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 Huang¹, Shu-Ting Cho¹, Choon-Meng Tan², Yi-Ching Chiu²,³, Jun-Yi Yang²,³,⁴*, Chih-Horng Kuo¹,³,⁵,⁶*

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
Introduction

Phytoplasmas are uncultivated bacteria associated with plant diseases in several hundred species (Lee et al., 2000; Hogenhout et al., 2008; Bertaccini et al., 2014; Namba, 2019). In infected plants, phytoplasma cells are restricted to phloem tissues and can secrete effector proteins that cause developmental abnormalities of the hosts (Sugio et al., 2011b).

Typical symptoms of phytoplasma infections include stunting, dwarfism, virescence (i.e., greening of flowers), phyllody (i.e., abnormal development of floral parts into leaf-like tissues), and witches’ broom (i.e., proliferation of stems and leaves), which result in substantial agricultural losses.

For classification of these uncultivated bacteria, a system based on restriction fragment length polymorphism (RFLP) analysis of their 16S rRNA genes was developed in the 1990s (Lee et al., 1993, 1998; Gundersen and Lee, 1996) and at least 33 16S rRNA gene RFLP (16Sr) groups have been described (Zhao et al., 2009; Zhao and Davis, 2016). Later, a provisional genus-level taxon ‘Candidatus Phytoplasma’ was proposed to accommodate these bacteria (The IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma taxonomy group, 2004) and at least 41 ‘Ca. P.’ species have been described or proposed (Bertaccini and Lee, 2018). Based on analysis of 16S rRNA and other conserved genes, phytoplasmas are divided into three major phylogenetic clusters (Hogenhout and Seruga Music, 2009; Chung et al., 2013; Seruga Music et al., 2019). Early genomics studies were mainly conducted for clusters I (i.e., ‘Ca. P. asteris’ of group 16Srl and ‘Ca. P. australiense’ of 16SrXII) (Oshima et al., 2004; Bai et al., 2006; Tran-Nguyen et al., 2008) and II (i.e., ‘Ca. P. mali’ of 16SrX) (Kube et al., 2008). In comparison, cluster III harbors the highest level of diversity, yet has received limited attention for comparative genomics studies (Chung et al., 2013; Wang et al., 2018a).
To improve our understanding of phytoplasma genomes, we conducted whole genome sequencing of a ‘Ca. P. luffae’ strain collected in Taiwan. The species-level taxon ‘Ca. P. luffae’ belongs to group 16SrVIII in cluster III and is associated with witches’ broom disease of loofah (Luffa aegyptiaca) (Davis et al., 2017). The availability of a complete genome sequence from this taxon provides a complete view of its gene content, which can facilitate the study of its pathogenesis mechanisms and other aspects of biology. More importantly, with increased availability of genome sequences from diverse phytoplasmas (Table 1), we performed genus-level comparative analysis to obtain a more comprehensive picture of their genomic diversity. This improves upon previous works that are limited to comparisons of closely related taxa or have sparse sampling (Bai et al., 2006; Tran-Nguyen et al., 2008; Kube et al., 2012; Saccardo et al., 2012; Andersen et al., 2013; Chung et al., 2013; Orlovskis et al., 2017; Wang et al., 2018a; Cho et al., 2019, 2020a; Seruga Music et al., 2019; Kirdat et al., 2021; Zhao et al., 2021). Furthermore, our focused analysis of the potential mobile units (PMUs) (Bai et al., 2006) revealed the influential roles of these mobile genetic elements in the evolution of phytoplasma genome organization and effector gene content.

Materials and Methods

Biological Materials

The strain NCHU2019 was collected from a naturally infected loofah plant found in Dacheng Township (Changhua County, Taiwan; 23.861860 N, 120.291322 E) on July 4th, 2019 (Figure 1A). After initial collection, the bacterium was transferred to lab-grown loofah plants (cultivar A-Jun, Known-You Seed Co., Kaohsiung, Taiwan) through grafting and maintained in a plant growth facility in the National Chung Hsing University (Taichung, Taiwan).
Taiwan) (Figure 1B). To confirm the presence and identity of this phytoplasma strain, a partial sequence of the rRNA operon was PCR amplified using the phytoplasma-specific primer set P1/P7 for Sanger sequencing as described (Chung et al., 2013).

