INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE

Rodrigues Jardim *et al.*, *Int. J. Syst. Evol. Microbiol.* 2023;73:005977

DOI 10.1099/ijsem.0.005977

MICROBIOLOGY SOCIETY

OPEN ACCESS

The observation of taxonomic boundaries for the 16SrII and 16SrXXV phytoplasmas using genome-based delimitation

Bianca Rodrigues Jardim[1,2,*], Lucy T. T. Tran-Nguyen[3], Cherie Gambley[4], Abdullah M. Al-Sadi[5], Ali M. Al-Subhi[5], Xavier Foissac[6], Pascal Salar[6], Hong Cai[7], Jun-Yi Yang[8,9], Richard Davis[10], Lynne Jones[10], Brendan Rodoni[1,2] and Fiona E. Constable[1,2]

Abstract

Within the 16SrII phytoplasma group, subgroups A–X have been classified based on restriction fragment length polymorphism of their 16S rRNA gene, and two species have been described, namely '*Candidatus* Phytoplasma aurantifolia' and '*Ca*. Phytoplasma australasia'. Strains of 16SrII phytoplasmas are detected across a broad geographic range within Africa, Asia, Australia, Europe and North and South America. Historically, all members of the 16SrII group share ≥97.5% nucleotide sequence identity of their 16S rRNA gene. In this study, we used whole genome sequences to identify the species boundaries within the 16SrII group. Whole genome analyses were done using 42 phytoplasma strains classified into seven 16SrII subgroups, five 16SrII taxa without official 16Sr subgroup classifications, and one 16SrXXV-A phytoplasma strain used as an outgroup taxon. Based on phylogenomic analyses as well as whole genome average nucleotide and average amino acid identity (ANI and AAI), eight distinct 16SrII taxa equivalent to species were identified, six of which are novel descriptions. Strains within the same species had ANI and AAI values of >97%, and shared ≥80% of their genomic segments based on the ANI analysis. Species also had distinct biological and/or ecological features. A 16SrII subgroup often represented a distinct species, e.g., the 16SrII-B subgroup members. Members classified within the 16SrII-A, 16SrII-D, and 16SrII-V subgroups as well as strains classified as sweet potato little leaf phytoplasmas fulfilled criteria to be included as members of a single species, but with subspecies-level relationships with each other. The 16SrXXV-A taxon was also described as a novel phytoplasma species and, based on criteria used for other bacterial families, provided evidence that it could be classified as a distinct genus from the 16SrII phytoplasmas. As more phytoplasma genome sequences become available, the classification system of these bacteria can be further refined at the genus, species, and subspecies taxonomic ranks.

DATA SUMMARY

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

Author affiliations: [1]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia; [2]Agriculture Victoria Research, Department of Energy, Environment and Climate Action, AgriBio, Bundoora, Victoria, Australia; [3]Plant Health Australia, Deakin, Australian Capital Territory, Australia; [4]Horticulture and Forestry Science, Department of Agriculture and Fisheries Maroochy Research Facility, Nambour, Queensland, Australia; [5]Department of Plant Sciences, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat, Oman; [6]University of Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, 33140, Bordeaux, Villenave d'Ornon, France; [7]The Key Laboratory for Plant Pathology, Yunnan Agricultural University, Kunming 650201, PR China; [8]Institute of Biochemistry, National Chung Hsing University, Taichung 402, Taiwan, ROC; [9]Advanced Plant Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan, ROC; [10]Northern Australia Quarantine Strategy, Department of Agriculture, Fisheries and Forestry, Canberra, Australian Capital Territory 2601, Australia.

***Correspondence:** Bianca Rodrigues Jardim, 20171100@student.latrobe.edu.au; bianca.rodriguesjardim@agriculture.vic.gov.au; b.rodriguesjardim.agric@gmail.com

Keywords: AAI; ANI; genus; *Mollicutes*; peanut witches'-broom group; phylogenomics; phytopathogen; species; species delimitation; subspecies.

Abbreviations: AAI, average amino acid identity; AF, alignment fraction; ANI, average nucleotide identity; BoLL, *Bonamia* little leaf; CAWB, cocky apple witches'-broom; gCF, gene concordance factor; MLSA, multilocus sequence analysis; MLSA, multilocus sequence analysis; PLL, pigeon pea little leaf; RFLP, restriction fragment length polymorphism; sCF, site concordance factor; SPLL, sweet potato little leaf; TBB, tomato big bud; UDI, unique dual index; WTWB, weeping teatree witches'-broom.

Four supplementary tables and two supplementary figures are available with the online version of this article.

005977 © 2023 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

Impact Statement

Species delimitation is important because our understanding of an organism's biology and taxonomy are interdependent. Phytoplasmas are a group of phytopathogenic bacteria that have not yet been cultured *in vitro*. This has impeded the development of a cohesive and taxonomically significant system for these bacteria. Instead, their taxonomy has largely focused on the 16S rRNA gene, which does not provide enough resolution to determine whether two individuals are distinct species. The work presented here has identified species and subspecies boundaries within the 16SrII phytoplasma taxonomic group, and genus boundaries between the 16SrII taxon and the closely related 16SrXXV-A taxon used as an outgroup. Six novel species are identified within the 16SrII taxon and described based on the high-resolution analyses afforded by whole genome sequences. The species descriptions are verified using multiple sequence data analysis approaches and are supported by analysing multiple individuals within the delimited species in some cases. This work contributes towards refining the taxonomy within the '*Candidatus* Phytoplasma' provisional genus, and expands upon the species delimitation criteria for the phytoplasmas. The work also significantly contributes to available phytoplasma sequence data, with draft genome sequences of up to 15 individuals within a species deposited in public repositories.

INTRODUCTION

Phytoplasmas are wall-less, insect-vectored bacteria that are associated with hundreds of plant diseases globally [1]. These bacteria inhabit the phloem tissue of plants and colonize the gut, haemocoel, salivary glands, and many other tissue types in their insect hosts [2]. Culture-dependent characteristics for bacterial taxonomic delineation such as morphology, chemistry, and physiology are not possible for phytoplasmas, for which an effective culture system is yet to be developed. Instead, phytoplasmas have been classified as a distinct taxon within the class *Mollicutes*, and are assigned to the provisional genus '*Candidatus* Phytoplasma' based on comparative DNA sequence analysis of the 16S rRNA gene of phytoplasmas, related mollicutes, and other prokaryotes [3, 4].

Species delimitation within the provisional '*Ca*. Phytoplasma' genus is traditionally based on the 16S rRNA gene. In 2004, the first guidelines for novel '*Ca*. Phytoplasma' species descriptions were published [5]. These guidelines specified a 97.5% 16S rRNA nucleotide sequence similarity threshold for a novel species description (Rule b) based on a >1200 bp region of the gene (Rule a). These guidelines also stated that a novel species could be described in cases where the 16S rRNA genes of two taxa share more than 97.5% sequence similarity but the phytoplasmas have unique vectors, distinct host ranges or symptomology in a host, and unique molecular or serological characteristics (Rule c, i–iii) [5]. In 2022, the guidelines for novel '*Ca*. Phytoplasma' species descriptions were revised [6], and state that the new sequence similarity threshold of the 16S rRNA gene is 98.65%, and at least 95% of the full 16S rRNA sequence must be used in these analyses (>1500 bp region). Rule c from the 2004 guidelines is still in place in the updated guidelines for cases where the sequence similarity of the 16S rRNA gene for two phytoplasmas exceeds the 98.65% threshold. The updated guidelines also specified threshold values for nucleotide sequence analyses of additional housekeeping gene regions based on previous '*Ca*. Phytoplasma' species descriptions, as well as whole genome average nucleotide identity (ANI) values based on the threshold set for bacterial species (within-species ANI of above 95 or above 96%) [7, 8].

A second classification system exists for the phytoplasmas, with the designation of groups and subgroups based on sequence similarity and restriction fragment length polymorphism (RFLP) of the 16S rRNA gene [5, 9]. The 2004 guidelines specify that phytoplasmas which share >97.5% nucleotide sequence similarity at their 16S rRNA gene are considered members of the same group [5]. By using a specific set of 17 restriction enzymes in separate reactions in the laboratory or *in silico*, the individual RFLP patterns, and associated similarity coefficient value, would determine whether the taxon belonged to an existing or a novel subgroup [5, 9]. However, the taxonomic rank of different subgroups is unknown (e.g., genus, species, or subspecies) due to the low resolution afforded by the nucleotide sequence of the 16S rRNA gene [10, 11]. The resolution is even lower when RFLP is applied as fewer nucleotide positions are considered in these analyses [4]. Phytoplasmas also have two 16S rRNA genes in their genome, with some showing sequence heterogeneity between copies and resulting in conflicting subgroup assignments based on RFLP [12]. There are no mentions or recommendations for group/subgroup classifications in the updated '*Ca*. Phytoplasma' species description guidelines [6].

The 16SrII phytoplasma group is one of the most diverse phytoplasma groups described to date; there are 24 defined subgroups (16SrII-A–16SrII-X [13]) and two delineated species, namely '*Ca*. Phytoplasma aurantifolia' (subgroup 16SrII-B [14]) and '*Ca*. Phytoplasma australasia' (subgroup 16SrII-D [15]). Several phytoplasma strains have not been formally assigned a subgroup designation within the 16SrII group although they meet the requirements, including the sweet potato little leaf (SPLL) [16, 17], pigeon pea little leaf (PLL) [18], cocky apple witches'-broom (CAWB) [19], and *Bonamia* little leaf (BoLL) [20] phytoplasmas, which are all found in Australia. Phytoplasmas within the 16SrII group have been detected in a large range of hosts over an extensive geographic range, spanning Africa, Asia, Australia, Europe, South America, and the Middle East [21–24]. Globally, confirmed and putative vectors of the 16SrII phytoplasma strains often include leafhoppers, with species within the genera

Austroagallia, *Batracomorphus*, *Empoasca*, *Hishimonus*, and *Orosius* commonly implicated [20, 25–27]. The vectors of many 16SrII phytoplasma subgroups, however, remain to be determined or confirmed.

There is a need to accurately delimit the species boundaries of members within the 16SrII group of phytoplasmas for which competent vectors are largely undetermined. While the 16S rRNA gene is the most sequenced molecular marker for bacterial taxonomy and is abundant in public databases compared to alternative markers [28], taxonomy based on multilocus sequence analysis (MLSA) and whole genome sequences provide more resolution to define the taxonomic ranks at the genus, species, and subspecies level [29–32]. Here, we applied the recently updated phytoplasma species description recommendations [6] as well as genome-based delimitation criteria for high resolution species delimitation within the 16SrII phytoplasma group. Genome-based criteria included phylogenomic analyses to reconstruct putative evolutionary relationships, as well as whole genome average nucleotide identity (ANI) and average amino acid identity (AAI) using within-species threshold values specified for culturable bacteria (i.e. >95% ANI and AAI values) [8, 30]. To this end, we sequenced 36 phytoplasma genomes of strains collected in Australia, Burkina Faso, China, Oman, Taiwan, and Thailand, including the original strain from which '*Ca*. Phytoplasma aurantifolia' was described [14], and one 16SrXXV-A phytoplasma strain collected in Australia. Publicly available phytoplasma genome sequences were retrieved for six individual phytoplasma strains within the 16SrII group collected from Australia, India, and Taiwan. Together, this dataset represented 11 subgroups or subgroup variants within the 16SrII phytoplasma group and an outgroup taxon. A total of eight distinct 16SrII phytoplasma species were delimited (six novel descriptions) based on the whole genome criteria in this study.

METHODS

Sample collection, DNA extraction, and DNA concentration estimations

Phytoplasma strains from previous studies performed outside Australia [33–35], from the phytoplasma collection held at INRAE, Bordeaux, France, and from strains collected prior to 2019 in Australia were supplied as total DNA extracts (Table S1). Fresh symptomatic plant samples were also collected in Australia after March 2019 (Table S1) and total DNA was extracted using a modified CTAB-DNeasy protocol [36]. All fresh samples were freeze dried at −50 °C for at least 72 h using the FreeZone 2.5 l benchtop freeze dry system (Labconco). Freeze-dried material of representative samples were deposited at the Victorian Plant Pathology Herbarium (VPRI) in Bundoora, Victoria, Australia.

DNA quality and concentrations were estimated using a NanoDrop spectrophotometer and a Qubit 2.0 fluorometer (Invitrogen) with the Qubit 1X dsDNA HS Assay Kit (Invitrogen), respectively. All total DNA extracts were stored at −80 °C until use.

Phytoplasma screening and identification by PCR, Sanger sequencing, and *i*PhyClassifier

Generic bacterial 16S rRNA primers [37] were used in PCR to determine whether PCR inhibitors were present in the DNA extracts. To detect and identify the phytoplasma taxon infecting each sample, universal phytoplasma P1/P7 and R16F2n/m23sr primers were used in PCR according to [38], and amplicons of the expected size were sent to Macrogen (Seoul, Republic of Korea) for purification and direct Sanger sequencing in both directions. The identities of the Sanger sequenced amplicons were confirmed using BLASTn analyses [39]. BLASTn analyses and alignments with the representative sequences of each subgroup as mentioned in [40–43] were used to determine the 16Sr subgroup identity of each strain analysed in this study based on bit-scores and sequence similarity, respectively. The 16S rRNA sequences were also submitted to *i*PhyClassifier for an additional phytoplasma subgroup assignment when the 16S rRNA amplicon sequences met the length requirements of the online program ([9] https://plantpathology.ba.ars.usda.gov/cgi-bin/resource/iphyclassifier.cgi, last accessed October 2022).

Illumina sequencing and read quality filtering

Total DNA was normalized to 10 ng or 100 ng and used as input for library preparation. Library preparation was done using either the NEXTFLEX Rapid XP DNA-Seq Kit (PerkinElmer) with the compatible Unique Dual Index (UDI) barcodes [NEXTFLEX 384 UDI Barcodes version 19.06 (PerkinElmer) or NEXTFLEX UDI Barcodes 1–1536 (3072 rxn; PerkinElmer)] or using the Nextera DNA Flex Library Preparation Kit with the IDT for Nextera DNA Unique Dual Indexes (Illumina). The manufacturer protocols were followed to produce libraries of fragments between 250 and 500 bp in size, depending on the chosen sequencing platform. The 2200 TapeStation with HSD1000 ScreenTapes (Agilent Technologies) were used to estimate the molarity of all libraries. The final libraries were pooled and then size selected to remove fragments smaller than 200 bp using the ProNex Size Selective Purification System (Promega). Most libraries were sequenced with the NovaSeq SP Flow Cell (2×250 bp), while a few were sequenced using the NovaSeq S1 Flow Cell (2×150 bp) or the HiSeq 2500 Dual Flow Cell (2×150 bp) (Table S1). All libraries sequenced on the NovaSeq 6000 platform were treated with the Illumina Free Adapter Blocking Reagent to reduce aberrant sequencing results.

Bioinformatic analyses

The quality control of the Illumina sequence reads, their metagenomic assembly, phytoplasma-derived contig identification and retrieval was done according to our previous pipeline [44] which implements FastP [45], metaSPAdes 3.15.2 [46, 47], BLASTn

analyses using BLAST+ version 2.11.0 [39], and a custom *grep* script. The phytoplasma genome sequences were imported to Geneious Prime version 2022.2.1 (www.geneious.com) to remove contigs shorter than 1000 bp, and to estimate the genome size and G+C content. The phytoplasma genome sequences were analysed in metaQUAST [48] to estimate the genome N50 values. Protein-coding, tRNA, and rRNA genes were annotated and counted using Prokka with default parameters [49] and specifying RNAmmer for rRNA annotation [50]. The 16S rRNA sequences obtained by Sanger sequencing and from the phytoplasma genome assemblies were aligned to ensure that both sequences matched prior to further analyses according to proposed minimum standards of using whole genome sequence data for prokaryote taxonomy [51]. Unless sequence data was available for fewer than six strains in a 16SrII subgroup, only genome sequences with genome N50 values >20000 bp, approximately 30 contigs above 1000 bp in size, and encoding more than 19 tRNA genes were selected for further analyses (Table S1). Reference sequences for each taxon was chosen when they met the above criteria in addition to having more than 100× read coverage as estimated using BBSplit implemented in the BBMap version 38.61b software suite [52].

Sequence analysis of the 16S rRNA gene

The 16S rRNA gene was extracted from the RNAmmer analysis of the 42 phytoplasma genome sequences analysed in this study (Table S1) and were trimmed to the length of '*Candidatus* Phytoplasma asteris' (*Oenothera* phytoplasma 86–7; 1535 bp; GenBank accession no. M30790.1) according to the updated '*Candidatus* Phytoplasma' description guidelines [6] (Table S2).

