 was applied to remove spurious

832 hits. The peaks located in the 531 to 544 kb region correspond to rRNA gene clusters. (3)

833 Potential mobile unit (PMU) regions (orange). (4) Two duplicated chromosomal segments

834 (green; positions: 316-391 and 391-467 kb).

835

836 **Figure S2.** Genome size estimation based on k-mer analysis. (A) A representative frequency

837 distribution of k-mer depth (k = 21). The peak depth is 51. (B) Correspondence between

838 genome size estimates and k-mer depths.

839

840 **Figure S3.** Protein sequence alignments of four characterized phytoplasma effectors.

841 Individual effectors are identified using locus tags. Shaded background colors indicate the

842 levels of sequence conservation. (A) SAP05. (B) SAP11/SWP1. (C) SAP54/PHYL1. (D) TENGU.

843

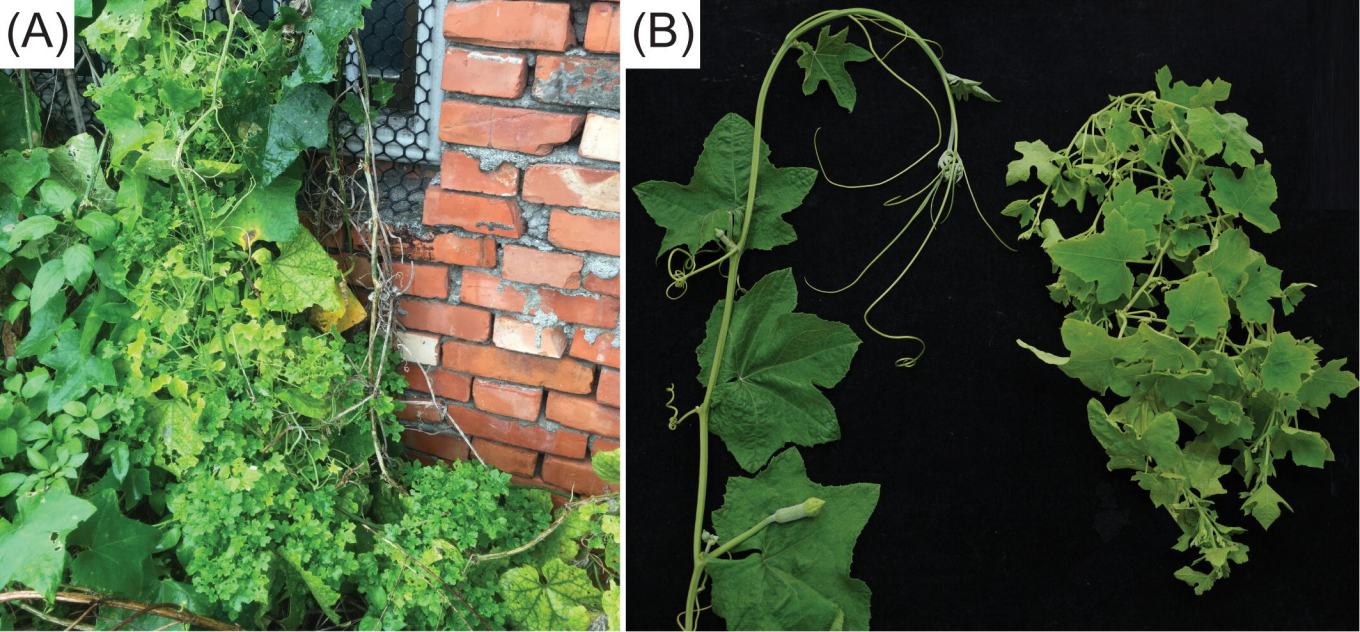
844 **Data S1.** Concatenated multiple sequence alignment of conserved single copy-genes used