**Genome Sequencing**

The procedures for genome sequencing and analysis were based on those described in our previous work on phytoplasma genomes (Chung et al., 2013; Cho et al., 2019, 2020a, 2020b). All kits were used according to the manufacturer’s protocols and all bioinformatics tools were used with the default settings unless stated otherwise.

For whole genome shotgun sequencing, leaves from one artificially infected plant exhibiting typical symptoms (i.e., small leaves and witches’ broom) (Figure 1B) were collected for total genomic DNA extraction using the Wizard Genomic DNA Purification Kit (A1120; Promega, USA). For Illumina sequencing, the DNA sample was processed using the KAPA Library Preparation Kit (KK8234) and the Invitrogen SizeSelect Gels (G6610-02) to obtain ~550 bp fragments, followed by MiSeq 2x300 bp paired-end sequencing (v3 chemistry). For Oxford Nanopore Technologies (ONT) sequencing, the library was prepared using the ONT Ligation Kit (SQK-LSK109) without shearing or size selection, followed by MiniION sequencing (R9.4.1 chemistry) and Guppy v3.3.3 base-calling.

The first round of *de novo* assembly was performed using Velvet v1.2.10 (Zerbino and Birney, 2008) with only the Illumina reads. The draft assembly was queried against the NCBI non-redundant protein database (Benson et al., 2018) using BLASTX v2.10.0 (Camacho et al., 2009) to identify putative phytoplasma contigs. These contigs were used as the reference for extracting phytoplasma reads from the Illumina data set using BWA v0.7.17 (Li and Durbin, 2009) with an alignment score cutoff of 30 and from the ONT data set using
Minimap2 v2.15 (Li, 2018) with an alignment score cutoff of 1000. The extracted reads from both data sets were processed together using Unicycler v0.4.9b for the second round of de novo assembly. The putative phytoplasma contigs were identified by BLAST searches and used as the starting point for an iterative process of assembly improvement. In each iteration, all Illumina and ONT raw reads were mapped to the draft assembly as described. The mapping results were programmatically checked using SAMTOOLS v1.9 (Li et al., 2009) and manually inspected using IGV v2.5.0 (Robinson et al., 2011) to identify assembly errors and to close gaps.

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

Comparative Analysis

For comparative analysis with other representative phytoplasma genomes (Table 1), homologous gene clusters were identified using OrthoMCL (Li et al., 2003). Multiple
sequence alignments of homologous genes were prepared using MUSCLE v3.8.31 (Edgar, 2004) and visualized using JalView v2.11 (Waterhouse et al., 2009). Maximum likelihood phylogenies were inferred using PhyML v3.3 (Guindon and Gascuel, 2003) and visualized using FigTree v1.4.4. PHYLIP v3.697 (Felsenstein, 1989) was used for bootstrap analysis. Classification of the 16S RFLP group was performed using iPhyClassifier (Zhao et al., 2009).

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

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

**Results and Discussion**

**Genomic Characterization of NCHU2019**

The shotgun sequencing generated \(~4.7\) Gb of Illumina raw reads and \(~3.5\) Gb of ONT raw reads. As expected, most of the reads are derived from the plant host. By mapping to the finalized phytoplasma genome assembly, \(~1.0\)% of the Illumina reads and \(~5.1\)% of the ONT reads are derived from phytoplasma DNA, which corresponds to \(~51\)- and \(~80\)-fold
sequencing depth, respectively. The de novo genome assembly produced one complete
circular chromosome that is 769,143 bp in size with 23.3% G+C content; no plasmid was
found (Figure 2).

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

Other than the low G+C content and reduced gene content, phytoplasma genomes
are generally known to be repetitive, partly due to the presence of PMUs (Sugio and
Hogenhout, 2012). Interestingly, the genome of this strain is far more repetitive than other
phytoplasmas that have been characterized. On average, PMU-associated genes account for
~4.7% (Std. Dev. = 2.7%) of the genome size among those 19 representative phytoplasma genomes analyzed (Table 1). For strain NCHU2019, there are 117 PMU-associated genes that are organized into at least 13 distinct PMU regions (Figure 2) and account for 11% of the chromosome length. Additionally, a pair of chromosomal segments, each is ~75 kb in size, were found to be duplicated (positions: 315,975-391,140 and 391,446-466,612). These parts of the assembly were verified based on manual inspection of the ONT long reads mapping result. Together, these two duplicated segments and the 13 PMUs account for 25% of the chromosome length. The explanation for the high genome repetitiveness of strain NCHU2019 compared to other phytoplasmas is unclear.