All sequences extracted from the genome sequences or from publicly sequence repositories from the original descriptions [40–43] that met the length criteria specified in the updated guidelines were used to determine the pairwise nucleotide sequence similarity (Table S3). When the publicly available 16S rRNA gene sequences were shorter than the recommended length [6], they were only used in the maximum-likelihood analyses and were not used in estimates of nucleotide sequence similarity for the gene. In cases where the 16S rRNA gene was identical in sequence for more than three samples obtained in this study, only the one representative sequence for each taxon was selected for phylogenetic analyses to improve the ease of interpretation. All alignments were performed using MAFFT [53]. Alignments were used for sequence similarity estimates and as input for the reconstruction of maximum-likelihood trees. For maximum-likelihood tree reconstruction, the alignment of 16S rRNA gene sequences from subgroup representatives (genome-extracted and publicly available) was submitted to the Smart Model Selection tool [54] to determine the best-fit substitution model for the maximum-likelihood analyses. Maximum-likelihood analyses were performed using RAxML version 8.2.11 [55] implemented in Geneious Prime version 2022.2.1 (Figs 1 and S1, available in the online version of this article).

All representative 16S rRNA gene sequences that were of sufficient length were submitted to *i*PhyClassifier [9] to confirm 16Sr subgroup classification based on the *in silico* RFLP similarity coefficient.

Phylogenomic analysis using single-copy orthologs (SCOs)

SCOs were identified using OrthoFinder 2 [56, 57] and their functional annotations recorded (Table S4). Each of the SCO amino acid sequences were aligned using MAFFT [53]. The SCO amino acid sequence alignments were concatenated in IQ-TREE version 2.0.5 [58], and -m TEST was specified to determine and select the optimum substitution model for each partition using the Bayesian information criterion for use in the inference of the maximum likelihood tree [59, 60]. Bootstrapping with 1000 replicates was specified [61]; however, since the bootstrap support approaches 100% as the number of sequences are increased [62], bootstraps are not a suitable statistic for phylogenomic analyses. As such, the gene concordance factor (gCF) and site concordance factor (sCF) were also calculated using IQ-TREE to determine the proportion of genes and amino acid sites, respectively, that are concordant with specific branches in the concatenated tree generated by IQ-TREE version 2.0.5.

Overall genome relatedness index analysis

ANI

The ANI values between phytoplasma genome sequences were calculated using FastANI version 1.33 [8]. FastANI also provides the number of genomic segments shared between two genomes. The mean ANI and percent of shared genomic segments (alignment fraction, AF) in reciprocal genome sequence comparisons were calculated as described according to [63].

AAI

As an additional comparison and to improve the resolution of more distantly related taxa that might share a lower percent of shared genomic segments during ANI analyses, the AAI values between genome sequences were calculated at http://enve-omics.ce.gatech.edu/aai/ using Prokka-annotated amino acid sequences for all genomes [30]. One or two representative whole genome sequences were chosen for these AAI analyses, under the following selection criteria: (i) if whole genome sequences were available for the reference strain (equivalent of type strain for culturable bacteria) of a subgroup, it was selected for the comparison, (ii) if the genome assembly of the reference strain was poor (i.e. low N50, high contig number), an additional, higher quality genome assembly was selected as the secondary representative sequence for the taxon, and (iii) in the absence of whole genome sequences

Table 1. Details, comparisons, and comments of the subgroup designations of this study compared to those generated using the *i*PhyClassifier-based in silico restriction fragment length polymorphism (RFLP) similarity coefficient for all the phytoplasma genomes analysed in this study as well as four reference GenBank accession sequences. Reference strains for each taxon are highlighted in bold font

Phytoplasma strains in this study	Original 16Sr group/ subgroup classification based on literature*	16Sr group/ subgroup classification based on *i*PhyClassifier	*i*PhyClassifier-generated message and similarity coefficients
BAWM-THA-CLP, **NCHU2014**	16SrII-A†	16SrII-V	**For the 16S rRNA gene of all 16SrII-A strains used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group II, subgroup V (GenBank accession: KY568717). The phytoplasma under study is a member of 16SrII-V. **For the 16S rRNA gene from the original description of 16SrII-A phytoplasma (GenBank accession no. L33765):** L33765 – The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group II, subgroup V (GenBank accession: KY568717). The phytoplasma under study is a member of 16SrII-V.
BAWM-OMN-P210, BAWM-OMN-P75, **BAWM-OMN-WBDL**	16SrII-B‡	16SrII-C var	**For the 16S rRNA gene of all 16SrII-B strains used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is most similar to the reference pattern of the 16Sr group II, subgroup C (GenBank accession: AJ293216), with a pattern similarity coefficient of 0.99. The phytoplasma under study is a variant of 16SrII-C. **For the 16S rRNA gene from the original description of '*Ca.* Phytoplasma aurantifolia' (GenBank accession no. U15442):** U15442 – The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group II, subgroup B (GenBank accession: U15442). The phytoplasma under study is a member of 16SrII-B.
BAWM-CHN-EU009550, BAWM-CHN-EU009551, **PR34**	16SrII-C§	16SrII-C	**For the 16S rRNA gene of all 16SrII-C strains used in this study and strain (GenBank accession no. AJ293216):** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group II, subgroup V (GenBank accession: KY568717). The phytoplasma under study is a member of 16SrII-V. **For the 16S rRNA gene from strain FBP (GenBank accession no. X83432):** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is most similar to the reference pattern of the 16Sr group II, subgroup C (GenBank accession: AJ293216), with a pattern similarity coefficient of 0.99. The phytoplasma under study is a variant of 16SrII-C.
BAWM-003, BAWM-055, BAWM-058, BAWM-176, BAWM-181, BAWM-206, BAWM-215, BAWM-246, BAWM-260, o4P, BAWM-OMN-P26, BAWM-OMN-P3, BAWM-AUS- TBB, **PR08,** SS02	16SrII-D	16SrII-D	**For the 16S rRNA gene of all 16SrII-D strains used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group II, subgroup D (GenBank accession: Y10097). The phytoplasma under study is a member of 16SrII-D.
BAWM-BFA-CoWB	16SrII-F	16SrII-C var	**For the 16S rRNA gene of strain BAWM-BFA-CoWB used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is most similar to the reference pattern of the 16Sr group II, subgroup C (GenBank accession: AJ293216), with a pattern similarity coefficient of 0.99. The phytoplasma under study is a variant of 16SrII-C.
BAWM-TWN	16SrII-V	16SrII-V	**For the 16S rRNA gene of strain BAWM-TWN used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group II, subgroup V (GenBank accession: KY568717). The phytoplasma under study is a member of 16SrII-V.
BAWM-OMN-P53	16SrII-W‖	16SrII-C var	**For the 16S rRNA gene of strain BAWM-OMN-P53 used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is most similar to the reference pattern of the 16Sr group II, subgroup C (GenBank accession: AJ293216), with a pattern similarity coefficient of 0.99. The phytoplasma under study is a variant of 16SrII-C.
BAWM-225	Subgroup unassigned (BoLL)	New 16SrII subgroup	**For the 16S rRNA gene of strain BAWM-225 used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is different from the reference patterns of all previously established 16Sr groups/subgroups. The most similar is the reference pattern of the 16Sr group II, subgroup D (GenBank accession: Y10097), with a similarity coefficient of 0.93, which is less than or equal to 0.97. This strain may represent a new subgroup within the 16Sr group II.
BAWM-156b	Subgroup unassigned (CAWB)	**New 16SrII subgroup**	**For the 16S rRNA gene of strain BAWM-156b used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is different from the reference patterns of all previously established 16Sr groups/subgroups. The most similar is the reference pattern of the 16Sr group II, subgroup D (GenBank accession: Y10097), with a similarity coefficient of 0.94, which is less than or equal to 0.97. This strain may represent a new subgroup within the 16Sr group II.
BAWM-025, BAWM-026, **BAWM-027,** BAWM-028, BAWM-322	Subgroup unassigned (PLL)	16SrII-D var	**For the 16S rRNA gene of all PLL strains used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is most similar to the reference pattern of the 16Sr group II, subgroup D (GenBank accession: Y10097), with a pattern similarity coefficient of 0.98. The phytoplasma under study is a variant of 16SrII-D.

Continued

Table 1. Continued

Phytoplasma strains in this study	Original 16Sr group/ subgroup classification based on literature*	16Sr group/ subgroup classification based on *i*PhyClassifier	*i*PhyClassifier-generated message and similarity coefficients
BAWM-082, BAWM-174, BAWM-250, BAWM-301, BAWM-312, BAWM-330, BAWM-333, **o7C**	Subgroup unassigned (SPLL)	16SrII-V var	**For the 16S rRNA gene of all SPLL strains used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is most similar to the reference pattern of the 16Sr group II, subgroup V (GenBank accession: KY568717), with a pattern similarity coefficient of 0.98. The phytoplasma under study is a variant of 16SrII-V.
BAWM-155c	16SrXXV-A (WTWB)	16SrXXV-A	**For the 16S rRNA gene of strain BAWM-155c used in this study:** The virtual RFLP pattern derived from the query 16S rDNA F2nR2 fragment is identical (similarity coefficient 1.00) to the reference pattern of 16Sr group XXV, subgroup A (GenBank accession: AF521672). The phytoplasma under study is a member of 16SrXXV-A.

*The 16Sr subgroups are listed based on [40–43] and match the designations in Table S1 in this study.
†All 16SrII-A strains listed in the *i*PhyClassifier database (https://plantpathology.ba.ars.usda.gov/cgi-bin/resource/phytoclass.cgi, last accessed 13 October 2022 and last updated 26 August 2015) are classified as 16SrII-V members (GenBank accession nos: GU113161, GU113160, L33765, KP027534) or variants (GenBank accession no: JQ067649), or as new 16SrII subgroups (GenBank accession nos.: GU004373, L33770, EF193357, KF923393, KF923391). The 16SrII-V phytoplasma (GenBank accession no. KY568717) according to [41] does not produce a distinct RFLP pattern from the 16SrII-A phytoplasma (GenBank accession no. L33765) when RLFP patterns are generated using the *i*PhyClassifer program (data not shown). However, the original 16SrII subgroup classification was kept to refer to the different 16SrII-A and 16SrII-V strains throughout this study.
‡No other sequences on the NCBI have 100% identity to the 16S rRNA sequence of the 16SrII-B original description (GenBank no. U15442) by Zreik *et al.* [14]. Additionally, we obtained PCR and whole genome sequence data of the same strain described in the original description. The original sequence and ours produced from the same strain are not 100% identical, likely due to minor sequencing errors in the original sequence used to describe '*Candidatus* Phytoplasma aurantifolia' due to the sequencing methodology of the time.
§We considered subgroup classification according to [40–43], apart from the 16SrII-C representative strain Faba Bean Phyllody (FBP) phytoplasma (GenBank accession no. X83432), as the sequence contains an 'N' (an unresolved, ambiguous nucleotide). No sequences on the NCBI have 100% nucleotide sequence identity with the FBP phytoplasma (GenBank accession no. X83432). We therefore used the *i*PhyClassifier 16SrII-C representative, which considers the cactus witches'-broom phytoplasma strain YN01 (GenBank accession no. AJ293216) as the reference strain.
‖The *in silico* RFLP subgroup assignment does not match the identified phytoplasma subgroup due to all the 16S rRNA gene sequences available for 16SrII-W being too short for *i*PhyClassifier. These short sequences are, therefore, not stored in the *i*PhyClassifier database for comparison (https://plantpathology.ba.ars.usda.gov/cgi-bin/resource/phytoclass.cgi?strainsearchbox=&countrysearchbox=&diseasesearchbox=&hostsearchbox=&group_popup=II&subgroup_popup=all&submit_group_subgroup_popup=search accessed 13 October 2022). BLASTn, sequence similarity, and phylogenetic analyses in this study, however, support that the sequence under investigation matches those of 16SrII-W phytoplasmas as described by Al-Subhi *et al.* [42].

for a reference strain, the best genome assembly of a subgroup or strain was selected as the representative sequence (i.e. genome assemblies with the highest N50 value, highest read coverage, and lowest number of contigs).

RESULTS

Initial phytoplasma screening by PCR of the 16S rRNA gene

No PCR inhibitors were present in the DNA extracts obtained in this study based on the visualization of amplicons of the expected size using the generic bacterial fD1/rP2 16S rRNA primers in the PCR (data not shown). The 16S rRNA genes of all 36 samples were amplified when using the phytoplasma-specific universal primers. Sanger sequencing of the phytoplasma-specific amplicons were successful in all except two cases, and were identified to be phytoplasma-derived based on BLASTn analyses (data not shown). The PCR amplicons for strains BAWM-CHN-EU009550 and BAWM-CHN-EU009551 appeared as a smear on the agarose gel, and Sanger sequencing failed for both. The DNA of these two strains were still used to prepare Illumina sequencing libraries because of their geographic significance since they are the only strains in this study originating from China (Table S1) and because they were analysed previously [64].

Of the 34 strains for which the 16S rRNA gene was successfully Sanger sequenced in this study, BLASTn results showed that 33 strains showed ≥97.5% nucleotide sequence similarity to at least one sequence of the 16SrII phytoplasmas. The 16S rRNA gene of strain BAWM-155, however, showed <97.5% sequence similarity to all known 16SrII phytoplasma strains, and shared 100% nucleotide identity with that of the weeping teatree witches'-broom (WTWB) phytoplasma (16SrXXV-A, GenBank accession no. AF521672). The 16S rRNA gene and genome sequences of strain BAWM-155 were included in further analyses to serve as an outgroup taxon as previous phylogenetic analyses based on the 16S rRNA gene has indicated that the WTWB phytoplasma (16SrXXV-A) is a sister taxon to the 16SrII group [65].

The Sanger-sequenced phytoplasma 16S rRNA genes of the strains received in this study were only used for initial phytoplasma screening and to determine that the strain identified during PCR matched the strain after Illumina sequencing. This is because the R16F2n/m23sr sequenced region represents a shorter segment of the 16S rRNA gene than is proposed in the updated '*Ca.* Phytoplasma' species description guidelines [6].

Strain information, 16Sr classification, and genome characteristics

A total of 36 phytoplasma genomes were sequenced in this study, and six additional genomes were retrieved from the NCBI (www.ncbi.nlm.nih.gov, last accessed July 2022; Table S1). Eleven distinct 16SrII subgroups or strains were identified based

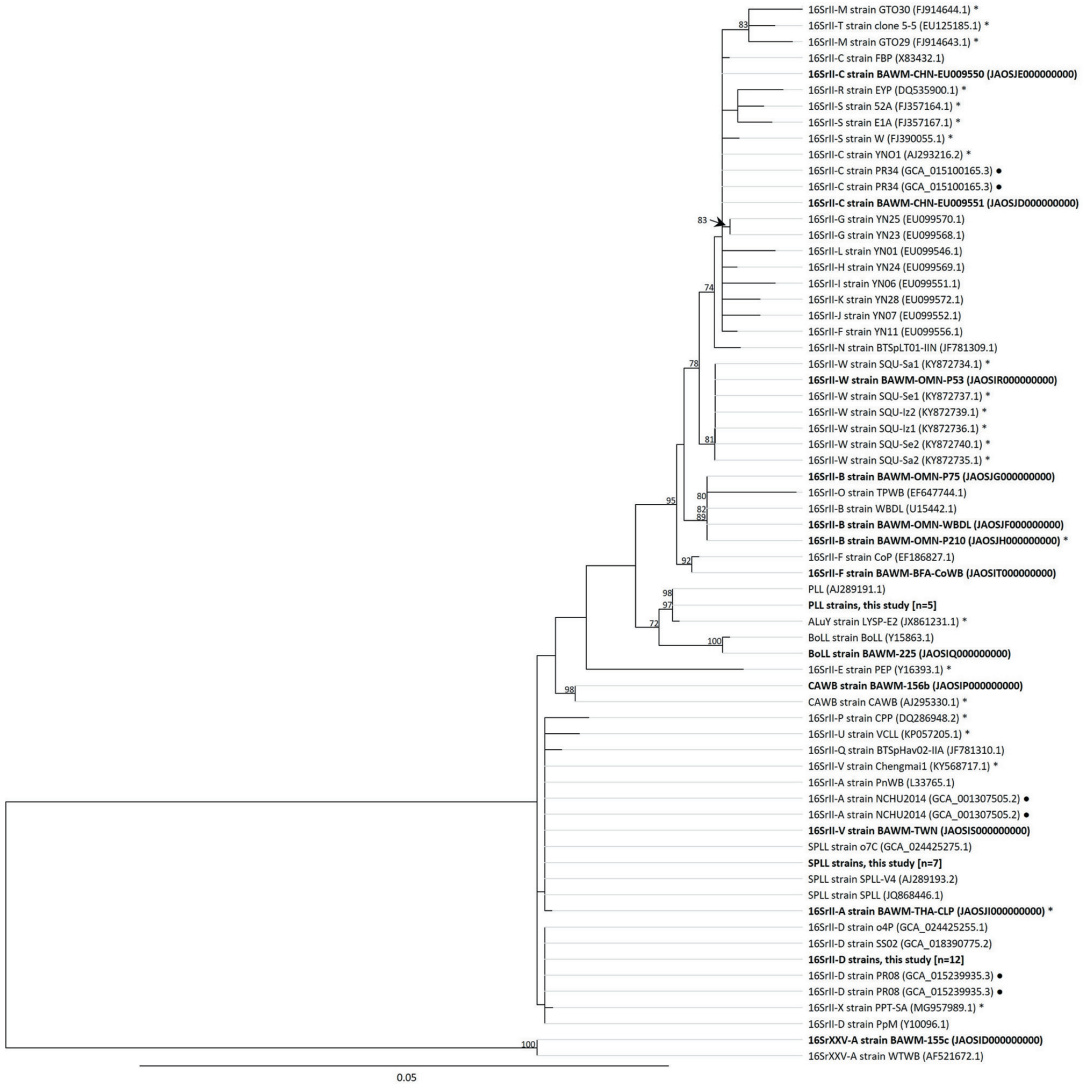

Fig. 1. Maximum-likelihood tree inferred using 16S rRNA nucleotide sequence alignments of representatives of 16SrII phytoplasma subgroups A to X (publicly available) as well as additional representative 16S rRNA sequences extracted from whole genome assemblies analysed in this study. The weeping teatree witches'-broom phytoplasma in the 16SrXXV-A subgroup served as the outgroup taxon. Analyses were done by specifying the GTR +G substitution model in RaxML, implemented in Geneious Prime 2022.2.1. Numbers at branch nodes indicate the percentage of bootstrap support associated with clustered taxa, and only support percentages of 70% and above of the 1000 bootstrap iterations are shown. Branch lengths indicate the number of nucleotide substitutions per site (see bar). Taxa were labelled with subgroup designations according to [40–43]. Accession numbers are shown in brackets. An asterisk (*) indicates 16S rRNA sequences which are too short according to the most recent species description recommendations (less than 95% of the full 16S rRNA length [6],). A solid circle (●) indicates sequences derived from complete genome assemblies. Names in bold indicate samples sequenced in this study. Numbers in square brackets indicate the number of samples obtained in this study with identical 16S rRNA nucleotide sequences.