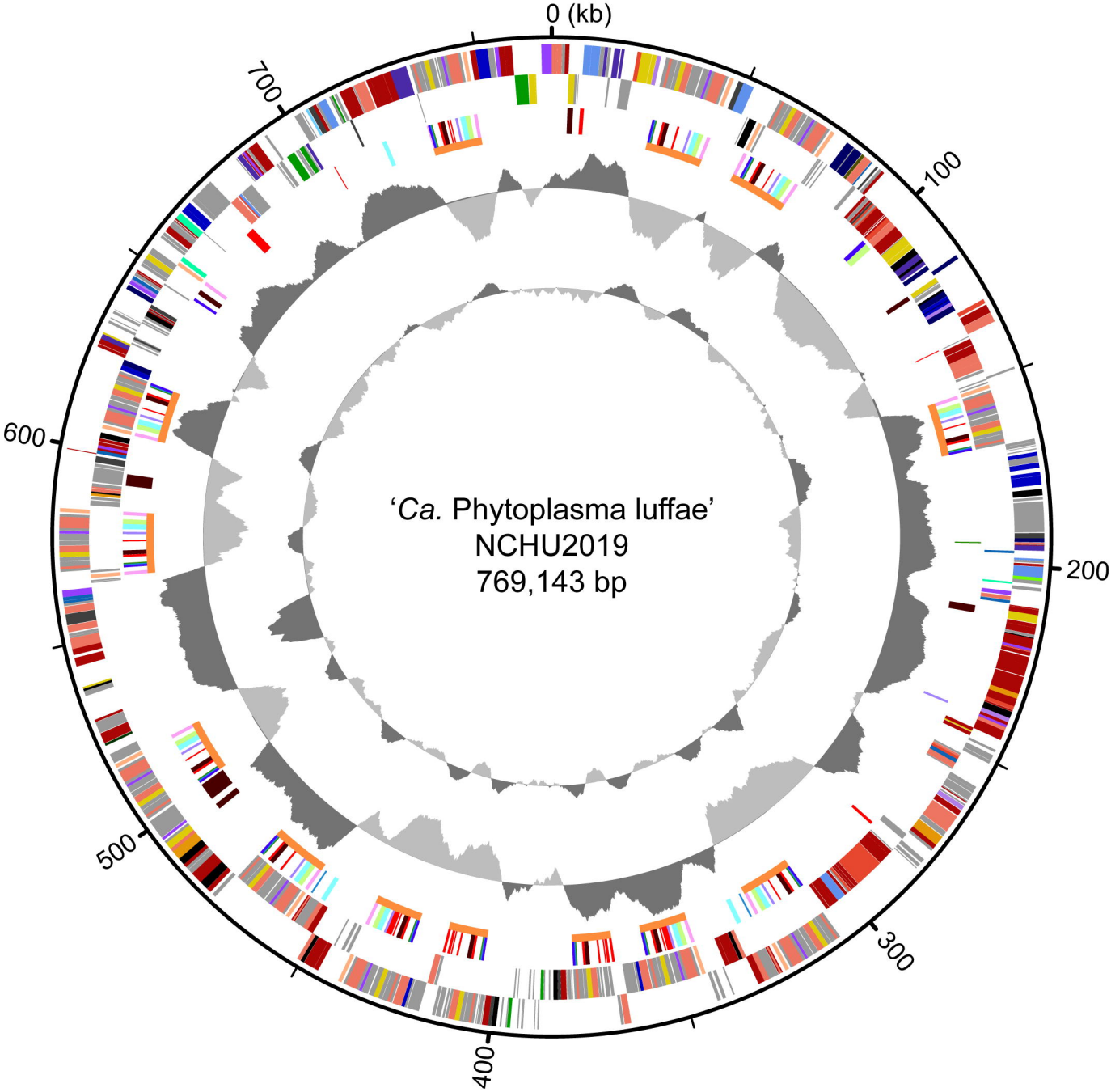
845 for inferring the molecular phylogeny shown in Figure 3. Simple text file in PHYLIP format.

846

Table 1. List of the genome assemblies analyzed. For each strain, information regarding the 16S rRNA gene (16Sr) group, genome accession number, assembly status, genome size, coding sequence (CDS) count, potential mobile unit (PMU) gene count, and combined length of PMU genes are provided. The PMU gene information is based on homologs of eight core genes (i.e., *tra5*, *dnaB*, *dnaG*, *tmk*, *hflB*, *himA*, *ssb*, and *rpoD*); other genes such as those encode hypothetical proteins or putative secreted proteins are excluded. *Acholeplasma laidlawii* is included as an outgroup.