Comparisons with Closely Related Phytoplasmas

Based on the established classification scheme and a previous study of 16S rRNA gene phylogeny (Davis et al., 2017), ‘Ca. P. luffae’ belongs to group 16SrVIII and is most closely related to ‘Ca. P. malaysianum’ (group 16SrXII; GenBank accession EU371934) (Nejat et al., 2013). These two species-level taxa have 95.9% sequence identity (i.e., 1,463/1,526 aligned nucleotides) in their 16S rRNA genes. However, no genome sequence is available for ‘Ca. P. malaysianum’ for comparative analysis.

Other than ‘Ca. P. malaysianum’, the next closest relatives of ‘Ca. P. luffae’ include phytoplasmas belonging to groups 16SrV (‘Ca. P. vitis’ and ‘Ca. P. ziziphi’), 16SrVI (‘Ca. P. sudamericanum’ and ‘Ca. P. trifolii’), and 16SrVII (‘Ca. P. fraxini’). Among these, one complete genome sequence (GenBank accession CP025121) is available for ‘Ca. P. ziziphi’ strain Jwb-nky (Wang et al., 2018a), which represents the most closely related lineage for comparative analysis (Figure 3). Comparison based on the 16S rRNA gene sequences indicated that ‘Ca. P. luffae’ NCHU2019 and ‘Ca. P. ziziphi’ Jwb-nky have 94.9% sequence
identity (i.e., 1,447/1,524 aligned nucleotides). For whole-genome comparison, only 49% of
the chromosomal segments can be mapped between these two strains and these segments
have < 80% ANI. Pairwise genome alignment indicated that the most of the conserved
regions correspond to PMUs (Figure 4), which further supports that the sequence
divergence between these two genomes is too high for nucleotide-level comparisons. The
lack of chromosome-level synteny conservation was commonly reported in previous
comparisons between complete genome sequences of phytoplasmas (Bai et al., 2006; Tran-
Nguyen et al., 2008; Andersen et al., 2013; Orlovskis et al., 2017; Wang et al., 2018a), even
for ‘Ca. P. asteris’ strains sharing > 99.9% 16S rRNA gene sequence identity and > 98.1%
genome-wide ANI (Cho et al., 2020a). These observations may be attributed to the strong
nucleotide composition biases, the high mutation accumulation rates, and the influence of
PMUs (Cho et al., 2020a).

Detailed Characterization of PMUs

To better understand the roles of PMUs in phytoplasma genome evolution, we
conducted detailed characterization of these mobile genetic elements. Among the 13 PMUs
identified in ‘Ca. P. luffae’ NCHU2019 (Figure 5A), 11 are considered as complete ones and
range from 14 to 18 kb in size. The remaining two (i.e., #6 and #7) are both 12 kb in size and
appear to be truncated. These PMUs are interspersed across the entire chromosome and
there is no obvious pattern of clustering (Figures 2 and 4). Two sets of PMUs (i.e., #4-6 and
#7-9) are associated with the two 75-kb repeat regions. Notably, one set of junctions for
these large repeats (chromosomal positions 315,975 and 466,612) are located inside PMU
#4 (positions 308,032..323,983) and PMU #9 (positions 458,604..475,923). This finding
suggests that homologous recombination facilitated by PMUs may have facilitated the
segmental duplication of this chromosome, which in turn further increased the PMU copy
number.