on alignment analyses of the 16S rRNA gene extracted from these genomes against the reference strains of each subgroup as designated in [40–43] (Tables 1, S1 and S3). These included members within the 16SrII-A (*n*=2), 16SrII-B (*n*=3), 16SrII-C (*n*=3), 16SrII-D (*n*=15), 16SrII-F (*n*=1), 16SrII-V (*n*=1), and 16SrII-W (*n*=1) sub-groups, which originated from six countries, including Australia, Burkina Faso, China, Oman, Taiwan, and Thailand. The phytoplasma strains that lacked formal subgrouping within the 16SrII group, but which met the criteria to be contained therein, included BoLL (*n*=1), CAWB (*n*=1), PLL (*n*=5), and SPLL (*n*=8), and were all sampled in Australia (Table S1). One 16SrXXV-A phytoplasma, strain BAWM-155, was detected from the single *Melaleuca* sp. (*Myrtaceae*) sample collected in Cairns, Queensland, Australia, with symptoms similar to those associated with infection by the WTWB phytoplasma (16SrXXV-A; GenBank accession no. AF521672 [66]; Table S1). The placement of all these strains was confirmed based on the maximum-likelihood analyses of their 16S rRNA gene with other '*Candidatus* Phytoplasma' species (Fig. S1).

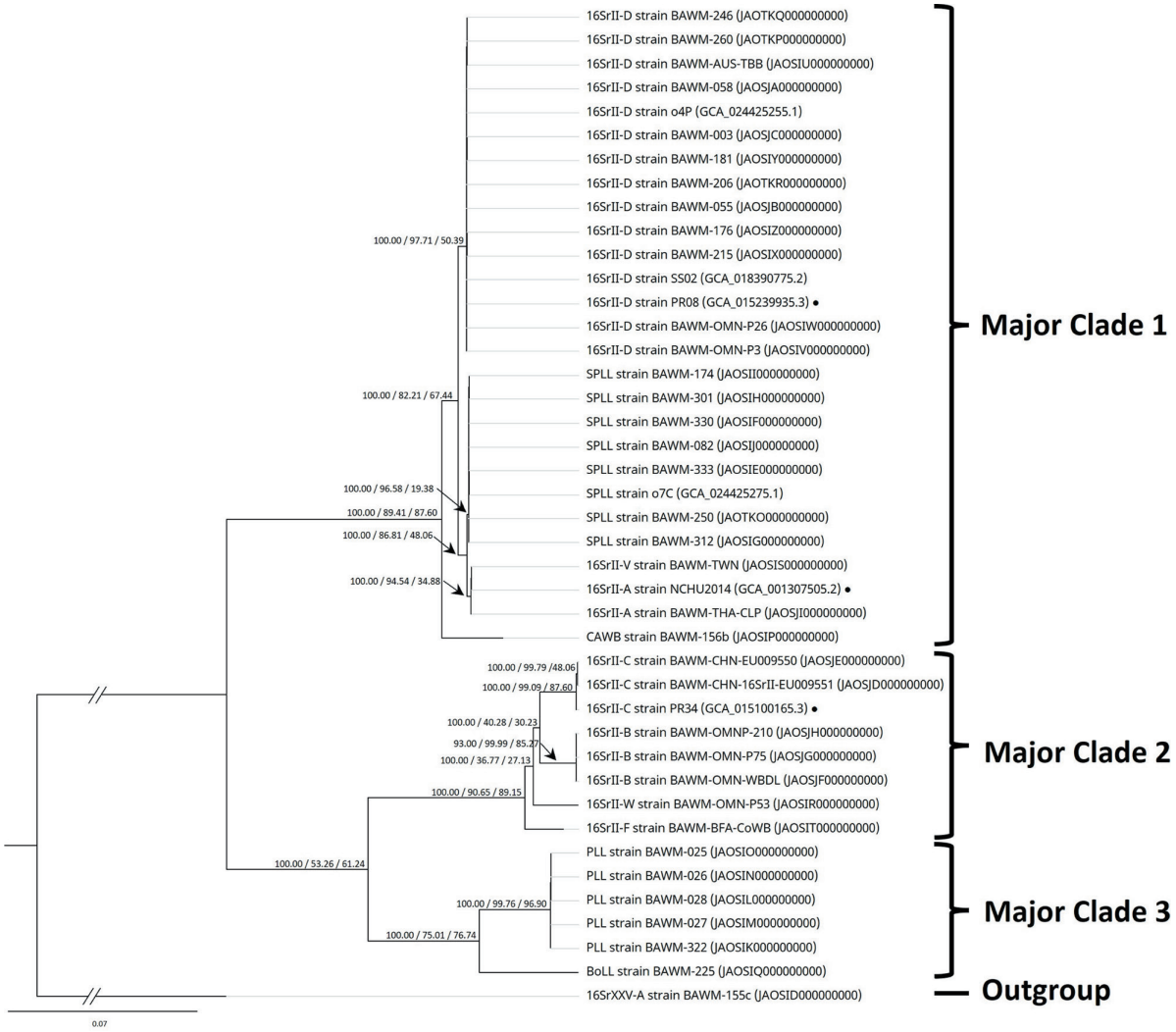

Fig. 2. Maximum-likelihood tree of the concatenated dataset of 215 single-copy ortholog (genes identified to be shared across each of the 42 genomes analysed in this study inferred using IQ-TREE version 2.0.5. Support values include bootstrap percentages after 1000 iterations, followed by the site (sCF) and gene concordance factors (gCF), i.e. (bootstrap/sCF/gCF). Branch lengths indicate the number of amino acid substitutions per site (see bar). Two forward slashes indicate branches that were trimmed, with the original branch length representing 0.47 amino acid substitutions per site. A solid circle (●) indicates sequences derived from complete genome assemblies.

*i*PhyClassifier confirmed the previous assignment of the 16SrII-D (*n*=15), 16SrII-V (*n*=1), and 16SrXXV-A (*n*=1) subgroup members, but classified the 16SrII-B (*n*=3), 16SrII-F (*n*=1), and 16SrII-W (*n*=1) phytoplasmas as 16SrII-C variants with similarity coefficients of 0.99 (Table 1). Additionally, the 16SrII-A subgroup members (*n*=2) were consistently identified as members of the 16SrII-V subgroup (Table 1). *i*PhyClassifier classified PLL and SPLL as variants of the 16SrII-D and 16SrII-V subgroups, respectively (Table 1). The BoLL and CAWB phytoplasmas were considered members of new 16SrII subgroups, with the highest RFLP similarity coefficients of 0.93 and 0.94, respectively, with the 16SrII-D subgroup (Table 1).

The phytoplasma genomes analysed in this study were of variable levels of completeness. Only the genomes of strains NCHU2014 (16SrII-A), PR34 (16SrII-C), and PR08 (16SrII-D) were considered complete genome assemblies and were 635584, 614574, and 588746 bp in size, respectively (Table S1). The 16SrII genomes investigated in this study ranged in size from 495918 bp (16SrII-C phytoplasma strain BAWM-CHN-EU009550) to 876297 bp (CAWB phytoplasma strain BAWM-156) (Table S1).

The complete phytoplasma genomes encoded 24, 27, and 28 tRNA genes for the 16SrII-A strain NCHU2014, 16SrII-C strain PR34, and the 16SrII-D strain PR08, respectively (Table S1). The fewest number of tRNA genes for the draft genome sequences was 17 per genome and was annotated from the publicly available genome of strain SS02 (16SrII-D), followed by 24 tRNA genes per genome, annotated from the genomes of strains BAWM-OMN-WBDL (16SrII-B) and BAWM-CHN-16SrII-EU009551

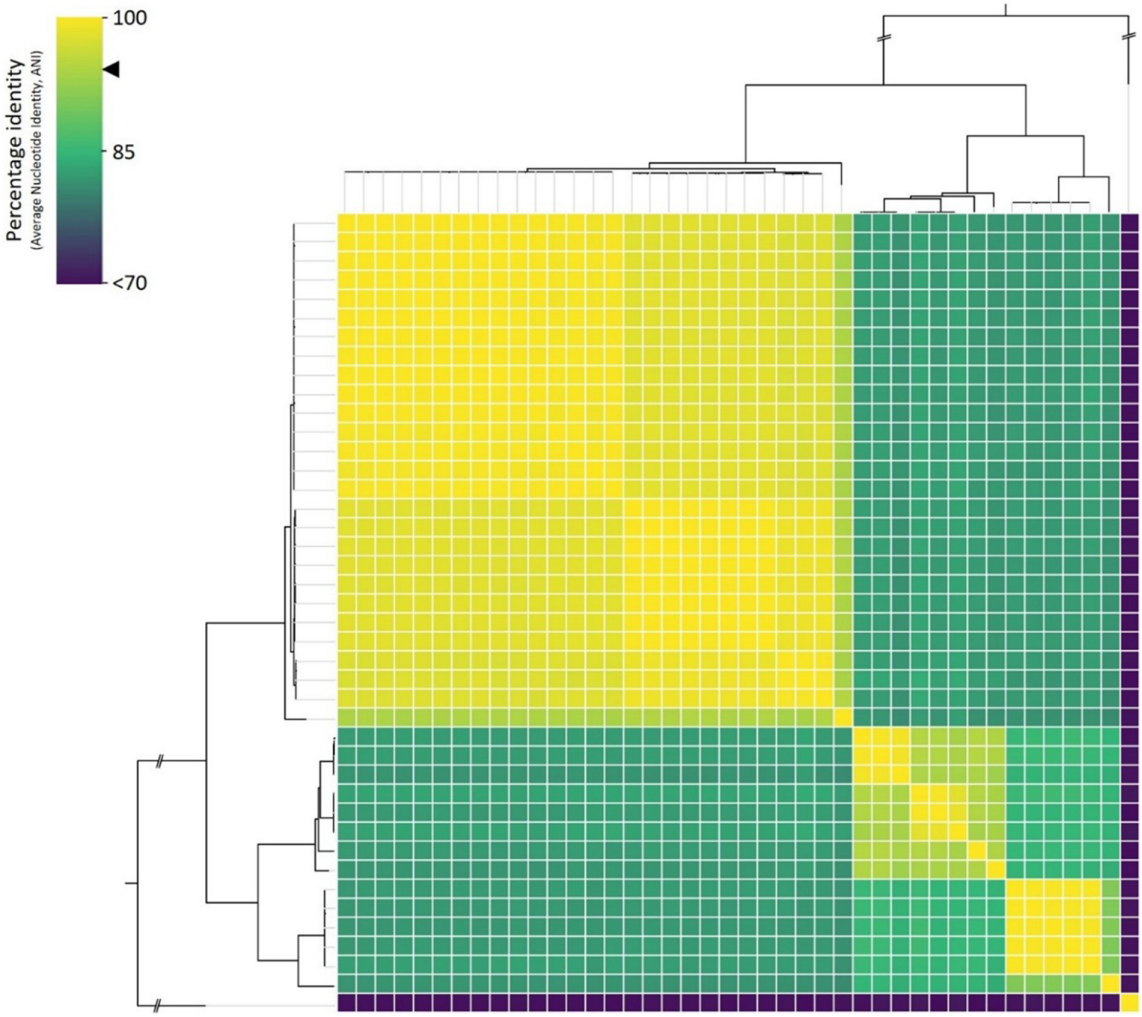

Fig. 3. Reciprocal comparisons of the average nucleotide identity (ANI) values plotted against the single copy ortholog phylogenetic tree from Fig. 2. The ANI values are depicted by a heat map (see key at top left). The black arrowhead at the gradient key indicates the 96% ANI within-species threshold.

(16SrII-C) (Table S1). Strain BAWM-026 (PLL) had a total of 28 tRNA genes, which was the most tRNA genes annotated from the draft genomes in this study (Table S1).

The complete 16SrII phytoplasma genome assemblies encoded 538, 501, and 488 protein-coding genes for 16SrII-A strain NCHU2014, 16SrII-C strain PR34, and 16SrII-D strain PR08, respectively (Table S1). A similar number of protein-coding genes were annotated from the 16SrII phytoplasma draft genomes, with the fewest annotated from strain BAWM-OMN-WBDL (16SrII-B, 404 genes) and the most from strain BAWM-BFA-CoWB (16SrII-F, 572 genes) (Table S1).

The WTWB phytoplasma strain BAWM-155 (16SrXXV-A) that was used as the outgroup taxon had a draft genome assembly size of 504639 bp (Table S1). Gene annotations of this genome recovered 27 tRNA genes, one 16S rRNA gene, one 23S rRNA, and 408 protein-coding genes (Table S1).