Strain	16Sr Group	Accession	Assembly	Genome Size (bp)	CDS Count	PMU Gene Count	PMU Gene Length (bp)
'*Ca.* P. asteris' AYWB	I	GCF_000012225.1	Complete	706,569	514	112	49,188
'*Ca.* P. asteris' De Villa	I	GCF_004214875.1	Complete	600,116	521	21	11,958
'*Ca.* P. asteris' DY2014	I	GCA_005093185.1	8 contigs	824,596	775	89	48,761
'*Ca.* P. asteris' LD1	I	GCF_001866375.1	8 contigs	599,264	513	29	15,441
'*Ca.* P. asteris' M3	I	GCF_001712875.1	Complete	576,118	490	27	14,541
'*Ca.* P. asteris' OY-M	I	GCF_000009845.1	Complete	853,092	708	137	72,399
'*Ca.* P. tritici' WBD	I	GCF_000495255.1	6 contigs	611,462	471	45	19,549
'*Ca.* P. aurantifolia' NCHU2014	II	GCA_001307505.2	Complete	635,584	471	23	16,378
'*Ca.* P. pruni' CX	III	GCF_001277135.1	46 contigs	598,511	434	31	22,474
'*Ca.* P. ziziphi' Jwb-nky	V	GCF_003640545.1	Complete	750,803	641	69	49,824
'*Ca.* P. luffae' NCHU2019	VIII	GCA_018024475.1	Complete	769,143	725	117	84,321
'*Ca.* P. mali' AT	X	GCF_000026205.1	Complete	601,943	495	20	15,084
'*Ca.* P. oryzae' Mbita	XI	GCF_001578535.1	28 contigs	533,195	432	17	14,259
'*Ca.* P. sacchari' SCGS	XI	GCF_009268105.1	28 contigs	502,197	402	15	13,620
'*Ca.* P. australiense' NZSb11	XII	GCF_000397185.1	Complete	959,779	828	131	78,642
'*Ca.* P. australiense' PAa	XII	GCF_000069925.1	Complete	879,959	699	108	55,565
'*Ca.* P. solani' SA-1	XII	GCA_003698095.1	19 contigs	821,322	709	66	45,234
'*Ca.* P. cynodontis' LW01	XIV	GCF_009268075.1	23 contigs	483,935	425	13	11,856
'*Ca.* P. pini' MDPP	XXI	GCF_007821455.1	16 contigs	474,136	392	15	11,685
Acholeplasma laidlawii PG-8A	N/A	GCF_000018785.1	Complete	1,496,992	1,378	0	0


(A)
(B)



INFORMATION

- [J] Translation
- [K] Transcription
- [L] Replication

POORLY CHARACTERIZED

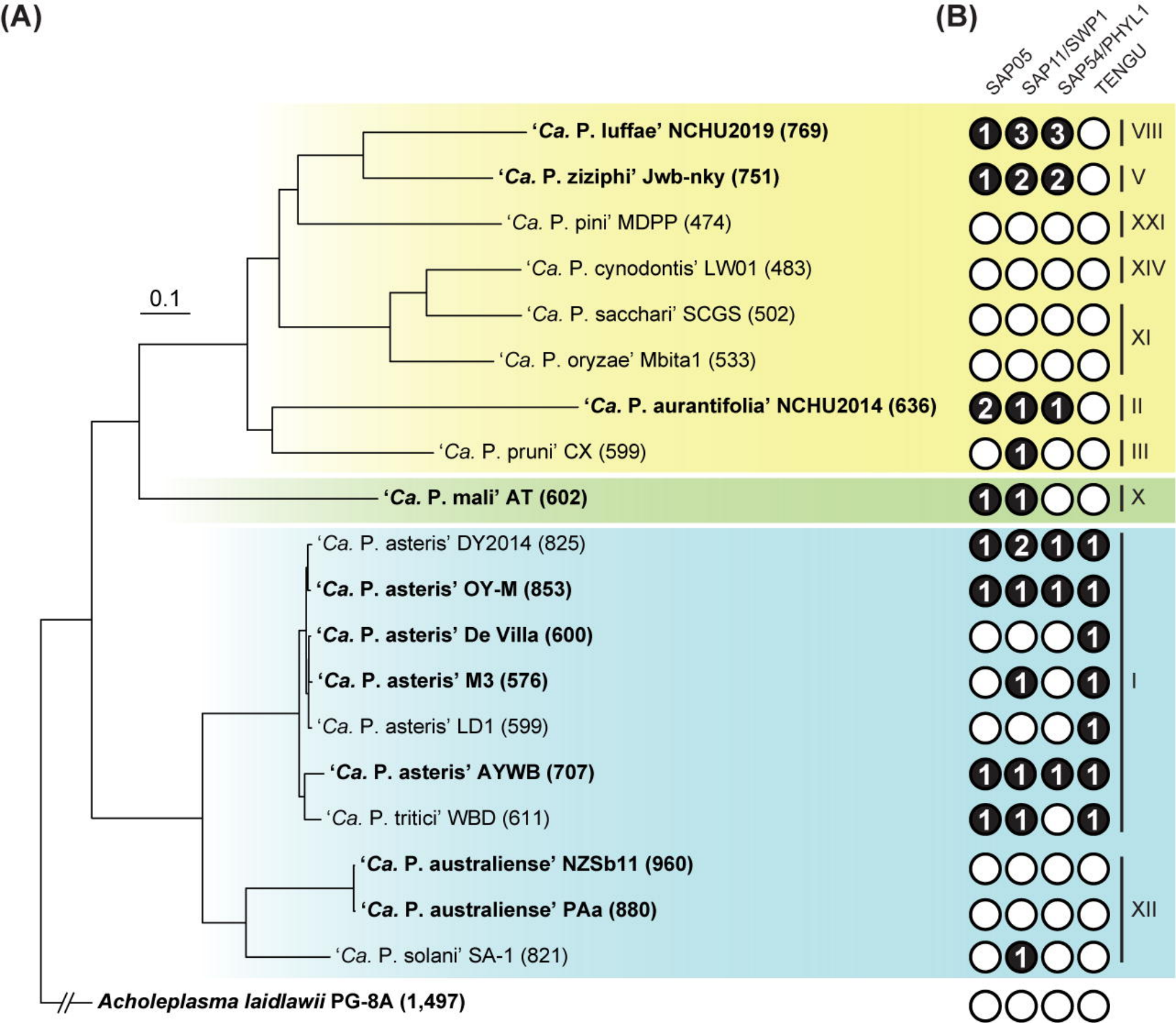
- [R] General prediction only
- [S] Function unknown
- [?] No COG assignment