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

Based on the presence/absence pattern and order of eight PMU core genes defined
previously (Bai et al., 2006), we found that the PMUs in these representative phytoplasma
genomes can be classified into three major types (Figure 5B). Type A PMUs, in which tmk is
upstream of dnaB, are the most common ones that include PMUs found in phylogenetic
clusters I (e.g., ‘Ca. P. asteris’ and ‘Ca. P. australiense’) and III (e.g., ‘Ca. P. luffae’, ‘Ca. P.
ziziphi’, and ‘Ca. P. aurantifolia’). Type B PMUs are shorter, have tmk downstream of dnaB,
and are found only in ‘Ca. P. australiense’ that belongs to cluster I. Type C is the rarest type,
with only one representative found in ‘Ca. P. mali’ that belongs to cluster II, and has hflB
and dnaG located in between tmk and dnaB. In addition to the gene organization, molecular
phylogenies of *dnaB* and *tmk* also revealed divergence among homologs from different PMU types (Figure 6). These patterns provide further support for our PMU classification scheme. It is interesting to note that regardless of PMU copy numbers, most of the phytoplasmas with genome sequences available harbor only one single type of PMUs. ‘Ca. P. australiense’ is the only exception that harbors both type A and B PMUs. Future improvements in sampling more diverse phytoplasma genomes, particularly cluster II lineages, are necessary to provide a more comprehensive understanding of PMU diversity.

PMU and Phytoplasma Genome Size Variation

One notable observation regarding phytoplasma genomes is the extensive size variation at both between- and within-species levels (Table 1). Previous within-species comparisons suggested that PMU abundance is an important factor (Bai et al., 2006; Andersen et al., 2013). To further test if this pattern holds true for genus-level analysis, we performed regression analysis to examine the correlation between the combined length of PMU core genes and genome size. Strikingly, when all 19 representative phytoplasma genome assemblies were examined together, PMU gene length explains 79% of the variance in genome sizes (*r* = 0.89, *p* = 3.7e-07) (Figure 7). Due to the concern that draft assemblies cannot provide accurate information regarding these two metrics, we also performed regression analysis with only the 10 complete assemblies and obtained similar results (*r* = 0.87, *p* = 9.7e-04). Given that the evolution of bacterial genome sizes may be affected by multiple factors (Konstantinidis and Tiedje, 2004; Ochman and Davalos, 2006; Kuo et al., 2009; Novichkov et al., 2009; McCutcheon and Moran, 2012; Lo et al., 2016; Sabater-Muñoz et al., 2017; Weinert and Welch, 2017), the strong correlation between PMU abundance and phytoplasma genome size is surprising. Notably, as genome reduction appears to be a
common and recurring theme of symbiont evolution, the roles of PMU in genome expansion of some phytoplasmas require further investigation. Because PMUs are known to be associated with effector genes and likely can transfer horizontally between closely- (Cho et al., 2019; Seruga Music et al., 2019) or distantly-related phytoplasmas (Chung et al., 2013; Ku et al., 2013), the involvement of PMUs in phytoplasma effector gene content evolution is particularly important.

Effector Genes

An important feature of phytoplasmas is their ability to modulate host plant development through effectors, which are small secreted proteins (Sugio et al., 2011b; Sugio and Hogenhout, 2012). To date, four phytoplasma effectors have been experimentally characterized, including SAP05 (Gamboa et al., 2019; Huang and Hogenhout, 2019; Huang et al., 2021), SAP11/SWP1 (Bai et al., 2009; Sugio et al., 2011a; Lu et al., 2014; Chang et al., 2018; Wang et al., 2018c, 2018b), SAP54/PHYL1 (MacLean et al., 2011; Maejima et al., 2014; Orlovskis and Hogenhout, 2016), and TENGU (Hoshi et al., 2009; Sugawara et al., 2013; Minato et al., 2014). The expanded genome sequence availability allowed us to investigate the phylogenetic distribution of homologous effector genes among diverse phytoplasmas. The highly variable distribution patterns suggest that the effector gene content may have rapid turnover during the diversification of phytoplasmas (Figure 8). For example, at within-species level comparison among ‘Ca. P. asteris’ strains, the patterns of presence/absence and gene copy number are variable for three of these effectors as we reported previously (Cho et al., 2020a). At genus-level, TENGU appears to be restricted to and conserved among 16Srl lineages in cluster I (i.e., ‘Ca. P. asteris’ and ‘Ca. P. tritic’), while the other three effectors are variable. Based on these patterns, it is likely that TENGU originated in the most
recent common ancestor (MRCA) of 16Srl phytoplasmas. However, for the other three effectors, it is unclear if the MRCA of all extant phytoplasmas harbor these genes or not. If yes, then multiple independent gene losses are required to explain the distribution of these genes among extant phytoplasmas. Alternatively, multiple independent origins, likely mediated by PMU-mediated horizontal gene transfer (Chung et al., 2013; Cho et al., 2019; Seruga Music et al., 2019), are required to explain the gene distribution pattern. Another interesting observation is that several phytoplasmas do not possess any of these four effector genes. It is likely that these phytoplasmas harbor novel effector genes that are yet to be characterized and further investigation is required to obtain a more complete picture of phytoplasma effector diversity.