Classification of taxa within the 16SrII phytoplasma

Nucleotide sequence analyses of the 16S rRNA gene

A total of 87 16S rRNA sequences were available for the 16SrII and 16SrXXV-A phytoplasmas when considering the genomes analysed in this study and sequences of representative strains (Tables S1 and S2, Fig. 1). Only 60 of these 87 16S rRNA gene sequences were >95% complete and could be used for pairwise nucleotide sequence similarity analysis according to the most recent phytoplasma description guidelines [6] (Tables S1 and S2). Pairwise nucleotide sequence comparisons indicated that all the phytoplasma strains classified within the 16SrII group showed at least 97.59% nucleotide sequence similarity to another 16SrII sequence (Table S2). The near-complete 16S rRNA genes of strains BAWM-155 and WTWB in the 16SrXXV-A phytoplasma

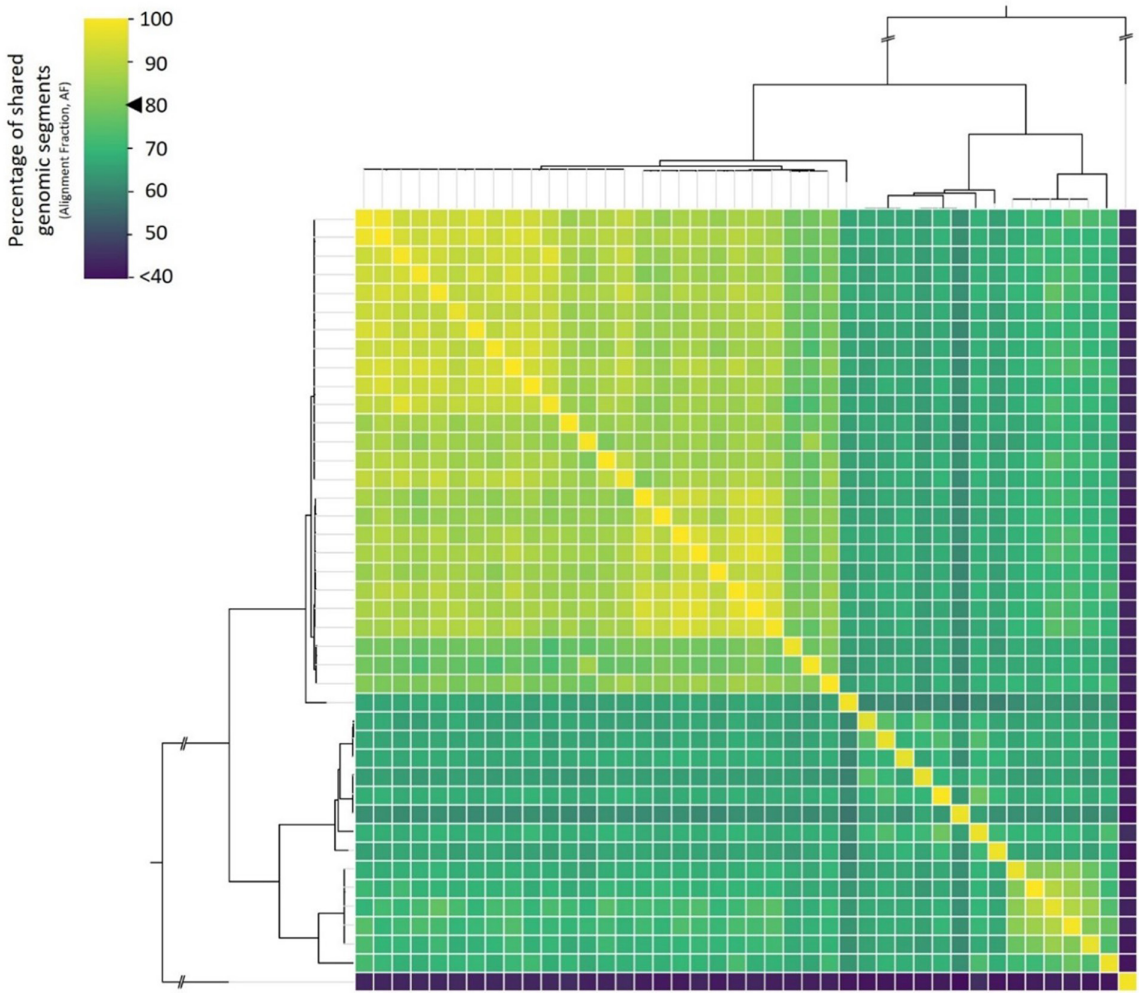

Fig. 4. Reciprocal comparisons of the mean percentage of genomic segments (alignment fraction, AF) shared between the 16SrII and 16SrXXV-A genomes plotted against the single-copy ortholog phylogenetic tree from Fig. 2. The percentages values are depicted by a heat map (see key, top left). The black arrowhead at the gradient key indicates 80% AF.

group showed <91.00% nucleotide sequence similarity to any of the 16SrII phytoplasma strains (Table S2). Of the 87 16S rRNA sequences obtained in this study, 44 were identical in sequence (Table S2).

Nine distinct clades within the 16SrII phytoplasma group were well supported (bootstrap >70%) in the maximum-likelihood tree reconstructed using nucleotide sequence alignments of the 16S rRNA gene obtained from the strains sequenced in this study, publicly available 16SrII genome sequences, and from subgroup representatives that were publicly available (Fig. 1). These clades included those comprising the 16SrII-M and 16SrII-T strains (83% bootstrap), the 16SrII-G strains (83% bootstrap), the clade splitting the 16SrII-N strain from 21 other 16SrII strains (74% bootstrap), the 16SrII-W strains (81% bootstrap), the clade comprising the 16SrII-B and 16SrII-O strains (89% bootstrap), the 16SrII-F clade with phytoplasma strains CoP and BAWM-BFA-CoWB (92% bootstrap), the PLL and ALuY phytoplasma strains (97% bootstrap), the BoLL phytoplasma strains (100% bootstrap), and the CAWB phytoplasma strains (98% bootstrap). The position of phytoplasma strains in subgroups 16SrII-A, 16SrII-C, 16SrII-D, 16SrII-E, 16SrII-P, 16SrII-Q, 16SrII-U, 16SrII-V, and 16SrII-X in the phylogenetic tree was poorly supported (bootstrap support <70%, Fig. 1). Lastly, the clade comprising the two 16SrXXV-A strains was well supported (bootstrap 100%).

Species delimitation using phylogenomics and whole genome comparisons

A total of 215 SCO genes were identified to be shared between the 16SrII and 16SrXXV-A phytoplasma genomes analysed in this study (Table S3). The maximum-likelihood tree reconstructed with the concatenated alignment of these 215 SCO gene amino acid sequences distinguished three major clades with high support values (bootstrap above 70%, sCF and gCF values above 50%; Fig. 2). The first clade consisted of strains classified as members within the 16SrII-A, 16SrII-D, and 16SrII-V subgroups, as well as SPLL strains, and the single CAWB phytoplasma strain (Fig. 2). The second clade comprised members within the 16SrII-B,

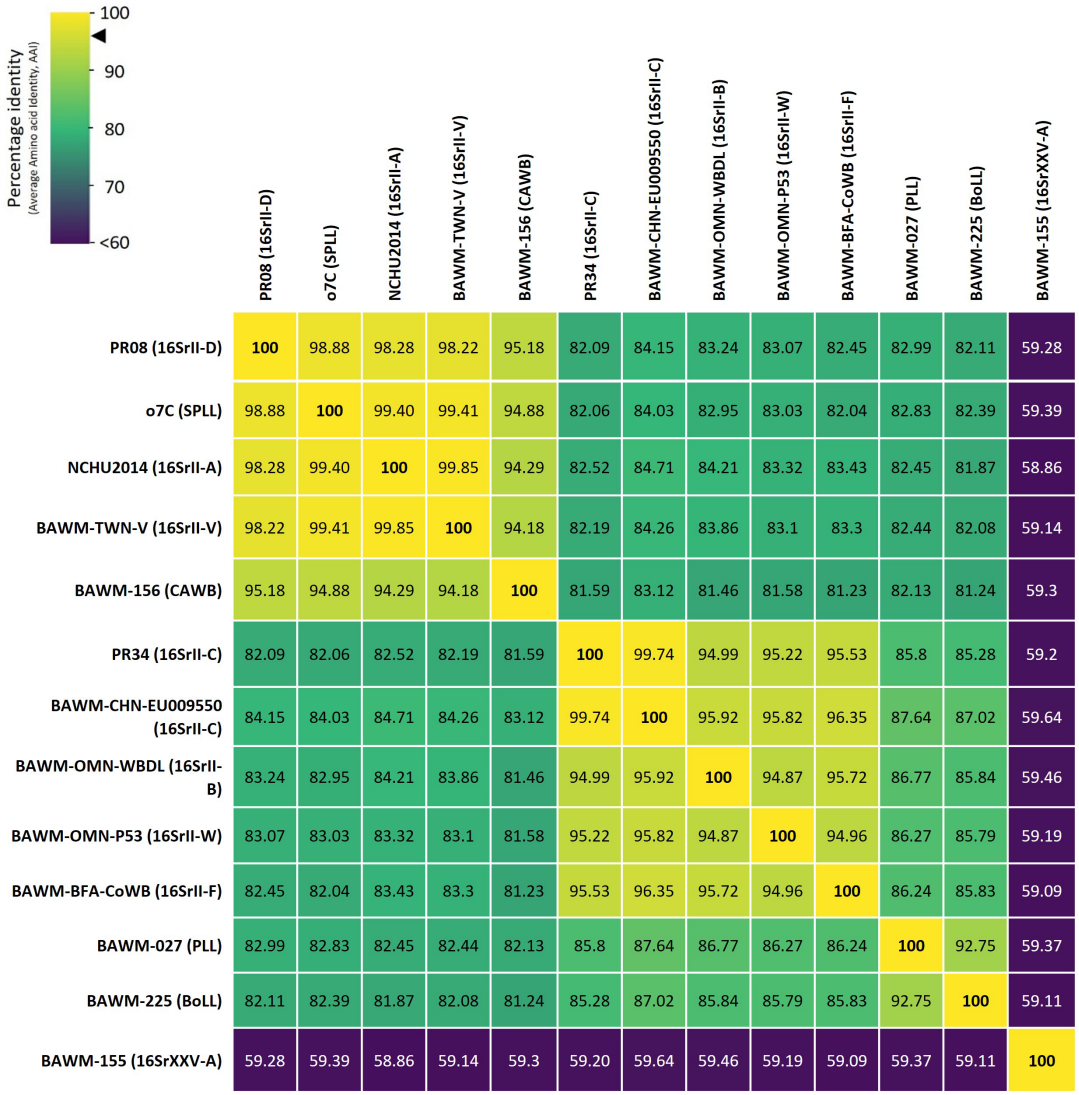

Fig. 5. Average amino acid identity (AAI) values estimated using the protein-coding genes annotated from genome sequences of the representative strain for the 16SrII and 16SrXXV-A phytoplasma strains analysed in this study. The AAI values are depicted by a heat map according to the key. The 16Sr group and subgroup or strain classifications for each representative are indicated in brackets. The black arrowhead at the gradient key indicates the 96% AAI within-species threshold.

16SrII-C, 16SrII-F, and 16SrII-W subgroups (Fig. 2). The last major clade consisted of the PLL and BoLL phytoplasma strains (Fig. 2). All members assigned within a specific 16SrII subgroup or subgroup variant based on the 16S rRNA gene clustered with each other in this phylogenomic analysis, and showed little sequence divergence, as seen by the short or flat branch lengths at these tips (Fig. 2). Together, these results suggest that each of the 16SrII subgroups or subgroup variants represent distinct taxonomic units (species or subspecies).

Six branches in Major Clade 1 and Major Clade 2 showed low gCF and/or sCF values, indicating that the relationships could not be accurately resolved (Fig. 2). These included the branches of the 16SrII-A, 16SrII-V, and SPLL part of Major Clade 1 (gCF values all below 50%, Fig. 2), as well as the clade consisting of phytoplasma strains in subgroups 16SrII-B, 16SrII-C, 16SrII-F, and 16SrII-W in Major Clade 2 (gCF and sometimes sCF values below 50%, Fig. 2). Nevertheless, the 16SrII-B and 16SrII-C are distinct taxa, demonstrated by their high bootstrap, gCF, and sCF values, as well as sequence divergence between these subgroups being evident (Fig. 2). The 16SrII-A, 16SrII-D, 16SrII-V, and SPLL phytoplasma strains in Major Clade 1 produced shorter branch lengths than the taxa in Major Clade 2.

The ANI results supported the close genetic relationship between the 16SrII subgroups and strains within each of the three major clades observed in (Fig. 2). At an ANI threshold of >97%, eight distinct 16SrII phytoplasma species were delimited (yellow clusters in Fig. 3). The strains classified as members within the 16SrII-A, 16SrII-D, and 16SrII-V subgroups, as well as the SPLL phytoplasma strains fulfilled the criteria to be classified as members belonging to a single species. The single CAWB strain BAWM-156 was identified as the second distinct species. The three strains classified as 16SrII-C members were identified as the third distinct species, followed by the 16SrII-B members, the single 16SrII-W phytoplasma strain BAWM-OMN-P53, and single 16SrII-F phytoplasma strain BAWM-BFA-CoWB, the PLL phytoplasma strains, and the single BoLL phytoplasma strain BAWM-225 as the fourth to eighth species within the 16SrII group. These ANI results were further supported as >80% of their genomic segments aligned (AF) within each species cluster delimited by the ANI threshold (Fig. 3). Within each of the three major clades, some phytoplasma species shared borderline ANI values (approximately 96% ANI, indicated in green in Fig. 3), but with an AF of <80% (shown in blue in Fig. 4), indicating high genomic divergence in these comparisons. An example of this can be seen within Major Clade 1 from Fig. 1 where the CAWB phytoplasma strain BAWM-156 shared ~96% ANI with the 16SrII-D members (Fig. 3) but with <80% genomic segments aligned (Fig. 4). Also within Major Clade 1, all the 16SrII-D strains shared 100% ANI with each other, and the same was true for all the SPLL strains (Fig. 3). Both the 16SrII-A and 16SrII-V members also shared 100% ANI with each other. The ANI in comparisons between strains classified within the subgroups 16SrII-D, SPLL, 16SrII-A, and 16SrII-V, however, was less than 100%, at approximately 98–99% ANI, suggesting that each represent distinct taxonomic units within this species designation (Fig. 3). The 16SrXXV-A phytoplasma strain shared mean ANI values and percent of shared genomic segments of between 77.08–77.85% and 26.84–37.94%, respectively, with all of the genome sequences of members assigned to the 16SrII group (Figs 3 and 4).

The clustering pattern observed for all the 16SrII or 16SrXXV-A phytoplasma strains analysed in this study using the >97% ANI threshold was mirrored at an AAI threshold of >97% (Fig. 5). Additionally, all the 16SrII phytoplasma genomes analysed here shared >80% AAI with each other, but they all shared <60% AAI with the 16SrXXV-A phytoplasma strain BAWM-155c (Fig. 5). These reciprocal AAI comparisons involved 295 to 465 shared protein-coding genes between a pair of phytoplasma genomes (Fig. S2, above the diagonal). This represented a minimum of 65% of the total protein-coding genes encoded by the phytoplasma genomes being compared in the reciprocal AAI analyses (Fig. S2, below the diagonal). In some cases, over 90% of the total protein-coding genes between a pair of phytoplasma genomes could be compared (Fig. S2, below the diagonal).

DISCUSSION

Contribution towards expanding sequence data for the phytoplasmas

The 36 genomes sequenced in this study contribute significantly towards expanding the genomic sequence data available for the phytoplasmas, with draft genome sequences being provided for the first time for six taxa classified within the 16SrII group, namely 16SrII-F, 16SrII-V, 16SrII-W, BoLL, CAWB, and PLL phytoplasma strains, and for one member within the 16SrXXV-A group (WTWB). Twelve additional 16SrII-D genomes and seven additional SPLL genomes were uploaded to the NCBI, representing the largest amount of intraspecies sequence data for a phytoplasma taxon to date.

The draft phytoplasma genomes obtained in this study represented assemblies of variable completeness (Table S1). Phytoplasma genomes exhibit high plasticity which contributes to natural variations in genome sizes and their respective gene compositions, even between closely related taxa [67, 68]. While we expect natural variations in phytoplasma genome sizes due to this high genomic plasticity, we do not expect the draft genome size estimates in Table S1 to be accurate. This is due to phytoplasma genomes being rich in homopolymer regions (AT-rich) [68, 69] which impact short read sequencing approaches [70]. Additionally, gene duplications are common in phytoplasma genomes [68, 69] which affects the performance of genome assembly programs using short reads due to duplicated sequences being collapsed into a single contig sequence [71]. Taken together, these factors warrant investigation into the genome size and gene compositions among the 16SrII strains analysed in this study should their draft genome assemblies be completed in the future. These genomes could be completed using hybrid genome assembly methods which has been effective at obtaining complete phytoplasma genome assemblies [72, 73]. This would entail using the Illumina sequence data generated in this study together with sequence data obtained by long read sequencing technologies, such as Oxford Nanopore Technology. The same reference strains could be re-sequenced with the long read technologies as total nucleic acid and/or preserved plant material is available for the strains investigated in this study. The draft genomes assembled in this study, however, were suitable for whole genome comparative assessments based on the similar number of tRNA genes, recovery of rRNA gene sequences, and the adequate quality metrics for most of the genomes (Table S1, [74]).

Taxonomic delimitation within the 16SrII phytoplasma group and with the 16SrXXV-A outgroup taxon

This study observed species and subspecies taxonomic boundaries within the 16SrII phytoplasma group, with eight species and three subspecies described. A genus boundary between the 16SrII and 16SrXXV phytoplasma groups was also observed in this study. These boundaries were consistently observed using three separate whole genome comparative approaches, namely phylogenomics, ANI with AF, and AAI, and by applying criteria and thresholds specified for culturable bacteria where applicable

[7, 8, 30]. These findings were possible due to the inclusion of multiple individuals assigned to a subgroup or subgroup variant from diverse geographic and host ranges, as well as the inclusion of a suitable outgroup taxon to the 16SrII phytoplasma group (Table S1). The 16SrXXV-A phytoplasma was an ideal outgroup taxon as it fulfilled two criteria: (i) the strain belonged to a separate taxonomic designation than all the ingroup taxa under study, and (ii) the outgroup taxon was a close relative to the ingroup taxa. This was based on previous nucleotide sequence analysis of the 16S rRNA gene from a 16SrXXV-A phytoplasma strain which was identified to be one of the closest known relatives to all strains classified within the 16SrII group [65]. With 16S rRNA nucleotide- and whole genome-based sequence analyses, the current study confirmed that the 16SrXXV-A phytoplasma strains can be considered as close relatives of the 16SrII phytoplasma strains but are not contained within this group (Figs 1-5).