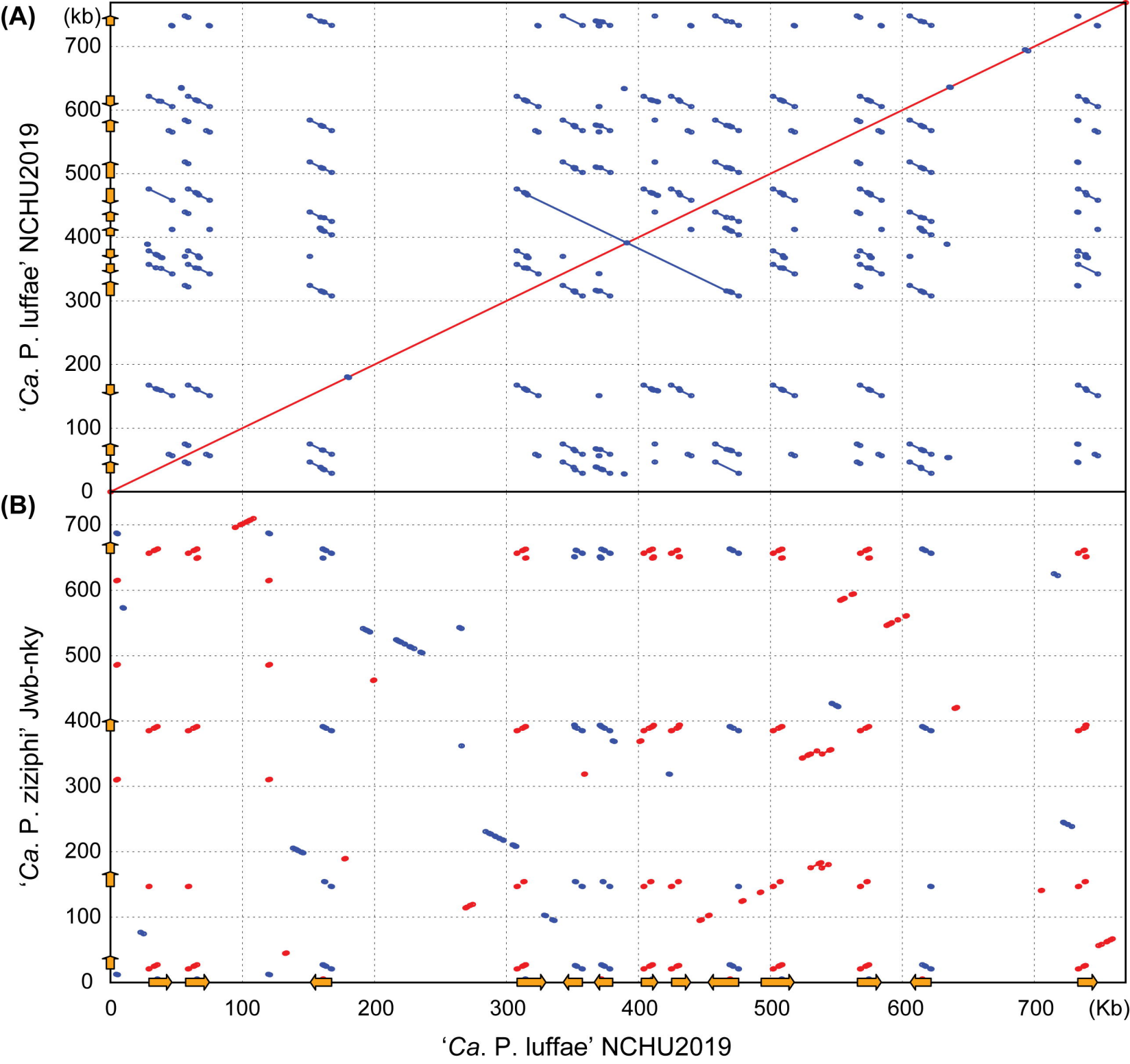
CELL PROCESS

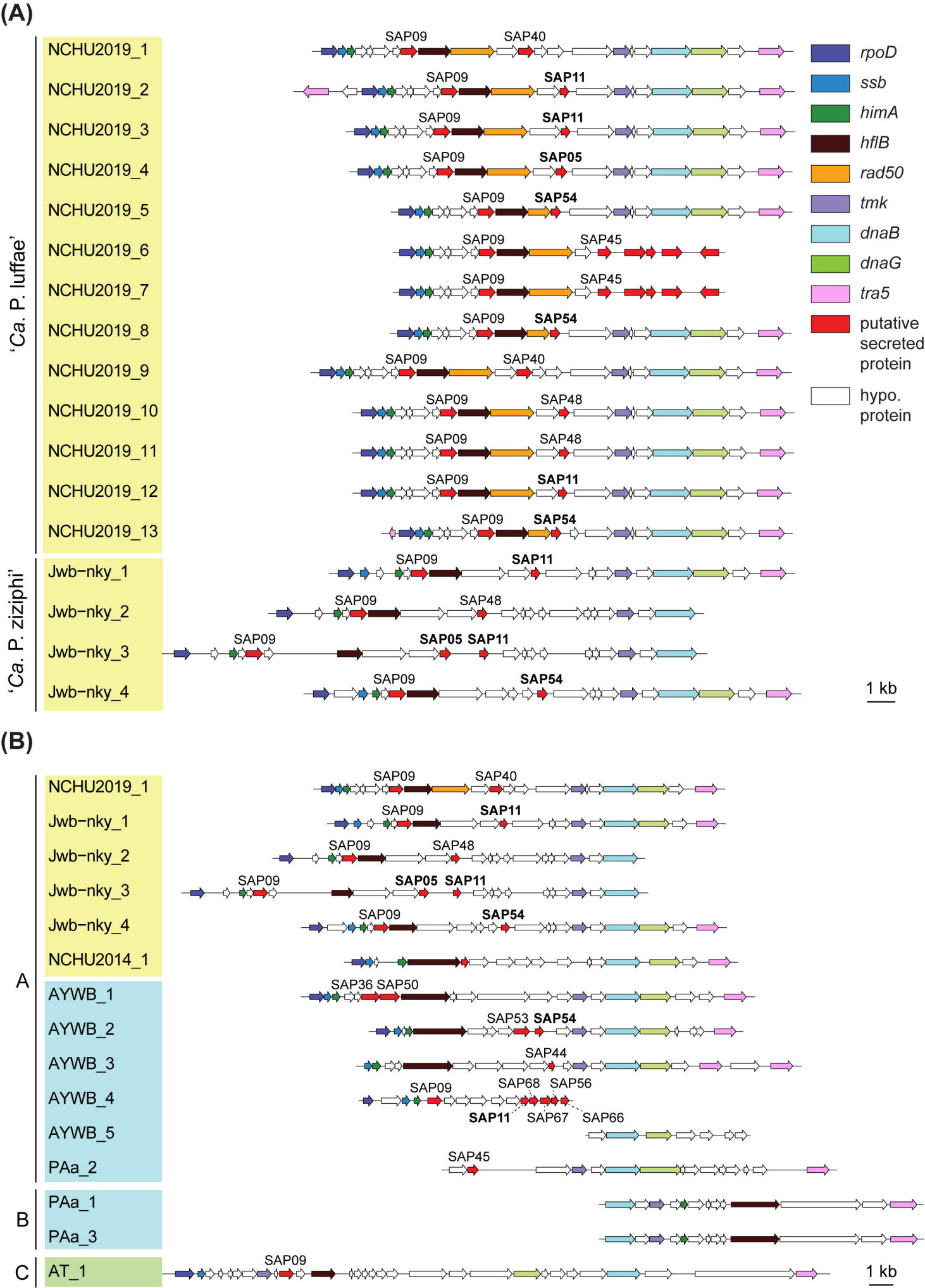
- [D] Cell cycle
- [V] Defense
- [T] Signal transduction
- [M] Cell wall, membrane, & envelope
- [N] Motility
- [W] Extracellular structure
- [U] Trafficking, secretion, & transport
- [O] Posttranslational modification
- [X] Mobilome