For closer inspection of these four known effectors, we performed multiple sequence alignments to examine the protein sequence divergence among homologs (Figure 9). Consistent with the expectation derived from their phylogenetic distribution patterns, the three effector genes found among diverse phytoplasmas have higher levels of sequence divergence compared to phylogenetically restricted TENGU homologs.

For copy number variation, ‘Ca. P. luffae’ NCHU2019 stands out as having the highest copy numbers for SAP11 and SAP54 (Figure 8). These homologous genes are all located within PMUs (Figure 5) and have nearly identical sequences (Figure 9), which suggest that recent intra-genomic PMU duplications are responsible for expansions in effector gene copy numbers. Similar patterns were observed for SAP11 and SAP54 homologs in ‘Ca. P. ziziphi’ Jwb-nky. Experimental characterization of the two SAP11 homologs in ‘Ca. P. ziziphi’ demonstrated that both can stimulate lateral bud outgrowth for witches’ broom symptoms when expressed in Nicotiana benthamiana (Zhou et al., 2021). Intriguingly, in phytoplasma-infected jujube plants, these two SAP11 homologs have different expression patterns.
among tissue types. These findings suggest that such gene duplication events may lead to
neofunctionalization or subfunctionalization, thus promoting the genetic diversity of
effectors.

Conclusion

In this work, we determined the complete genome sequence of an uncultivated ‘Ca.
P. luffae’ strain associated with the witches’ broom disease of loofah. This assembly
provides the first representative genome sequence for the 16SrVIII group of phytoplasmas
and improves the taxon sampling of these diverse plant-pathogenic bacteria. For
comparative genomics analysis conducted at genus level, we provided a global view of the
PMUs (i.e., phytoplasma-specific mobile genetic elements) and identified three major PMU
types that differ in gene organization and phylogenetic distribution. Importantly, statistical
analysis revealed that PMU abundance explains nearly 80% of the variance in phytoplasma
genome sizes, thus providing a quantitative estimate on the importance of these elements.
Finally, our investigation of effector genes highlighted the genetic diversity associated with
phytoplasma virulence and established the roles of PMUs in shaping such diversity.

Data Availability

This genome sequencing project was deposited in the NCBI under BioProject
PRJNA636624. The raw reads were deposited in the Sequence Read Archive (SRA) under the
accessions SRR11921288 (Illumina MiSeq paired-end reads) and SRR11921289 (ONT MinION
reads). The complete and annotated genome sequence of ‘Ca. P. luffae’ NCHU2019 was
deposited in GenBank under the accession CP054393.
Author Contribution

JYY and CHK conceptualized the study, acquired funding, and supervised the project.

JYY, YCC, and CMT provided the biological materials. CMT and STC coordinated the sequencing. CMT conducted the initial genome assembly. STC completed 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.
References

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

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

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

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

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

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

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

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

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

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

Cho, S.-T., Kung, H.-J., Huang, W., Hogenhout, S. A., and Kuo, C.-H. (2020a). Species boundaries and molecular markers for the classification of 16SrI phytoplasmas inferred by genome analysis. *Front. Microbiol.* 11, 1531.
doi:10.3389/fmicb.2020.01531.
Cho, S.-T., Lin, C.-P., and Kuo, C.-H. (2019). Genomic characterization of the periwinkle leaf yellowing (PLY) phytoplasmas in Taiwan. *Front. Microbiol.* 10, 2194. doi:10.3389/fmicb.2019.02194.