The first whole genome sequence approach used in this study to analyse the strains classified within the 16SrII and 16SrXXV phytoplasma groups was a phylogenomic assessment. The phylogenomic assessment was performed using the SCO genes identified to be common to all strains analysed in this study. This analysis was done to reconstruct and interpret the evolutionary relationships between taxa, and to estimate their molecular divergence. In the phylogenomic analysis, all eight 16SrII taxa that were described as novel species had high sCF and gCF support values (above 50%) indicating high confidence in the delimitation of these taxa as distinct species (Fig. 2). However, some nodes were not confidently resolved as they had low gCF and/or sCF values (below 50%). For example, the terminal node comprising all SPLL strains had a gCF value of 19.38%, while the node that comprised the 16SrII-A and 16SrII-V subgroups had a gCF value of 34.88% (Fig. 2). Poor node support within clades may be attributed to limited sequence divergence between the taxa, indicating that the representative sequences are from strains belonging to the same species. For example, there was enough evidence to suggest that the clade consisting of members within the 16SrII-A, 16SrII-D, and 16SrII-V subgroups and the SPLL subgroup variants belonged to the same species and was made up of three separate subspecies (Figs 2 and 3). Indeed, we observed that the 16SrII-V strain and 16SrII-A strains phytoplasmas lacked any major sequence divergence at all and are likely strains within the same taxon (Fig. 2). Additionally, limited taxon sampling occurred in the Major Clade 2 (Fig. 2) as whole genome sequence data of representative strains has not been made publicly available and was not obtained in this study for 15 known 16SrII phytoplasma subgroups (16SrII-E, 16SrII-G–T). Future work should focus on generating publicly available genome sequences for these missing 16SrII subgroup members to determine their taxonomic rank and evolutionary relationships with other phytoplasma strains within the putative genus represented by the 16SrII group.

The second and third whole genome species delimitation approaches involved pairwise comparisons of whole genome ANI and AF, as well as pairwise AAI estimates. The currently specified within-species thresholds for bacteria include an ANI of >95 or >96% [8] with ≥80% AF [7], and an AAI of >95% [75]. The ANI values between distinct bacterial genera are specified to be below 75%, with less than 40% AF [29], and <60% pairwise AAI [30]. We investigated both indices of ANI and the AF during the process of species delimitation because when the AF percentage is low, spurious ANI values are often obtained [30, 76]. Additionally, both ANI and AAI approaches were chosen as ANI values are only considered to be robust in the 80–100% range, while AAI values offer more resolution and are more suitable for comparisons between highly divergent taxa where ANI values are below 80% [28]. The robustness of the AAI approach is demonstrated in Fig. S2, where generally over 75% of the protein-coding genes could be used in the pairwise AAI comparisons, even when the 16SrXXV-A phytoplasma strain was compared to the 16SrII phytoplasma strains. When using ANI and AAI within-species thresholds of >97%, with >80% AF in the ANI analyses, to define members belonging to the same species, further support to the phylogenomic assessments was given to the delineate eight distinct 16SrII phytoplasma species (Figs 3 -5). It also added additional support to the phylogenomic results that members of the 16SrII-V subgroup are unlikely to be a distinct taxon from members in the 16SrII-A subgroup. All 16SrII phytoplasma strains met the within-genus threshold values for ANI and AF [30].

The ANI with AF, and AAI between the 16SrXXV-A phytoplasma strain and all of the 16SrII phytoplasma strains were below 40, 70, and 60%, respectively (Fig. 5), supporting a putative genus boundary between these two phytoplasma groups [30]. A similar ANI analysis was carried out by Cho *et al.* [63] for 11 strains within the 16SrI phytoplasma group, all of which shared ANI and AAI values above 92%, with >60% AF [63]. These results by Cho *et al.* [63], therefore, classifies these 16SrI strains to be members within the same genus. However, no taxon outside of the group 16SrI phytoplasma group was compared in their analyses of the 16SrI phytoplasma strains [63]. Together, the results in our study along with those by Cho *et al.* [63] suggest that phytoplasma 16Sr groups, delimited by the 97.5% 16S rRNA nucleotide sequence similarity threshold, might represent genus-level designations.

The clades we observed in our phylogenomic and whole genome comparative analyses were consistent with previous analyses of 16SrII phytoplasma strains using MLSA-based approaches [33, 77, 78], suggesting that the genes used in these MLSA-based studies are sufficient to identify distinct species within the 16SrII phytoplasma group. Interestingly, our results are also consistent with those from Southern blot analyses conducted in the late 1990s [79]. These Southern blot analyses identified that most of the genomic regions of the SPLL and tomato big bud (TBB; 16SrII-D) phytoplasma strains hybridized with each other, but also that some regions present in the TBB strains did not hybridize with SPLL strains, showing evidence of sequence divergence between the two phytoplasma genomes. The molecular sequence-based species delimitation criteria are also supported by the distinct biological and ecological features of each of these phytoplasma taxa (see "*Candidatus* Phytoplasma' species descriptions' section). For example, the 16SrII-B phytoplasma strains are distinct from the 16SrII-D phytoplasma strains based on sequence data, but they also have distinct natural host ranges and are naturally vectored,

confirmed or putatively, by insect species within different genera [20, 24, 80, 81]. In a separate case, the phytoplasma strains in subgroups 16SrII-B (detected in Oman, Iran, the UAE), 16SrII-C (detected in China, India), 16SrII-F (detected in Burkina Faso), and 16SrII-W (detected in Oman) consistently showed a close phylogenomic relationship (Figs 2–5); however, three of these four taxa have narrow host and/or geographic ranges.

Members assigned to a specific 16SrII subgroup or subgroup variant based on nucleotide sequence analysis of the 16S rRNA gene clustered with each other in our genome-based analyses (Figs 2–5). However, data analyses of the 16S rRNA gene (i.e., nucleotide sequence similarity, maximum-likelihood tree, and *i*PhyClassifier) were unable to accurately resolve the taxonomic rank of each of the subgroups or subgroup variants. For example, the PLL phytoplasma strains were identified as 16SrII-D variants by *i*PhyClassifier (Table 1), but were described as a distinct species from the 16SrII-D phytoplasma strains based on the whole genome sequence analyses reported here, and were not closely related to this subgroup based on the phylogenomic maximum-likelihood tree, ANI with AF, and AAI results (Figs 2–5). Several of the distinct species identified in this study shared >98.65% nucleotide sequence similarity at their 16S rRNA gene, which is higher than the most recent delimitation criteria suggested for phytoplasma species delimitation [6]. For example, the near full length 16S rRNA gene of members within the 16SrII-B, 16SrII-C, and 16SrII-W subgroups share 98.87–99.60% nucleotide sequence similarity, Table S2, but are distinct species based on the whole genome sequence analyses (Figs 3–5). Additionally, many of the nodes in the 16S rRNA maximum-likelihood tree had poor bootstrap support, resulting in the inability to resolve many of the evolutionary relationships between strains classified within the 16SrII group (Fig. 1). These results indicate that, similar to other bacteria [75, 82], 16S rRNA sequence similarity assessments are sufficient to assign a taxon to a genus. However, for accurate taxon delimitation below the genus level, i.e., species and subspecies, more in-depth assessments of the molecular divergence between phytoplasma strains and species are required than those based on the 16S rRNA gene alone. Genome-based methods for species and subspecies delimitation are more robust and provide more phylogenetic resolution than the 16S rRNA gene alone or MLSA approaches based on analyses done for other bacteria genera, including those that are phytopathogenic [32, 83, 84].

The subgroup classification of the 16SrII-A, 16SrII-V, and 16SrII-F phytoplasma strains in this study were difficult to resolve when considering published literature and *i*PhyClassifier results. The most noticeable case was the 16SrII-A and 16SrII-V phytoplasma strains, which were determined to be the same subspecies in our analyses (Figs 2–5). This error could stem from the *i*PhyClassifier database lacking any sequence data of the 16SrII-A phytoplasma strains prior to the description of the 16SrII-V phytoplasma subgroup [41], leading the authors to believe they had identified a novel subgroup. Alternatively, the 16SrII-V description could have overwritten the 16SrII-A description due to their high sequence similarity. This is plausible as no sequence analysed in this study was assigned as a member or variant of the 16SrII-A subgroup by *i*PhyClassifier (Table 1), despite being identified as such previously in the literature [40, 85, 86] and in the *i*PhyClassifier database available for users (accessible at: https://plantpathology.ba.ars.usda.gov/cgi-bin/resource/phytoclass.cgi). It is difficult, however, to confirm these theories as the classification database available for users of *i*PhyClassifier was last updated in 2015, prior to the description of the 16SrII-V phytoplasma [41]. Nevertheless, the phytoplasma reference sequences of the members belonging to the subgroups 16SrII-A (GenBank accession no.: L33765) and 16SrII-V (GenBank accession no.: KY568717) on the accessible database and/or in the literature have identical 16S rRNA nucleotide sequence similarities (apart from a few bases at the beginning of the gene) with indistinguishable RFLPs with all enzymes on *i*PhyClassifier. These two phytoplasma subgroups should, therefore, be considered as members of the same 16Sr subgroup. To clarify this, we recommend the abolishment of the 16SrII-V subgroup.

The current phytoplasma guidelines do not recommend subspecies and genus designations between phytoplasma strains [4, 5, 87]. Instead, a single provisional genus exists for the phytoplasmas, namely the '*Candidatus* Phytoplasma' genus, and Rule d in the first '*Ca.* Phytoplasma' species description guidelines recommends against the use of the subspecies rank [5]. These decisions were made due to the lack of suitable criteria and sufficient sequence data to define phytoplasmas at the subspecies level at the time [4]. For example, RFLP of the 16S rRNA gene was deemed to lack sufficient resolution for subspecies delineation, there was uncertainty around whether the 16S rRNA gene provided sufficient resolution below the species level, and there was a lack of additional gene sequences available for many phytoplasma taxa or genome sequences for any phytoplasma strain [4]. The informal subgroup classification was, however, supported. Nevertheless, it was encouraged that additional conserved genes be identified to improve phytoplasma classification at the different taxonomic levels [4]. The present study analysed phytoplasma genome sequences, which serves as a high-resolution approach. Phylogenomic and comparative genomic approaches, such as those implemented in the present study, has allowed for the identification of genus-, species-, and subspecies-level taxonomic relationships within other bacterial taxa [29, 31]. Therefore, our results represent a more accurate delimitation of these taxonomic ranks between members within the 16SrII and 16SrXXV-A phytoplasma groups in this study.

The taxonomic ranks of genus, species, and subspecies are acceptable for culturable bacteria according to the International Code of Nomenclature of Prokaryotes [88]. These taxonomic ranks have also been described for unculturable bacteria [89]. For example, the rank of subspecies has been described for the plant pathogenic species '*Ca.* Liberibacter africanus' (class Alphaproteobacteria), and for the species '*Ca.* Mycoplasma haemomuris' (class Mollicutes). It would therefore be valuable and appropriate to include these three taxonomic ranks in any updates to the guidelines used to describe '*Ca.* Phytoplasma' taxa. Previous studies have shown that some closely related phytoplasma strains present different host and/or vector ranges and are distinguishable at the molecular

level, with the term 'vectotype' being assigned in at least one case [90]. Assigning relevant nomenclature at all taxonomic levels allows for accurate and clear communication of the taxon or group of taxa. With the assignment of and reference to phytoplasma 16Sr groups (putative genera), species, and subspecies names, more accurate investigations regarding phytoplasma biology, evolution, and taxonomy can begin and can be communicated more clearly. Together, these aid in the development of management practices for these plant pathogenic bacteria.

'*Candidatus* Phytoplasma' species descriptions

Members within or related to '*Candidatus* Phytoplasma aurantifolia' and the adoption of '*Ca.* Phytoplasma citri' as the recommended nomenclature (16SrII-B)

'*Candidatus* Phytoplasma citri' (ci′tri. L. gen. n. citri meaning 'of *Citrus*', referring to the plant genus in which the phytoplasma was discovered). The reference sequence for '*Ca.* Phytoplasma citri' is the draft whole genome sequence of the original description's reference strain by Zreik *et al.* [14] originally identified from *Citrus aurantifolia* (strain WBDL, genome assembly no.: JAOSJF000000000) originating from Oman.

'*Candidatus* Phytoplasma citri' [(Mollicutes) NC; NA; O, wall-less; NAS (genome assembly no.: JAOSJF000000000); G+C content, 23.8 mol%; highest genome ANI values with a representative '*Candidatus* Phytoplasma' species of 95.63% ANI (69.37% AF) with phytoplasma strain BAWM-BFA-CoWB (AAI is 95.72% for this comparison); oligonucleotide sequences of unique regions within the 16S rRNA gene are (positions are in reference to nucleotide sequence alignments with M30790): AGAAATAAGGCA TCTTTTTTC (175–195 bp), AGTGATAGGTATACTTTAGGAGGGGCTTGCGCCATATTAGTTAGTTGGTA (209–261 bp), CAAAACGGTAGCCTAACTCGTTTATCGAGAGGGCGCTGTCTAAGGTAGGGTCG (1431–1484 bp); P (*Citrus aurantifolia*, phloem); M]. The reference sequence of this species represents a draft genome assembly.

'*Candidatus* Phytoplasma citri' is generally associated with witches'-broom disease of *Citrus* species and has been detected in Oman, Iran, India, and the United Arab Emirates [14, 80]. Sequence analysis using *i*PhyClassifier [9] for a number of previous '*Ca.* Phytoplasma aurantifolia', 16SrII-B, or related phytoplasma strains in a few alternative hosts in the Middle East showed that these strains were more closely related to the '*Ca.* Phytoplasma australasiaticum' (synonym '*Ca.* Phytoplasma australasia') clade; for example, in *Solanum tuberosum* L. 'Spunta' (potato), *Aerva javanica* [91], *Manilkara zapota* [92] (data not shown). Putative vectors of '*Ca.* Phytoplasma citri' include *Diaphorina citri* Kuwayama (Hemiptera: Psyllidae) and *Hishimonus phycitis* Distant (Hemiptera: Cicadellidae) [93]. '*Ca.* Phytoplasma citri' fulfils Rule c [5] as molecular divergence from other phytoplasma taxa, including its closest known relative strain BAWM-BFA-CoWB, is observed using the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct geographic distribution, plant host range, and putative vectors of this species.

As previously suggested [89], the original name of '*Ca.* Phytoplasma citri', namely '*Candidatus* Phytoplasma aurantifolia' [14], violates the International Code of Nomenclature of Prokaryotes and its Orthography appendix [89, 94] and should not be used in further literature and communication. The correct usage of '*Ca.* Phytoplasma citri' is to describe members classified within or related to the 16SrII-B subgroup only (based on 16S rRNA sequence similarity and phylogeny, as well as additional gene or genomic sequences). In this study, no sequence data was reported for strains related to '*Ca.* Phytoplasma citri', and their existence remains to be determined. '*Ca.* Phytoplasma citri' should not be used to refer to the 16SrII group as a whole, as was done previously for '*Ca.* Phytoplasma aurantifolia'. Additionally, '*Ca.* Phytoplasma citri' should not be used for taxa which are more closely related to '*Ca.* Phytoplasma australasiaticum' (synonym '*Ca.* Phytoplasma australasia'), including those classified as 16SrII-A, 16SrII-D, and 16SrII-V members or 16SrII-V variants (SPLL), so as to remove the confusion between these two, distinct species.

Members within or related to '*Candidatus* Phytoplasma australasia' and the adoption of '*Ca.* Phytoplasma australasiaticum' as the recommended nomenclature (16SrII-A, 16SrII-D, 16SrII-V, SPLL)

'*Candidatus* Phytoplasma australasiaticum' (aus.tral.a.si.a′ti.cum. N.L. neut. adj. australasiaticum meaning 'from Australasia', referring to the geographic location from where the species was first described). The reference sequence for '*Ca.* Phytoplasma australasiaticum' is the complete whole genome sequence of strain PR08 (GenBank assembly accession no.: GCA_015239935.3) originating from India.

'*Candidatus* Phytoplasma australasiaticum' [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: GCA_015239935.3); G+C content, 24.8 mol%; highest genome ANI values with another representative '*Candidatus* Phytoplasma' species of 95.88% ANI (73.53% AF) with strain BAWM-156b (CAWB phytoplasma), with an AAI of 95.18% for this comparison; oligonucleotide sequences of unique regions in the 16S rRNA gene are shown for '*Candidatus* Phytoplasma australasiaticum subsp. australasiaticum' below; P (*Parthenium* sp., phloem); M]. The reference sequence of this species represents a complete genome assembly.