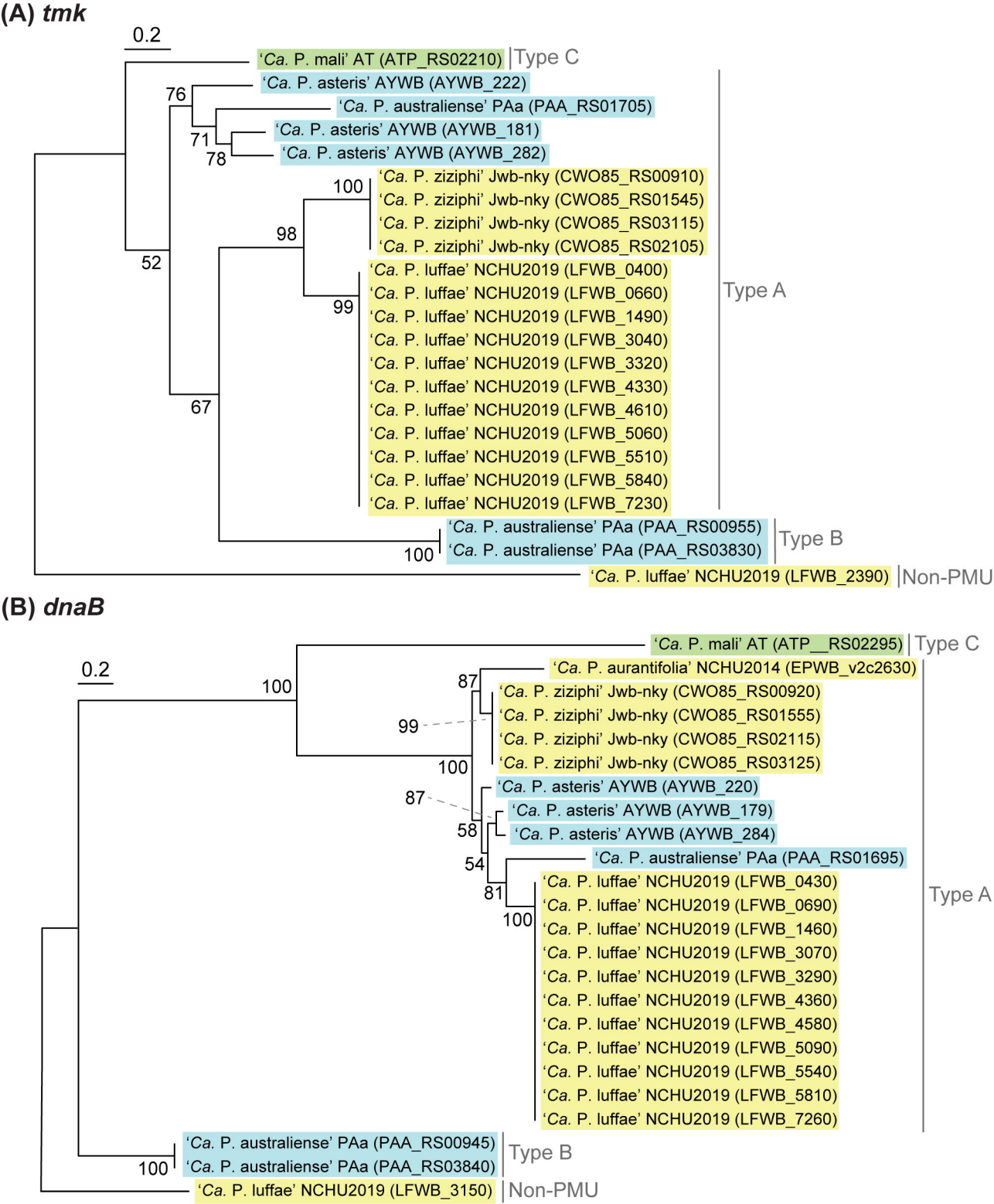
METABOLISM

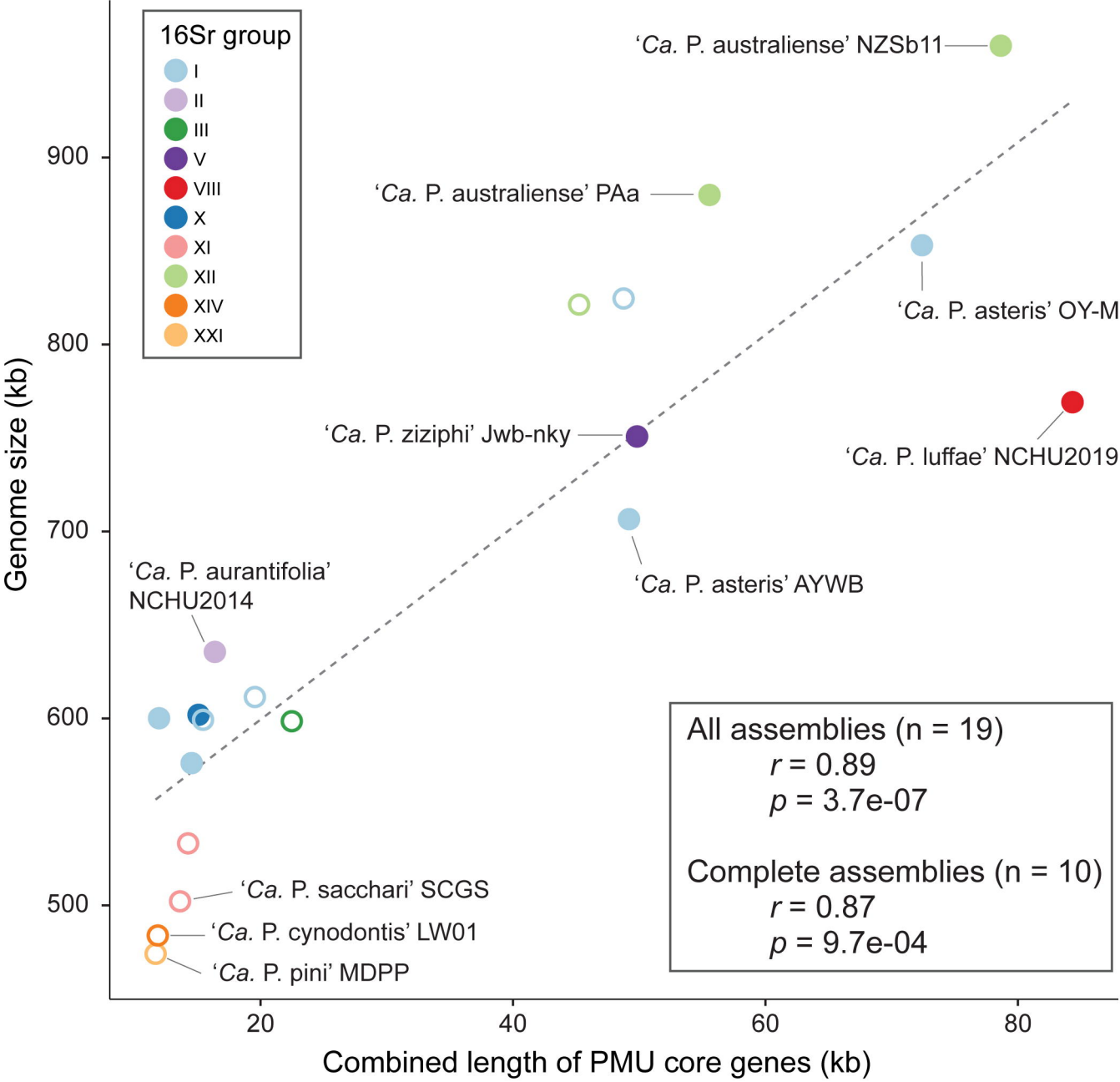
- [C] Energy
- [G] Carbohydrate
- [E] Amino acid
- [F] Nucleotide
- [H] Coenzyme
- [I] Lipid
- [P] Inorganic ion
- [Q] Secondary metabolites


(A)
(B)
SAP05
SAP11/SWP1
SAP54/PHYL1
TENGU
0.1
'Ca. P. luffae' NCHU2019 (769)
'Ca. P. ziziphi' Jwb-nky (751)
'Ca. P. pini' MDPP (474)
'Ca. P. cynodontis' LW01 (483)
'Ca. P. sacchari' SCGS (502)
'Ca. P. oryzae' Mbita1 (533)
'Ca. P. aurantifolia' NCHU2014 (636)
'Ca. P. pruni' CX (599)
'Ca. P. mali' AT (602)
'Ca. P. asteris' DY2014 (825)
'Ca. P. asteris' OY-M (853)
'Ca. P. asteris' De Villa (600)
'Ca. P. asteris' M3 (576)
'Ca. P. asteris' LD1 (599)
'Ca. P. asteris' AYWB (707)
'Ca. P. tritici' WBD (611)
'Ca. P. australiense' NZSb11 (960)
'Ca. P. australiense' PAa (880)
'Ca. P. solani' SA-1 (821)
Acholeplasma laidlawii PG-8A (1,497)
VIII
V
XXI
XIV
XI
II
III
X
I
XII


(A)
(kb)
700
600
500
400
300
200
100
0
'Ca. P. luffae' NCHU2019
(B)
700
600
500
400
300
200
100
0
'Ca. P. ziziphi' Jwb-nky
0
100
200
300
400
500
600
700
(Kb)
'Ca. P. luffae' NCHU2019