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

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

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

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

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

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

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

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

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

Hogenhout, S. A., and Seruga Music, M. (2009). "Phytoplasma genomics, from sequencing to comparative and functional genomics - what have we learnt?,” in *Phytoplasmas: genomes, plant hosts and vectors* (CABI), 19–36.
Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., et al. (2009). A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci.* 106, 6416–6421. doi:10.1073/pnas.0813038106.

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

Huang, W., MacLean, A. M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., et al. (2021). Parasite co-opts a ubiquitin receptor to induce a plethora of developmental changes. *bioRxiv*, 2021.02.15.430920. doi:10.1101/2021.02.15.430920.

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

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

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

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

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

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

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

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

Kube, M., Mitrovic, J., Duduk, B., Rabus, R., and Seemüller, E. (2012). Current view on phytplasma genomes and encoded metabolism. *Sci. World J.* 2012, 1–25. doi:10.1100/2012/185942.
Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A., et al. (2008). The linear chromosome of the plant-pathogenic mycoplasma “Candidatus Phytoplasma mali.” BMC Genomics 9, 306. doi:10.1186/1471-2164-9-306.

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

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

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

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

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

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

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

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

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

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

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

Lo, W.-S., Huang, Y.-Y., and Kuo, C.-H. (2016). Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol. Rev. 40, 855–874. doi:10.1093/femsre/fuw028.
Lowe, T., and Eddy, S. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi:10.1093/nar/25.5.955.

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

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

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

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

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

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

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

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

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

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

Orlovskis, Z., and Hogenhout, S. A. (2016). A bacterial parasite effector mediates insect vector attraction in host plants independently of developmental changes. *Front. Plant Sci.* 7, 885. doi:10.3389/fpls.2016.00885.
Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H.-Y., Wei, W., Suzuki, S., et al. (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* 36, 27–29. doi:10.1038/ng1277.

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

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

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

Seruga Music, M., Samarzija, I., Hogenhout, S. A., Haryono, M., Cho, S.-T., and Kuo, C.-H. (2019). The genome of *Candidatus Phytoplasma solani* strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Syst. Appl. Microbiol.* 42, 117–127. doi:10.1016/j.syapm.2018.10.008.

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

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

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

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

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

The IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma taxonomy group (2004). “*Candidatus Phytoplasma*”, a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* 54, 1243–1255. doi:10.1099/ijs.0.02854-0.
Tran-Nguyen, L. T. T., Kube, M., Schneider, B., Reinhardt, R., and Gibb, K. S. (2008). Comparative genome analysis of “Candidatus Phytoplasma australiensis” (subgroup \textit{tu}f-Australia I; \textit{rp}-A) and “Ca. Phytoplasma asteris” strains OY-M and AY-WB. \textit{J. Bacteriol.} 190, 3979–3991. doi:10.1128/JB.01301-07.

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

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

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

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

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

Zhao, Y., and Davis, R. E. (2009). Construction of an interactive online phytoplasma classification tool, \textit{iPhyClassifier}, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). \textit{Int. J. Syst. Evol. Microbiol.} 59, 2582–2593. doi:10.1099/ijs.0.010249-0.

Zhao, Y., Wei, W., Lee, I.-M., Shao, J., Suo, X., and Davis, R. E. (2009). The agent associated with blue dwarf disease in wheat represents a new phytoplasma taxon, \textit{Candidatus Phytoplasma tritici}. \textit{Int. J. Syst. Evol. Microbiol.} doi:10.1099/ijs.0.004604.

Zhao, Y., Wei, W., Lee, I.-M., Shao, J., Suo, X., and Davis, R. E. (2021). Criteria for phytoplasma 16Sr group/subgroup delineation and the need of a platform for proper registration of new groups and subgroups. \textit{Int. J. Syst. Evol. Microbiol.} 66, 2121–2123. doi:10.1099/ijsem.0.000999.
Zhou, J., Ma, F., Yao, Y., Deng, M., Chen, M., Zhang, S., et al. (2021). JWB phytoplasma effectors SJP1 and SJP2 induce lateral bud outgrowth by repressing the ZjBRC1-controlled auxin efflux channel. *Plant Cell Environ.* n/a. doi:10.1111/pce.14141.
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.