'*Candidatus* Phytoplasma australasiaticum' is associated with many diseases of a broad range of hosts and over a broad geographic range, including Australia, India, Iran, Oman, and Pakistan, and Turkey [21, 22, 25]. Insects in the family Cicadellidae have been identified as putative vectors for 16SrII-D subgroup members, including *Orosius lotophagorum* (Kirkaldy), *O. argentatus* (Lachlan),

O. orientalis (Matsumura), as well as *Austroagallia torrida* (Evans), *Neoaliturus* (*Circulifer*) *tenellus* (Baker), and *Neoaliturus haematoceps* (Mulsant and Rey) [20, 26, 95–99]. '*Ca.* Phytoplasma australasiaticum' is a novel species as it fulfils Rule c [5]. This includes evidence of molecular of '*Ca.* Phytoplasma australasiaticum' compared to its closest known relative, strain BAWM-156b (CAWB phytoplasma), based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as evidence of distinct geographic distributions and plant host ranges between these closely related species.

As suggested previously [89], the name '*Candidatus* Phytoplasma australasia' [15] violates the International Code of Nomenclature of Prokaryotes and its Orthography appendix [89, 94]. We propose that '*Ca.* Phytoplasma australasia' should not be used in further literature and communication and, instead, '*Ca.* Phytoplasma australasiaticum' should replace this name in subsequently published literature. '*Ca.* Phytoplasma australasiaticum' should not be used for taxa which are closer related to '*Ca.* Phytoplasma citri' (16SrII-B) so as to remove the confusion between these two distinct species. The correct usage of '*Ca.* Phytoplasma australasiaticum' is to describe members classified within the 16SrII-D subgroup only (based on 16S rRNA sequence similarity and phylogeny, as well as additional gene or genomic sequences), while '*Ca.* Phytoplasma australasiaticum'-related phytoplasma strains include those identified within the 16SrII-A, 16SrII-V (abolished), and SPLL phytoplasmas. '*Ca.* Phytoplasma australasiaticum'-related phytoplasma strains are distinct taxa in their own right, likely representing taxonomic units below the species level. The term 'subspecies', has not been adopted by the phytoplasma taxonomy group [5, 6, 87]. However, we propose the use of this designation for members within the '*Ca.* Phytoplasma australasiaticum' species, and is likely a relevant taxonomic level to define for other phytoplasma taxa.

Should the subspecies concept become accepted in phytoplasma taxonomy, we propose the following subspecies epithets and reference sequences for each:

'*Candidatus* Phytoplasma australasiaticum subsp. australasiaticum' (subspecies epithet: aus.tral.a.si.a′ti.cum. N.L. neut. adj. australasiaticum from Australasia, referring to the geographic location from where the species was first described). The unique oligonucleotide sequence of the 16S rRNA gene is (positions are in reference to alignments with M30790): AGATAAAGGCA TCTTTTATC (175–195 bp). '*Ca.* Phytoplasma australasiaticum subsp. australasiaticum' will refer to phytoplasma strains identified as 16SrII-D members based on 100% sequence similarity of their 16S rRNA gene or 100% ANI with the reference sequence of strain PR08 (GenBank assembly accession no.: GCA_015239935.3) originating from India. '*Ca.* Phytoplasma australasiaticum subsp. australasiaticum' reference strain PR08 shares the highest whole genome comparison values with phytoplasma strain o7C (98.70% ANI, 91.76% AF, 98.83% AAI).

'*Candidatus* Phytoplasma australasiaticum subsp. taiwanense' (subspecies epithet: tai.wan.en'se. N.L. neut. adj. taiwanense, referring to Taiwan, the geographic location where the reference strain was detected). The unique oligonucleotide sequence of the 16S rRNA gene is (positions are in reference to alignments with M30790): TTAATAGGTATGCTTTAGGAGGGG CTTGCGCCATATTAGTTAGTTGGTA (212–261 bp). '*Ca.* Phytoplasma australasiaticum subsp. taiwanense' will refer to all phytoplasma strains identified as 16SrII-A or 16SrII-V (abolished) members based on 100% sequence similarity of their 16S rRNA gene or 100% ANI with the reference sequence of strain NCHU2014 (GenBank assembly accession no.: GCA_001307505.2). '*Ca.* Phytoplasma australasiaticum subsp. taiwanense' reference strain NCHU2014 shares the highest whole genome comparison values phytoplasma strain o7C (99.28% ANI, 87.69% AF, 99.30% AAI).

'*Candidatus* Phytoplasma australasiaticum subsp. ipomoeae' (subspecies epithet: i.po.moe'ae. N.L. gen. fem. n. ipomoeae referring to *Ipomoea*, the plant genus of the phytoplasma's common name, i.e., SPLL phytoplasma). '*Ca.* Phytoplasma australasiaticum subsp. ipomoea' will refer to all phytoplasma strains identified as members of the SPLL phytoplasma based on 100% sequence similarity of their 16S rRNA gene or 100% ANI with the reference sequence of strain o7C (GenBank assembly accession no.: GCA_024425275), originating from Australia. '*Ca.* Phytoplasma australasiaticum subsp. ipomoeae' reference strain o7C shares the highest whole genome comparison values '*Ca.* Phytoplasma australasiaticum subsp. taiwanense' strain NCHU2014 (99.28% ANI, 87.69% AF, 99.30% AAI).

Description of '*Candidatus* Phytoplasma asiaticum' (16SrII-C)

'*Candidatus* Phytoplasma asiaticum' (a.si.a'ti.cum. N.L. neut. adj. asiaticum meaning 'from Asia', referring to Asia, the geographic location from where the phytoplasma has been discovered). The reference sequence for '*Ca.* Phytoplasma asiaticum' is the complete whole genome sequence of strain PR34 (GenBank assembly accession no.: GCA_015100165.3) originating from India.

'*Candidatus* Phytoplasma asiaticum' [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: GCA_015100165.3); G+C content, 24.6 mol%; highest genome ANI value with a representative '*Candidatus* Phytoplasma' species of 95.86% ANI (81.54% AF) with strain BAWM-OMN-P53 (16SrII-W), with an AAI value of 95.22% for this comparison; the unique oligonucleotide sequence of the 16S rRNA gene is (positions are in reference to nucleotide alignments with M30790): GTAATAGGTATACTTTAGGAGGGGCTTGCGCCAT (211–245 bp); P (*Parthenium hysterophorus*, phloem); M]. The reference sequence of this species represents a complete genome assembly.

'*Ca.* Phytoplasma asiaticum' has been reported from China [64, 100], India [26], and Iran [101], with whole genome sequence data available for strains originating from China (GenBank assembly accession nos: JAOSJE000000000 and JAOSJD000000000) and India (GenBank assembly accession no.: GCA_015100165.3). The phytoplasma has a broad plant host range, being detected in at least three plant families including the Cactaceae (cacti, *Opuntia* sp.), Apiaceae (carrots, *Daucus carota*), and Fabaceae (chickpeas, *Cicer arietinum*, and faba beans, *Vicia faba*). *Orosius albicinctus* (Distant) was identified as a competent vector of this phytoplasma to faba beans (*V. faba*) in two regions in Iran [101], and as a putative vector of these phytoplasma strains among *P. hysterophorus* plants in areas of Maharashtra, India [26]. The vector(s) of this phytoplasma species in *Opuntia* sp. in China remain to be determined. '*Ca.* Phytoplasma asiaticum' is a novel species as it fulfils Rule c [5]. This includes evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct geographic distribution and plant host ranges of '*Ca.* Phytoplasma asiaticum' and its closest known relative strain BAWM-OMN-P53 (16SrII-W).

Description of '*Candidatus* Phytoplasma gossypii' (16SrII-F, CoWB)

'*Candidatus* Phytoplasma gossypii' (gos.sy'pi.i. N.L. gen. neut. n. gossypii referring to *Gossypium*, the plant genus from which the phytoplasma was first discovered). The reference sequence for '*Ca.* Phytoplasma gossypii' is the whole genome sequence of strain BAWM-BFA-CoWB (GenBank assembly accession no.: JAOSIT000000000) originating from Burkina Faso.

'*Candidatus* Phytoplasma gossypii' [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: JAOSIT000000000); G+C content, 24.6mol%; highest whole genome comparison values with a representative '*Candidatus* Phytoplasma' species of 95.98% ANI (82.84% AF) with phytoplasma strain BAWM-OMN-P53 (16SrII-W), and an AAI of 94.96% for this comparison; the unique oligonucleotide sequence of the 16S rRNA gene is (positions are in reference to nucleotide sequence alignments with M30790): CAAAACGGTAGCCTAACTCGTTTATGCGAGAGGGCGCCGTCTAA GGTAGGGTCG (1431–1484 bp); P (*Gossypium* sp., phloem); M]. The reference sequence of this species represents a draft genome assembly.

To date '*Ca.* Phytoplasma gossypii' has only been detected in *Gossypium* and *Sida* species grown in the western regions of Africa. Prior to molecular characterization of the phytoplasma, phytoplasma-associated symptoms of virescence, stunting, as well as leaf yellowing and reddening have been documented in cotton in Burkina Faso, Ivory Coast, and Ghana since the 1960s and 1970s [102]. *Orosius cellulosus* (Lindberg) was identified as the vector of the virescence disease in cotton (*Gossypium* sp., family Malvaceae) and weeds in the genus *Sida* (family Malvaceae) in Burkina Faso [102]. Marzachi *et al.* [102] identified that, based on RFLP of the 16S rRNA gene using the restriction enzymes *Alu*I, *Hha*I, *Hpa*II, *Mse*I, the same phytoplasma was infecting *G. hirsutum* and *S. cordifolia* in Mali, supporting the epidemiological observations made in the 1970s in Burkina Faso. '*Ca.* Phytoplasma gossypii' is a novel species as it fulfils Rule c [5]. This includes evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct geographic distribution and plant host range of '*Ca.* Phytoplasma gossypii' compared to its closest known relative, strain BAWM-OMN-P53 (16SrII-W).

The 16S rRNA gene of the '*Ca.* Phytoplasma gossypii' reference strain, strain BAWM-BFA-CoWB, shows the highest sequence similarity to other phytoplasma strains detected in cotton (*Gossypium* species) from two north west African countries, namely Burkina Faso (100% with GenBank accession no.: JQ868439 of strain CoP) and Mali (99.93% with GenBank accession no.: EF363314 of strain CoP_M). There is some confusion in the literature, however, regarding the 16SrII subgroup designation of this phytoplasma. The cotton phyllody phytoplasma (referred to as strain CoP or CoWB) has been classified as a 16SrII-C member [102, 103] or as a member within the 16SrII-F subgroup [40, 42, 104, 105], while *i*PhyClassifier classified it as a variant of the 16SrII-C subgroup (Table 1). We referred to the strain as a 16SrII-F member in this study for consistency with the most recent updates in the literature. This taxon is, however, distinct from other phytoplasma strains classified as members of the subgroups 16SrII-C (FBP phytoplasma GenBank accession no. X83432 and those specified in [64]) and 16SrII-F phytoplasma strains (as specified in [64]) based on the full length of the 16S rRNA gene (Fig. 1). Therefore, we recommend that this taxon and related taxa should not be referred to using a 16SrII subgroup designation for clarification. Instead, referring to the species name, strain name, and as a member of the 16SrII group would provide more clarity.

Description of '*Candidatus* Phytoplasma crotalariae' (16SrII-W)

'*Candidatus* phytoplasma crotalariae' (cro.ta.la'ri.ae. N.L. gen. fem. n. crotalariae referring to *Crotalaria*, plant genus from which the phytoplasma was first discovered). The reference sequence for '*Ca.* Phytoplasma crotalariae' is the whole genome sequence of strain BAWM-OMN-P53 (GenBank assembly accession no.: JAOSIR000000000) originating from Oman.

'*Candidatus* phytoplasma crotalariae' [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: JAOSIR000000000); G+C content, 24.3mol%; highest genome ANI values with a representative '*Candidatus* Phytoplasma' species of 95.98% ANI (82.84% AF) with phytoplasma strain BAWM-BFA-CoWB (AAI is 95.22% for this comparison); the unique oligonucleotide sequence of the 16S rRNA gene is (positions are in reference to nucleotide sequence alignments with

M30790): GAAATAGGTATACTTTAGGAGGGGCTTGCGCCAT (211–245 bp); P (*Crotalaria aegyptiaca*, phloem); M]. The reference sequence of this species represents a draft genome assembly.

'*Ca.* Phytoplasma crotalariae' has only been detected in *Crotalaria aegyptiaca* (Benth.) showing witches'-broom symptoms and from unidentified *Orosius* species in the northern regions of Oman [42]. The plant host *C. aegyptiaca* grows primarily in desert or dry shrubland biomes, and is native to Oman [106], among other countries. '*Ca.* Phytoplasma crotalariae' is a novel species as it fulfils Rule c [5]. This is supported by evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct geographic distribution and plant host range of '*Ca.* Phytoplasma crotalariae' compared to its closest known relative, '*Ca.* Phytoplasma gossypii' (strain BAWM-BFA-CoWB).

Description of '*Candidatus* Phytoplasma bonamiae' (BoLL phytoplasma, no previous subgroup assignment)

'*Candidatus* phytoplasma bonamiae' (bon.ami'ae, N.L. gen. fem. n. bonamiae referring to *Bonamia*, the plant genus from which this phytoplasma has been detected). The reference sequence for '*Ca.* Phytoplasma bonamiae' is the whole genome sequence of strain BAWM-225 (GenBank assembly accession no.: JAOSIQ000000000) originating from Australia.

'*Candidatus* phytoplasma bonamiae' [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: JAOSIQ000000000); G+C content, 25.2 mol%; highest genome ANI values with a representative '*Candidatus* Phytoplasma' species of 93.45% ANI (86.85% AF) with phytoplasma strain BAWM-027 (AAI is 92.75% for this comparison); unique oligonucleotide sequences of the 16S rRNA gene are (positions are in reference to nucleotide sequence alignments with M30790): AAAACCTTCGGG TTTTAGTGGCGAACGGGTGAGTAACACGTAAGCAACCTACCCTA (68–129 bp), AGATATGAGGCATCTTGTACC (175–195 bp), CAAGGAAGAAAGGCAAGTGGTGAACCATTTGTTTGCCGGTACTTG (449–493 bp), CTAGTAAGTCAG TGGTGTAATGGCAATGCTTAACA (594–628 bp), CAAAACGGTAGCTTAACTCGTTTATTCGAGAGGGCGCCGTCTAA GGTAGGGTCG (1431–1484 bp); P (*Bonamia pannosa*, phloem); M]. The reference sequence of this species represents a draft genome assembly.

'*Ca.* Phytoplasma bonamiae' is an infrequently detected phytoplasma based on surveys conducted in Australia (referred to as *Bonamia* little leaf or BoLL phytoplasma [20, 107]), with 14 records in the Phytoplasma Collection held at the Northern Territory Department of Industry, Tourism and Trade, Darwin, Australia (previously at the Charles Darwin University, Darwin, Australia). All records are from regions in Australia located north of the Tropic of Capricorn, including North Kimberley, Western Australia [107] and near Katherine, Northern Territory [20]. This phytoplasma species has only been associated with the plant host *Bonamia pannosa* [20, 107], which is native to Australia and endemic to the north west and north central regions of Australia in particular [108]. In one known occasion, '*Ca.* Phytoplasma bonamiae' and '*Ca.* Phytoplasma stylosanthis' phytoplasmas were found in a mixed infection [20]. No putative or confirmed vectors of this phytoplasma are known. '*Ca.* Phytoplasma bonamiae' is a novel species as it fulfils Rule c [5]. This includes evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct geographic distribution and plant host range of '*Ca.* Phytoplasma bonamiae' compared to its closest known relative, strain BAWM-027 (PLL phytoplasma).

Description of '*Candidatus* Phytoplasma planchoniae' (CAWB phytoplasma, no previous subgroup assignment)

'*Candidatus* Phytoplasma planchoniae' (plan.chon'iae, N.L. gen. fem. n. planchoniae referring to *Planchonia*, the plant genus in which this phytoplasma has been detected). The reference sequence for '*Ca.* Phytoplasma planchoniae' is the whole genome sequence of strain BAWM-156b (GenBank assembly accession no.: JAOSIP000000000) originating from Australia.

'*Candidatus* Phytoplasma planchoniae' [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: JAOSIP000000000); G+C content, 24.0 mol%; highest genome ANI values with a representative '*Candidatus* Phytoplasma' species of 95.88% ANI (73.53% AF) with '*Ca.* Phytoplasma australasiaticum' strain PR08 (AAI is 95.72% for this comparison); unique oligonucleotide sequences of the 16S rRNA gene are (positions are in reference to nucleotide sequence alignments with M30790): TAGTAATAGGTATGCTTTAGGAGGGGCTTGCGCCATATTAGTTAGTTGGTAA (210–261 bp), A (736 bp), TAGTAATAGGTATGCTTTAGGAGGGGCTTGCGCCATATTAGTTAGTTGGTAA (210–261 bp), CAAAACGGTAGC CTAACTCGTTTATTCGAGAGGGCGTCGTCTAAGGTAGGGTCA (1431–1484 bp); P (*Planchonia careya*, phloem); M]. The reference sequence of this species represents a draft genome assembly.