(A)
'Ca. P. luffae'
NCHU2019_1
NCHU2019_2
NCHU2019_3
NCHU2019_4
NCHU2019_5
NCHU2019_6
NCHU2019_7
NCHU2019_8
NCHU2019_9
NCHU2019_10
NCHU2019_11
NCHU2019_12
NCHU2019_13
'Ca. P. ziziphi'
Jwb-nky_1
Jwb-nky_2
Jwb-nky_3
Jwb-nky_4
SAP09
SAP40
SAP11
SAP05
SAP54
SAP45
SAP48
rpoD
ssb
himA
hflB
rad50
tmk
dnaB
dnaG
tra5
putative secreted protein
hypo. protein
1 kb
(B)
A
NCHU2019_1
Jwb-nky_1
Jwb-nky_2
Jwb-nky_3
Jwb-nky_4
NCHU2014_1
AYWB_1
AYWB_2
AYWB_3
AYWB_4
AYWB_5
PAa_2
B
PAa_1
PAa_3
C
AT_1
SAP36
SAP50
SAP53
SAP44
SAP68
SAP56
SAP67
SAP66
1 kb


(A) tmk
0.2
'Ca. P. mali' AT (ATP_RS02210)
Type C
'Ca. P. asteris' AYWB (AYWB_222)
'Ca. P. australiense' PAa (PAA_RS01705)
'Ca. P. asteris' AYWB (AYWB_181)
'Ca. P. asteris' AYWB (AYWB_282)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS00910)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS01545)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS03115)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS02105)
'Ca. P. luffae' NCHU2019 (LFWB_0400)
'Ca. P. luffae' NCHU2019 (LFWB_0660)
'Ca. P. luffae' NCHU2019 (LFWB_1490)
'Ca. P. luffae' NCHU2019 (LFWB_3040)
'Ca. P. luffae' NCHU2019 (LFWB_3320)
'Ca. P. luffae' NCHU2019 (LFWB_4330)
'Ca. P. luffae' NCHU2019 (LFWB_4610)
'Ca. P. luffae' NCHU2019 (LFWB_5060)
'Ca. P. luffae' NCHU2019 (LFWB_5510)
'Ca. P. luffae' NCHU2019 (LFWB_5840)
'Ca. P. luffae' NCHU2019 (LFWB_7230)
Type A
'Ca. P. australiense' PAa (PAA_RS00955)
'Ca. P. australiense' PAa (PAA_RS03830)
Type B
'Ca. P. luffae' NCHU2019 (LFWB_2390)
Non-PMU
76
71
78
100
98
52
99
67
100
(B) dnaB
0.2
'Ca. P. mali' AT (ATP__RS02295)
Type C
'Ca. P. aurantifolia' NCHU2014 (EPWB_v2c2630)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS00920)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS01555)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS02115)
'Ca. P. ziziphi' Jwb-nky (CWO85_RS03125)
'Ca. P. asteris' AYWB (AYWB_220)
'Ca. P. asteris' AYWB (AYWB_179)
'Ca. P. asteris' AYWB (AYWB_284)
'Ca. P. australiense' PAa (PAA_RS01695)
'Ca. P. luffae' NCHU2019 (LFWB_0430)
'Ca. P. luffae' NCHU2019 (LFWB_0690)
'Ca. P. luffae' NCHU2019 (LFWB_1460)
'Ca. P. luffae' NCHU2019 (LFWB_3070)
'Ca. P. luffae' NCHU2019 (LFWB_3290)
'Ca. P. luffae' NCHU2019 (LFWB_4360)
'Ca. P. luffae' NCHU2019 (LFWB_4580)
'Ca. P. luffae' NCHU2019 (LFWB_5090)
'Ca. P. luffae' NCHU2019 (LFWB_5540)
'Ca. P. luffae' NCHU2019 (LFWB_5810)
'Ca. P. luffae' NCHU2019 (LFWB_7260)
Type A
'Ca. P. australiense' PAa (PAA_RS00945)
'Ca. P. australiense' PAa (PAA_RS03840)
Type B
'Ca. P. luffae' NCHU2019 (LFWB_3150)
Non-PMU
100
87
99
100
87
58
54
81
100
100


16Sr group
I
II
III
V
VIII
X
XI
XII
XIV
XXI
'Ca. P. australiense' NZSb11
'Ca. P. australiense' PAa
'Ca. P. asteris' OY-M
'Ca. P. ziziphi' Jwb-nky
'Ca. P. luffae' NCHU2019
'Ca. P. asteris' AYWB
'Ca. P. aurantifolia' NCHU2014
'Ca. P. sacchari' SCGS
'Ca. P. cynodontis' LW01
'Ca. P. pini' MDPP
All assemblies (n = 19)
r = 0.89
p = 3.7e-07
Complete assemblies (n = 10)
r = 0.87
p = 9.7e-04
Genome size (kb)
900
800
700
600
500
20
40
60
80
Combined length of PMU core genes (kb)