Figure Legends

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

Figure 2. Genome map of ‘Candidatus Phytoplasma luffae’ NCHU2019. Rings from outside in: (1) Scale marks (kb). (2 and 3) Coding sequences on the forward and reverse strand, respectively. Color-coded by functional categories. (4) Genes associated with potential mobile units (PMUs), color-coded by annotation (see Figure 5). Gene clusters that represent individual PMUs are labeled by orange lines. Genes encode putative secreted proteins are in red. (5) GC skew (positive: dark gray; negative: light gray); (6) GC content (above average: dark gray; below average: light gray). One high GC peak located in the 531 to 544 kb region corresponds to two adjacent rRNA operons.
**Figure 3.** Evolutionary relationships among representative phytoplasmas with genome sequences available. The maximum likelihood phylogeny was inferred using a concatenated alignment of 80 conserved single-copy genes (25,340 aligned amino acid sites). All internal nodes received >80% bootstrap support based on 1,000 re-sampling. Strains with complete genome assemblies available are highlighted in bold. The number in parentheses following the strain name indicates the genome size (in kb). The 16S rRNA gene (16Sr) group assignments are labeled on the right. The three major phylogenetic clusters are indicated by colored backgrounds (I: blue; II: green; III: yellow). *Acholeplasma laidlawii* is included as the outgroup.

**Figure 4.** Pairwise genome alignments. The genome of ‘*Ca. P. luffae*’ NCHU2019 was used as the reference for comparison with (A) itself and (B) ‘*Ca. P. ziziphi*’ Jwb-nky. Matches on the same strand and the opposite strand are indicated in red and blue, respectively. Potential mobile units (PMUs) are illustrated as orange arrows on both axes.

**Figure 5.** Gene organization of potential mobile units (PMUs). Each individual PMU is labeled by the strain name and a numerical identifier. Background colors for the PMU identifiers indicate the three phylogenetic clusters of phytoplasmas (I: blue; II: green; III: yellow). Genes are drawn to scale and color-coded according to annotation. Homologs of putative effectors identified in the ‘*Ca. P. asteris*’ AYWB genomes are labeled, those with experimental evidence (i.e., SAP05, SAP11, and SAP54) are highlighted in bold. (A) All PMUs in ‘*Ca. P. luffae*’ NCHU2019 and ‘*Ca. P. ziziphi*’ Jwb-nky. Grouped by genomes. (B) Representative PMUs from selected phytoplasmas. Grouped by PMU types.
Figure 6. Maximum likelihood phylogenies of PMU core genes (A) \textit{tmk} (228 aligned amino acid sites). (B) \textit{dnaB} (531 aligned amino acid sites). In both panels, a non-PMU homolog is included as the outgroup. Numbers next to internal branches indicate the bootstrap support levels based on 1,000 re-sampling. Background colors for the gene identifiers indicate the three phylogenetic clusters of phytoplasmas (I: blue; II: green; III: yellow).

Figure 7. Correlation between the combined length of PMU core genes and genome size. Strains with complete and draft genome assemblies are indicated by filled and open circles, respectively. The linear regression line was based on all available assemblies.

Figure 8. Phylogenetic distribution of known effector genes. Gene presence and absence are indicated by filled (with copy numbers) and empty circles, respectively. The cladogram on the left is based on Figure 3. Strains with complete genome assemblies available, which have a higher confidence level for inferring gene presence and absence, are highlighted in bold.

Figure 9. Protein sequence alignments of four characterized phytoplasma effectors. Shaded background colors indicate the levels of sequence conservation. (A) SAP05. (B) SAP11/SWP1. (C) SAP54/PHYL1. (D) TENGU.
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., \textit{tra5, dnaB, dnaG, tmk, hflB, himA, ssb}, and \textit{rpoD}); other genes such as those encode hypothetical proteins or putative secreted proteins are excluded. \textit{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_000509318.1 | 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 | GCF_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 | XIX | GCF_007821455.1 | 16 contigs | 474,136 | 392 | 15 | 11,685 |
| \textit{Acholeplasma laidlawii} PG-8A | N/A | GCF_000018785.1 | Complete | 1,496,992 | 1,378 | 0 | 0 |