To date, '*Ca.* Phytoplasma planchoniae' has only been detected in *Planchonia careya* (F. Muell.) [19], a tree in the family Lecythidaceae that is common across the tropical north of Australia [109]. Symptoms associated with this phytoplasma include witches'-broom and little leaf. '*Ca.* Phytoplasma planchoniae' has been detected in at least 13 locations in the far north of Queensland, Australia [19]. Although no putative or confirmed vectors are known, it is postulated that this phytoplasma may be transmitted by an insect that is monophagous on *P. careya* [19]. '*Ca.* Phytoplasma planchoniae' is a novel species as it fulfils Rule c [5]. This includes evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct geographic

distribution and plant host range of '*Ca.* Phytoplasma planchoniae' compared to its closest known relative, '*Ca.* Phytoplasma australasiaticum' (strain PR08).

Description of '*Candidatus* Phytoplasma fabacearum' (PLL phytoplasma, no previous subgroup assignment)

'*Candidatus* Phytoplasma fabacearum' (fa.ba.ce.a'rum N.L. gen. fem. pl. n. fabacearum meaning of *Fabaceae*'s, the plant family in which this phytoplasma is frequently detected). The reference sequence for '*Ca.* Phytoplasma fabacearum' is the whole genome sequence of strain BAWM-027 (GenBank assembly accession no.: JAOSIM000000000) originating from Australia.

'*Candidatus* Phytoplasma fabacearum'. [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: JAOSIM000000000); G+C content, 24.9 mol%; highest genome ANI values with a representative '*Candidatus* Phytoplasma' species of 93.45% ANI (86.85% AF) with phytoplasma strain BAWM-225 (AAI is 92.75% for this comparison); unique oligonucleotide sequences of the 16S rRNA gene are (positions are in reference to nucleotide sequence alignments with M30790): AGATATGAGGCATCTTGTATC (175–195 bp), CAAAACGGTAGTCTAACTCGTTTATTCGAGAGGGCGCCGT CTAAGGTAGGGTCG (1431–1484 bp); P (*Bituminaria bituminosa*, phloem); M]. The reference sequence of this species represents a draft genome assembly.

'*Ca.* Phytoplasma fabacearum' is generally detected in plants within the family Fabaceae and is detected less frequently in species within other plant families (e.g., Asteraceae, Aizoaceae, and Apocynaceae) [21]. This phytoplasma is associated with little leaf, phyllody, witches'-broom, stunting, proliferation, and yellowing symptoms [20, 66, 107, 110–112]. No putative or confirmed vectors of this phytoplasma are recorded. '*Ca.* Phytoplasma fabacearum' is a novel species as it fulfils Rule c [5]. This includes evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the different known geographic distributions and plant host ranges of '*Ca.* Phytoplasma fabacearum' and its closest known relative, '*Ca.* Phytoplasma bonamiae' (strain BAWM-225).

Description of '*Candidatus* Phytoplasma melaleucae' (16SrXXV-A)

'*Candidatus* Phytoplasma melaleucae' (me.la.leu'cae. N.L. gen. fem. n. melaleucae, of *Melaleuca*, referring to the plant genus from which this phytoplasma has been detected). The reference sequence for '*Ca.* Phytoplasma melaleucae' is the whole genome sequence of strain BAWM-155c (GenBank assembly accession no.: JAOSID000000000) originating from Australia.

'*Candidatus* Phytoplasma melaleucae'. [(Mollicutes) NC; NA; O, wall-less; NAS (GenBank assembly accession no.: JAOSID000000000); G+C content, 23.7 mol%; highest genome ANI values with a representative '*Candidatus* Phytoplasma' species of 77.85% ANI (26.84% AF) with phytoplasma strain BAWM-OMN-WBDL (AAI is 59.46% for this comparison); unique oligonucleotide sequences of the 16S rRNA gene are (positions are in reference to nucleotide sequence alignments with M30790): ACTTTTTAGGCCTCTAGAAAG (175–195 bp), ATTTATAGGTATGCTTAAGGAGGGGCTTGCGCCATATTAGTTAGTTGG TAGGGTAAT (209–268 bp), GGGAATTTTTGGCAATGGAGGCAAC – TCTGACCAAGCAACGCC – GCGTGAACGATG AAGTACTTCGGTAT (364–433 bp), TAGGGAAGAAAAAAGAGTGGAAAAACTCTCTTGACGGTACTTA (449–493 bp), GAGT (550–553 bp), TTGGCAAGTCAATAGTTTAATGGCAGTGCTCAACATTGTCCTGCTATTGAAACTGTCAGACTA (594–656 bp), CCTGCTATTGAAACTGTCAGACTAGAGTA (633–661 bp); AGGTCAAACTG (839–349 bp), GTTTTTGC AAAGTTATAGTAATATAATGGAGGTCACCAGAAAC (995–1037 bp), GAAGTTCGTGCCGTGAGG (1059–1076 bp); TCTTAAGTGG (1082–1091 bp); TC (1110–1111 bp), GGGATGACGTCAAATCATCATGCCCCTTATGG (1181–1212 bp), CTAA (1270–1273 bp), ATTACAAAGAGTAGCTGAAACGCGAGTTTTTAGCCAATCTCAAAAAGATAGTCTCAGTT (1241–1300 bp), GCACTTAACTTTGTAAAAAGAGAGAACC (1439–1468 bp); P (*Melaleuca* sp., phloem); M]. The reference sequence of this species represents a draft genome assembly.

'*Ca.* Phytoplasma melaleucae' has been detected on various *Melaleuca* species showing little leaf and witches-broom symptoms. The plant host species of '*Ca.* Phytoplasma melaleucae' include *Melaleuca leucadendra*, *M. cajuputi* subsp. *cajuputi*, *M. cajuputi* subsp. *platyphylla* in far north Queensland [66] as well as *M. cajuputi* subsp. *platyphylla* in the Western Province of Papua New Guinea (PNG) (PNG National Agricultural Quarantine and Inspection Authority; unpublished data). '*Ca.* Phytoplasma melaleucae' has also been detected on various islands in the Torres strait associated with *M. viridiflora* and *M. stenostachya* (unpublished data). In 2012, '*Ca.* Phytoplasma melaleucae' was also detected in an alternative host genus for the first time, namely *Synsepalum dulcificum* (family Sapotaceae) showing symptoms of witches'-broom and chlorosis in in Mossman, Queensland, Australia, potentially representing an incidental host (R. Davis, L. Jones; personal communication). In 2022, '*Ca.* Phytoplasma melaleucae' was detected for the first time in Kununnurra, Western Australia (unpublished data). The putative or confirmed vectors of this phytoplasma are unknown, but may be an insect with a feeding preference for species in the genus *Melaleuca*. '*Ca.* Phytoplasma melaleucae' is a novel species due to the <98.65% nucleotide sequence identity of the 16S rRNA gene compared to other '*Candidatus* Phytoplasma' species [113] and as it fulfils Rule c [5]. Evidence that Rule c was fulfilled includes evidence of molecular divergence based on the phylogenomic (Fig. 2) and comparative genomic analyses used in this study (ANI, AF, AAI, percent shared genes, Figs 3–5 and S2), as well as the distinct host association (i.e., with Melaleuca species).

CONCLUSIONS

Using multiple lines of evidence (phylogenomics using SCOs, ANI with AF, and AAI analysis) and taking known biological and ecological features into consideration, this study observed species boundaries between closely related phytoplasma strains classified within the 16SrII group. The within-species ANI and AAI threshold values were both identified to be >97%, with AF values of >80% in the ANI analyses. A genus boundary between the 16SrII and 16SrXXV phytoplasmas was also delimited, serving as the first observation of the genus rank within the '*Candidatus* Phytoplasma'. Strains belonging to two separate genera consistently shared <80% ANI with <40% AF and <60% AAI. Due to the high-resolution analysis in this study, potential subspecies could also be delimited for the first time in phytoplasma taxonomy. Strains belonging to the same subspecies shared ≥99.5% ANI, while separate subspecies in the same species cluster shared ANI values above 98.4% but below 99.5%. These observations of ANI, AF, and AAI threshold values for genus, species and subspecies boundaries have been previously described for other bacteria [7, 8, 29, 30, 114]. Using these genome-based criteria, six novel '*Candidatus* Phytoplasma' species were described in this study, and the data supported the delimitations of '*Ca.* Phytoplasma citri' (synonym '*Ca.* Phytoplasma aurantifolia') [14] and '*Ca.* Phytoplasma australasiaticum' (synonym '*Ca.* Phytoplasma australasia') [15] as distinct species. The description of '*Ca.* Phytoplasma australasia' has been rejected several times in the literature over the years [6, 87]. This has resulted in multiple incorrect classifications and identifications of 16SrII-D phytoplasmas as '*Ca.* Phytoplasma aurantifolia' in the literature and in sequence repositories. Therefore, for future identifications, we propose the use of their updated names that conform with the International Code of Nomenclature of Prokaryotes and its Orthography appendix, as previously recommended [89], to clear the confusion between these two taxa. The two putative genera identified in this study were not named as (i) they are assigned to distinct phytoplasma 16Sr groups which makes them clearly distinguishable, and (ii) we believe that the nomenclature of phytoplasma genera should be discussed with any associated recommendations outlined by members of the International Committee on Systematic Bacteriology Subcommittee on the Taxonomy of Mollicutes. The results of this study also confirm that the 16S rRNA gene does not provide meaningful taxonomic discrimination within the phytoplasmas at the species and subspecies levels, although it may place taxa within a genus designation, as shown for other bacteria [75, 82]. This study also emphasizes that, in the absence of correct species descriptions and nomenclature within the phytoplasmas, confusion may arise in the literature between which strains are representatives of a subgroup, or variants thereof. Since accurate species delimitation and identification is important for robust diagnostics [63], as well as for research on basic biology [69], disease epidemiology, and disease management [90], we agree with the recommendations of other authors that the use of additional gene regions or whole genome sequences for phytoplasma species delimitation is required alongside the 16S rRNA gene sequence [63].

Genetically diverse 16SrII phytoplasmas were analysed in this study, from a wide range of hosts, geographic regions, and sampling years, and represent a rich dataset that is suitable for uncovering the genetic diversity and biology both within and between phytoplasma species. However, obtaining complete phytoplasma genome sequences remains a challenge to genome-based systems of classification, as well as for research on their basic biology. It is therefore important to continue to develop phytoplasma enrichment procedures as well as methods to obtain axenic phytoplasma cultures. Additionally, the DNA and/or preserved plant material should be retained for strains in which only one or a few gene regions have been sequenced or for which only a draft genome has been assembled, as done in the present study and in previous work [38]. This allows future work to focus on improving the draft genome assemblies by using additional sequencing platforms, such as the long read sequencing afforded by Oxford Nanopore Technologies, that have been effective at obtaining complete phytoplasma assemblies [35]. Future work should also aim to improve the number of taxa for which sequence data is available within and between phytoplasma species, which will be facilitated by the decreasing cost of sequencing. With these types of datasets, we can begin to explore the genus boundaries of phytoplasmas according to criteria applied to bacteria [29, 30], and to clarify their classification within the order *Acholeplasmatales* [115] and the class *Mollicutes*.

Funding information
This project was funded by Hort Innovation using vegetable industry levies with co-investment from the Queensland Department of Agriculture and Fisheries; Victorian Department of Jobs, Precincts and Regions; the Northern Territory Department of Industry, Tourism and Trade; the Western Australia Department of Primary Industries and Regional Development, and the University of Tasmania. Hort Innovation is the grower-owned, not-for-profit research, and development corporation for Australian horticulture. We would also like to acknowledge Agriculture Victoria Research and La Trobe University for financial support of BRJ through their provision of a Living Allowance Scholarship and the La Trobe Full Fee Research Scholarship (LTUFFRS), respectively. This research was conducted using the facilities of Agriculture Victoria.

Acknowledgements
We thank growers for allowing sample collection on their properties, as well as colleagues who assisted in collecting or sending samples for the present study, including: Monica Kehoe and Craig Webster (both Department of Primary Industries and Regional Development, Western Australia, Australia), Wycliff M. Kinoti and David Lovelock (both Agriculture Victoria Research, Victoria, Australia), Samantha Bond (Department of Industry, Tourism and Trade, Northern Territory, Australia), Nandita Pathania (Department of Agriculture, Fisheries and Forests, Queensland, Australia), and David Tenakanai (Papua New Guinea National Agricultural Quarantine and Inspection Authority, Western Province, Papua New Guinea). We acknowledge Denis Lacaze and Jean-Said Bey for their work in the plant production and maintenance of the phytoplasma collection in *Catharanthus roseus* at INRAE Bordeaux. We would also like to thank Aharon Oren who contributed to the accurate nomenclature and etymology of the novel phytoplasma species and subspecies prior to submission.

Author contributions

B.R.J., participated in the design of the study and carried out the laboratory work; as well as data curation, analysis, and interpretation; B.R.J., drafted the original manuscript. A.M.A.-S., A.M.A.-S., C.G., F.C., H.C., J.-Y.Y., R.D., L.J., P.S., and X.F., participated in the collection and/or provision of samples. F.C., B.R., C.G., and L.T.T.-N. participated in the conceptualization of the study and provided supervision. A.M.A.-S., A.M.A.-S., B.R., C.G., F.C., H.C., J.-Y.Y., L.T.T.-N., R.D., L.J., and X.F., contributed to interpretation, reviewing, and editing of the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Lee IM, Davis RE, Gundersen-Rindal DE. Phytoplasma: phytopathogenic mollicutes. *Annu Rev Microbiol* 2000;54:221–255.
2. Hogenhout SA, Oshima K, Ammar E-D, Kakizawa S, Kingdom HN, *et al*. Phytoplasmas: bacteria that manipulate plants and insects. *Mol Plant Pathol* 2008;9:403–423.
3. Seemüller E, Marcone C, Lauer U, Ragozzino A, Göschl M. Current status of molecular classification of the Phytoplasmas. *J Plant Pathol* 1998;80:3–26.
4. Bradbury JM. Minutes of the interim meeting, 12 and 18 July 1996, Orlando, Florida, U.S.A. *Int J Syst Bacteriol* 1997;47:174–176.
5. Firrao G, Andersen M, Bertaccini A, Boudon E, Bové JM, *et al*. "*Candidatus* Phytoplasma", a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int J Syst Evol Microbiol* 2004;54:1243–1255.
6. Bertaccini A, Arocha-Rosete Y, Contaldo N, Duduk B, Fiore N, *et al*. Revision of the '*Candidatus* Phytoplasma' species description guidelines. *Int J Syst Evol Microbiol* 2022;72.
7. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, *et al*. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 2015;43:6761–6771.
8. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
9. Zhao Y, Wei W, Lee I-M, Shao J, Suo X, *et al*. Construction of an interactive online phytoplasma classification tool, iPhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). *Int J Syst Evol Microbiol* 2009;59:2582–2593.
10. Kämpfer P, Glaeser SP. Prokaryotic taxonomy in the sequencing era–the polyphasic approach revisited. *Environ Microbiol* 2012;14:291–317.
11. Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J* 2021;15:1879–1892.
12. Zhao Y, Wei W, Davis RE, Lee IM, Bottner-Parker KD. The agent associated with blue dwarf disease in wheat represents a new phytoplasma taxon, "*Candidatus* Phytoplasma tritici." *Int J Syst Evol Microbiol* 2021;71:1–11.
13. Omar AF, Aljmhan KA, Alsohim AS, Pérez-López E. Potato purple top disease associated with the novel subgroup 16SrII-X phytoplasma. *Int J Syst Evol Microbiol* 2018;68:3678–3682.
14. Zreik L, Carle P, Bové JM, Garnier M. Characterization of the mycoplasmalike organism associated with witches'-broom disease of lime and proposition of a Candidatus taxon for the organism, "'*Candidatus* Phytoplasma aurantifolia." *Int J Syst Bacteriol* 1995;45:449–453.
15. White DT, Billington SJ, Walsh KB, Scott PT. DNA sequence analysis supports the association of phytoplasmas with papaya (*Carica papaya*) dieback, yellow crinkle and mosaic. *Austral Plant Pathol* 1997;26:28.
16. Gibb K, Padovan AC, Mogen BD. Studies on sweet potato little-leaf phytoplasma detected in sweet potato and other plant species growing in Northern Australia. *Mol Plant Pathol* 1996;85:169.
17. Davis RI, Schneider B, Gibb KS. Detection and differentiation of phytoplasmas in Australia. *Aust J Agric Res* 1997;48:535–544.
18. Schneider B, Padovan A, De La Rue S, Eichner R, Davis R, *et al*. Detection and differentiation of phytoplasmas in Australia: an update. *Aust J Agric Res* 1999;50:333–342.
19. Davis RI, Jacobson SC, Waldeck GJ, De La Rue SJ, Gibb KS. A witches' broom of cocky apple (*Planchonia careya*) in north Queensland. *Austral Plant Pathol* 2001;30:179.
20. Padovan AC, Gibb KS. Epidemiology of phytoplasma diseases in papaya in Northern Australia. *J Phytopathol* 2001;149:649–658.
21. Streten C, Gibb KS. Phytoplasma diseases in sub-tropical and tropical Australia. *Austral Plant Pathol* 2006;35:129–146.
22. Kumari S, Nagendran K, Rai AB, Singh B, Rao GP, *et al*. Global status of phytoplasma diseases in vegetable crops. *Front Microbiol* 2019;10.
23. Chaturvedi Y, Rao G, Tiwari A, Duduk B, Bertaccini A. Phytoplasma on ornamentals: detection, diversity and management. *Acta Phytopathol Entomol Hungarica* 2010;45:31–69.
24. Hemmati C, Nikooei M, Al-Subhi AM, Al-Sadi AM. History and current status of phytoplasma diseases in the Middle East. *Biology*2021;10:226.
25. Trivellone V. An online global database of Hemiptera-Phytoplasma-Plant biological interactions. *Biodivers Data J* 2019;7:e32910.
26. Yadav A, Thorat V, Bhale U, Shouche Y. Association of 16SrII-C and 16SrII-D subgroup phytoplasma strains with witches' broom disease of *Parthenium hysterophorus* and insect vector *Orosius albicinctus* in India. *Australasian Plant Dis Notes* 2015;10.
27. Alhudaib K, Arocha Y, Wilson M, Jones P. Molecular identification, potential vectors and alternative hosts of the phytoplasma associated with a lime decline disease in Saudi Arabia. *Crop Protection* 2009;28:13–18.
28. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, *et al*. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 2014;12:635–645.
29. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, *et al*. A genus definition for bacteria and Archaea based on a standard genome relatedness index. *MBio* 2020;11:1–20.
30. Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species. *Microbe Magazine* 2014;9:111–118.
31. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, *et al*. A complete domain-to-species taxonomy for bacteria and archaea. *Nat Biotechnol* 2020;38:1079–1086.
32. Denancé N, Briand M, Gaborieau R, Gaillard S, Jacques M-A. Identification of genetic relationships and subspecies signatures in Xylella fastidiosa. *BMC Genomics* 2019;20:239.
33. Al-Subhi AM, Hogenhout SA, Al-Yahyai RA, Al-Sadi AM. Detection, identification, and molecular characterization of the 16SrII-D phytoplasmas infecting vegetable and field crops in Oman. *Plant Dis* 2018;102:576–588.
34. El-Sisi Y, Omar AF, Sidaros SA, Elsharkawy MM, Foissac X. Multilocus sequence analysis supports a low genetic diversity among '*Candidatus* Phytoplasma australasia' related strains infecting vegetable crops and periwinkle in Egypt. *Eur J Plant Pathol* 2018;150:779–784.
35. Chen Y-M, Chien Y-Y, Chen Y-K, Liao P-Q, Tan C-M, *et al*. Identification of 16SrII-V phytoplasma associated with mungbean phyllody disease in Taiwan. *Plant Dis* 2021;105:2290–2294.
36. Green MJ, Thompson DA, MacKenzie DJ. Easy and efficient DNA extraction from woody plants for the detection of phytoplasmas by polymerase chain reaction. *Plant Dis* 2007;83:482–485.

37. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 1991;173:697–703.
38. Rodrigues Jardim B, Kinoti WM, Tran-Nguyen LTT, Gambley C, Rodoni B, *et al*. '*Candidatus* Phytoplasma stylosanthis', a novel taxon with a diverse host range in Australia, characterised using multilocus sequence analysis of 16S rRNA, secA, tuf, and rp genes. *Int J Syst Evol Microbiol* 2021;71:1–11.
39. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
40. Bertaccini A, Lee I-M. Phytoplasmas: an update. In: Rao GP, Bertaccini A, Fiore N and Liefting LW (eds). *Phytoplasmas: Plant Pathogenic Bacteria - I*. Springer; 2018. pp. 1–29.
41. Yang Y, Jiang L, Tian Q, Lu Y, Zhang X, *et al*. Detection and identification of a novel subgroup 16SrII-V phytoplasma associated with *Praxelis clematidea* phyllody disease. *Int J Syst Evol Microbiol* 2017;67:5290–5295.
42. Al-Subhi A, Hogenhout SA, Al-Yahyai RA, Al-Sadi AM. Classification of a new phytoplasmas subgroup 16SrII-W associated with *Crotalaria* witches' broom diseases in Oman based on multigene sequence analysis. *BMC Microbiol* 2017;17:221.
43. Cai H, Chen HR, Li F, Kong BH. First report of a phytoplasma associated with cactus witches-broom in Yunnan (China). *Plant Pathol* 2002;51:394.
44. Rodrigues Jardim B, Tran-Nguyen LTT, Gambley C, Rodoni B, Constable FE. Iodixanol density gradients as an effective phytoplasma enrichment approach to improve genome sequencing. *Front Microbiol* 2022;13:937648.
45. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
46. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo assembler. *Curr Protoc Bioinformatics* 2020;70:e102.
47. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824–834.
48. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 2016;32:1088–1090.
49. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
50. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007;35:3100–3108.
51. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, *et al*. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–466.
52. Bushnell B, Rood J, Singer E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* 2017;12:e0185056.
53. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
54. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 2017;34:2422–2424.
55. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
56. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;16:157.
57. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:238.
58. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al*. Corrigendum to: IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
59. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
60. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* 2016;65:997–1008.
61. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
62. Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, *et al*. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 2018;556:452–456.
63. Cho S-T, Kung HJ, Huang W, Hogenhout SA, Kuo CH. Species boundaries and molecular markers for the classification of 16SrI phytoplasmas inferred by genome analysis. *Front Microbiol* 2020;11:1531.
64. Cai H, Wei W, Davis RE, Chen H, Zhao Y. Genetic diversity among phytoplasmas infecting *Opuntia* species: virtual RFLP analysis identifies new subgroups in the peanut witches'-broom phytoplasma group. *Int J Syst Evol Microbiol* 2008;58:1448–1457.
65. Cao Y, Trivellone V, Dietrich CH. A timetree for phytoplasmas (*Mollicutes*) with new insights on patterns of evolution and diversification. *Mol Phylogenet Evol* 2020;149:106826.
66. Davis RI, Jacobson SC, Rue S, Tran-Nguyen L, Gunua TG, *et al*. Phytoplasma disease surveys in the extreme north of Queensland, Australia, and the island of New Guinea. *Austral Plant Pathol* 2003;32:269–277.
67. Music MS, Samarzija I, Hogenhout SA, Haryono M, Cho S-T, *et al*. The genome of "*Candidatus* Phytoplasma solani" strain SA-1 is highly dynamic and prone to adopting foreign sequences. *Syst Appl Microbiol* 2019;42:117–127.
68. Huang C-T, Cho S-T, Lin Y-C, Tan C-M, Chiu Y-C, *et al*. Comparative genome analysis of "*Candidatus* phytoplasma luffae" reveals the influential roles of potential mobile units in phytoplasma evolution. *Front Microbiol* 2022;13:773608.
69. Oshima K, Kakizawa S, Nishigawa H, Jung H-Y, Wei W, *et al*. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat Genet* 2004;36:27–29.
70. Olson ND, Treangen TJ, Hill CM, Cepeda-Espinoza V, Ghurye J, *et al*. Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Brief Bioinform* 2018;20:1140–1150.
71. Koren S, Phillippy AM. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* 2015;23:110–120.
72. Tan CM, Lin Y-C, Li J-R, Chien Y-Y, Wang C-J, *et al*. Accelerating complete phytoplasma genome assembly by immunoprecipitation-based enrichment and MinION-based DNA sequencing for comparative analyses. *Front Microbiol* 2021;12:766221.
73. Cho S-T, Zwolińska A, Huang W, Wouters RHM, Mugford ST, *et al*. Complete genome sequence of "*Candidatus* phytoplasma asteris" RP166, a plant pathogen associated with rapeseed phyllody disease in Poland. *Microbiol Resour Announc* 2020;9:e00760-20.
74. Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, *et al*. SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat Microbiol* 2022;7:1702–1708.
75. Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, *et al*. Microbial genomic taxonomy. *BMC Genomics* 2013;14:913.
76. Gosselin S, Fullmer MS, Feng Y, Gogarten JP. Improving phylogenies based on average nucleotide identity incorporating saturation correction and nonparametric bootstrap support. *Syst Biol* 2022;71:396–409.
77. Al-Abadi SY, Al-Sadi AM, Dickinson M, Al-Hammadi MS, Al-Shariqi R, *et al*. Population genetic analysis reveals a low

level of genetic diversity of "*Candidatus* phytoplasma aurantifolia" causing witches' broom disease in lime. *Springerplus* 2016;5:1701.

78. Siampour M, Izadpanah K, Martini M, Salehi M. Multilocus sequence analysis of phytoplasma strains of 16SrII group in Iran and their comparison with related strains. *Ann Appl Biol* 2019;175:83–97.
79. Schneider B, Gibb KS, Padovan A, Davis RI, De La Rue S. Comparison and characterization of tomato big bud- and sweet potato little leaf-group phytoplasmas. *J Phytopathol* 1999;147:31–40.
80. Jeger M, Bragard C, Candresse T, Chatzivassiliou E, *et al*. Pest categorisation of witches' broom disease of lime (*Citrus aurantifolia*) phytoplasma. *EFSA J* 2017;15:e05027.
81. Salehi M, Izadpanah K, Siampour M, Bagheri A, Faghihi SM. Transmission of "*Candidatus* phytoplasma aurantifolia" to bakraee *(Citrus reticulata* hybrid) by feral *Hishimonus phycitis* leafhoppers in Iran. *Plant Dis* 2007;91:466.
82. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* 2007;45:2761–2764.
83. Palmer M, Steenkamp ET, Coetzee MPA, Chan W-Y, van Zyl E, *et al*. Phylogenomic resolution of the bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella*. *Antonie van Leeuwenhoek* 2017;110:1287–1309.
84. Maayer PD, Aliyu H, Cowan DA. Reorganising the order Bacillales through phylogenomics. *Syst Appl Microbiol* 2019;42:178–189.
85. Chang S-H, Cho S-T, Chen C-L, Yang J-Y, Kuo C-H. Draft genome sequence of A 16SrII-a subgroup phytoplasma associated with purple coneflower (*Echinacea purpurea*) witches' broom disease in Taiwan. *Genome Announc* 2015;3:e01398-15.
86. Gundersen DE, Lee IM, Rehner SA, Davis RE, Kingsbury DT. Phylogeny of mycoplasmalike organisms (phytoplasmas): a basis for their classification. *J Bacteriol* 1994;176:5244–5254.
87. Wei W, Zhao Y. Phytoplasma taxonomy: nomenclature, classification, and identification. *Biology* 2022;11:1119.
88. Parker CT, Tindall BJ, Garrity GM. International code of nomenclature of prokaryotes. *Int J Syst Evol Microbiol* 2019;69:S1–S111.
89. Oren A, Garrity GM, Parker CT, Chuvochina M, Trujillo ME. Lists of names of prokaryotic *Candidatus* taxa. *Int J Syst Evol Microbiol* 2020;70:3956–4042.
90. Malembic-Maher S, Desqué D, Khalil D, Salar P, Bergey B, *et al*. When a palearctic bacterium meets a nearctic insect vector: genetic and ecological insights into the emergence of the grapevine Flavescence dorée epidemics in Europe. *PLoS Pathog* 2020;16:e1007967.
91. Hemmati C, Nikooei M, Bertaccini A. Identification and transmission of phytoplasmas and their impact on essential oil composition in *Aerva javanica*. *3 Biotech* 2019;9:310.
92. Bagheri A, Faghihi MM, Khankahdani HH, Seyahooei MA, Ghanbari N, *et al*. First report of a phytoplasma associated with sapodilla flattened stem disease in Iran. *Australasian Plant Dis Notes* 2017;12:16–18.
93. Noorizadeh S, Golmohammadi M, Bagheri A, Bertaccini A. Citrus industry: phytoplasma-associated diseases and related challenges for Asia, America and Africa. *Crop Protection* 2022;151:105822.
94. Oren A. A plea for linguistic accuracy - also for *Candidatus* taxa. *Int J Syst Evol Microbiol* 2017;67:1085–1094.
95. Pilkington LJ, Gurr GM, Fletcher MJ, Nikandrow A, Elliott E. Vector status of three leafhopper species for Australian lucerne yellows phytoplasma. *Aust J Entomol* 2004;43:366–373.
96. Salehi M, Rasoulpour R, Izadpanah K. Molecular characterization, vector identification and partial host range determination of phytoplasmas associated with faba bean phyllody in Iran. *Crop Protection* 2016;89:12–20.
97. Özdemir Z. Phytoplasmas of sesame and *Orosius orientalis* are genetically diverse based on 16S rDNA sequencing and PCR-RFLP in Turkey. *Arch Phytopathol Plant Prot* 2017;50:674–686.
98. Özdemir Z. Identification of phytoplasmas from *Neoaliturus haematoceps* associated with sesame phyllody disease in southwestern Turkey. *J Phytopathol* 2018;166:242–248.
99. Aslam M, Tanwir S, Akhtar ZR, Ahmad JN. First report of 16Srii-D phyllody phytoplasma and associated insect vectors infecting multi-flower inbred lines of sunflower (*Helianthus Annuus* L.) in Faisalabad, Pakistan. *Pakistan J Agric Sci* 2021;58:985–992.
100. Cai H, Chen HR, Kong BH, Yang GH, Liu T. First report of a phytoplasma associated with Christmas cactus witches' broom. *Plant Pathology* 2007;56:346– .
101. Salehi M, Esmailzadeh Hosseini SA, Salehi E, Bertaccini A. Molecular and biological characterization of a 16Srii phytoplasma associated with carrot witches' broom in Iran. *J Plant Pathol* 2016;98:83–90.
102. Marzachì C, Coulibaly A, Coulibaly N, Sangaré A, Diarra M, *et al*. Cotton virescence phytoplasma and its weed reservoir in Mali. *J Plant Pathol* 2009;91:717–721.
103. Martini M, Lee I-M, Bottner KD, Zhao Y, Botti S, *et al*. Ribosomal protein gene-based phylogeny for finer differentiation and classification of phytoplasmas. *Int J Syst Evol Microbiol* 2007;57:2037–2051.
104. Khan AJ, Botti S, Al-Subhi AM, Gundersen-Rindal DE, Bertaccini AF. Molecular identification of a new phytoplasma associated with alfalfa witches'-broom in oman. *Phytopathology* 2002;92:1038–1047.
105. Bertaccini A. Plants and phytoplasmas: when bacteria modify plants. *Plants* 2022;11:1425.
106. Ghazanfar SA. Flora of the Sultanate of Oman. In: *Scripta Botanica Belgica*. 2007.
107. Schneider B, Padovan A, De La Rue S, Eichner R, Davis R, *et al*. Detection and differentiation of phytoplasmas in Australia. *Aust J Agric Res* 1999;50:333.
108. Myint T, Ward DB. A taxonomic revision of the genus *Bonamia* (*Convulvulaceae*). *Phytologia* 1968;17:121–239.
109. Barrett RL. A review of *Planchonia* (Lecythidaceae) in Australia. *Aust Syst Bot* 2006;19:147–153.
110. Wilson D, Blanche KR, Gibb KS. Phytoplasmas and disease symptoms of crops and weeds in the semi-arid tropics of the Northern Territory, Australia. *Austral Plant Pathol* 2001;30:159.
111. De La Rue S, Padovan A, Gibb K. *Stylosanthes* is a host for several phytoplasmas, one of which shows unique 16S-23S intergenic spacer region heterogeneity. *J Phytopathol* 2001;149:613–619.
112. De La Rue SJ, Hopkinson R, Foster S, Gibb KS. Phytoplasma host range and symptom expression in the pasture legume *Stylosanthes*. *F Crop Res* 2003;84:327–334.
113. Wei W, Davis RE, Lee I-M, Zhao Y. Computer-simulated RFLP analysis of 16S rRNA genes: identification of ten new phytoplasma groups. *Int J Syst Evol Microbiol* 2007;57:1855–1867.
114. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–351.
115. Zhao Y, Davis RE, Wei W, Lee I-M. Should "*Candidatus* phytoplasma" be retained within the order *Acholeplasmatales*? *Int J Syst Evol Microbiol* 2015;65:1075–1082.