A metagenomic investigation of phytoplasma diversity in Australian vegetable growing regions

Bianca Rodrigues Jardim[1,2,*], Cherie Gambley[3], Lucy T. T. Tran-Nguyen[4], Craig Webster[5], Monica Kehoe[5], Wycliff M. Kinoti[2], Samantha Bond[6], Richard Davis[7], Lynne Jones[7], Nandita Pathania[8], Murray Sharman[9], Toni Chapman[10], Brendan C. Rodoni[1,2] and Fiona E. Constable[1,2]

Abstract

In this study, metagenomic sequence data was used to investigate the phytoplasma taxonomic diversity in vegetable-growing regions across Australia. Metagenomic sequencing was performed on 195 phytoplasma-positive samples, originating either from historic collections (*n*=46) or during collection efforts between January 2015 and June 2022 (*n*=149). The sampled hosts were classified as crop (*n*=155), weed (*n*=24), ornamental (*n*=7), native plant (*n*=6), and insect (*n*=3) species. Most samples came from Queensland (*n*=78), followed by Western Australia (*n*=46), the Northern Territory (*n*=32), New South Wales (*n*=17), and Victoria (*n*=10). Of the 195 draft phytoplasma genomes, 178 met our genome criteria for comparison using an average nucleotide identity approach. Ten distinct phytoplasma species were identified and could be classified within the 16SrII, 16SrXII (PCR only), 16SrXXV, and 16SrXXXVIII phytoplasma groups, which have all previously been recorded in Australia. The most commonly detected phytoplasma taxa in this study were species and subspecies classified within the 16SrII group (*n*=153), followed by strains within the 16SrXXXVIII group ('*Ca*. Phytoplasma stylosanthis'; *n*=6). Several geographic- and host-range expansions were reported, as well as mixed phytoplasma infections of 16SrII taxa and '*Ca*. Phytoplasma stylosanthis'. Additionally, six previously unrecorded 16SrII taxa were identified, including five putative subspecies of '*Ca*. Phytoplasma australasiaticum' and a new putative 16SrII species. PCR and sequencing of the 16S rRNA gene was a suitable triage tool for preliminary phytoplasma detection. Metagenomic sequencing, however, allowed for higher-resolution identification of the phytoplasmas, including mixed infections, than was afforded by only direct Sanger sequencing of the 16S rRNA gene. Since the metagenomic approach theoretically obtains sequences of all organisms in a sample, this approach was useful to confirm the host family, genus, and/or species. In addition to improving our understanding of the phytoplasma species that affect crop production in Australia, the study also significantly expands the genomic sequence data available in public sequence repositories to contribute to phytoplasma molecular epidemiology studies, revision of taxonomy, and improved diagnostics.

Impact Statement

Phytoplasmas are unculturable, plant pathogenic bacteria that infect and impact yield of many agriculturally important plant species. In this study, 16S rRNA gene detection and sequencing for triaging was coupled with metagenomic sequencing to determine the diversity of phytoplasma taxa and associated diseases in Australian vegetable growing regions. It is the first study to use metagenomic analysis to improve the understanding of phytoplasma diversity in Australia. None of the phytoplasma taxa that were detected were exotic to Australia, but host- and geographic-range expansions were recorded for some. Since the metagenomic approach obtains DNA sequences from all organisms in a sample, the identification of plant and insect host families, genera, or species was possible by DNA barcode analysis when they were undetermined based on morphology. This study has analysed the largest number of phytoplasma whole genomes to date (*n*=195) and significantly contributes to the available sequence data for these bacteria. The sequence and metadata provided in this study offer an improved understanding of the phytoplasma taxa present in the different states and territories of Australia, and contributes to improving phytoplasma taxonomy, molecular epidemiology, and diagnostics.

DATA SUMMARY

The phytoplasma genomes analysed in this study were downloaded from the NCBI or sequenced for this study. The list of genomes that were already available on the NCBI are listed in Tables S1 and S2, available in the online version of this article, while those added from this publication are listed in Table S1 only and were submitted under the BioProject PRJNA980440. Tables S1 and S2; Fig. S1 are available with the online version of this article.

INTRODUCTION

Phytoplasmas are a diverse monophyletic clade of unculturable bacteria in the class *Mollicutes* within the provisional genus 'Candidatus Phytoplasma' [1]. They are phloem-limited plant pathogenic bacteria that are transmitted by phloem-feeding hemipteran insects [2]. Diseases associated with phytoplasma infections have been described in over 700 plant hosts globally from agriculturally and horticulturally important crops, ornamental plants, weeds, and native plants [3–5].

In Australia, phytoplasma-like symptoms were first recorded in the 1900s from tomato (Tomato Big Bud, TBB) [6], lucerne (Lucerne Witches'-Broom) [7], and pasture legumes (Legume Little Leaf) [8]. All were thought to be vector-transmitted viruses until mycoplasma-like structures were identified by electron microscopy in fabaceous plants showing symptoms of little leaves and spindled stems [9]. By the 1990s, molecular detections of phytoplasmas were being made globally [10], including in Australia [11], by PCR amplification of the 16S rRNA gene region. The molecular, PCR-based techniques facilitated the screening of many more plants for phytoplasma infections than was previously possible by microscopy and serological techniques. These PCR-based techniques involved restriction fragment length polymorphism (RFLP) analysis and/or sequencing for phytoplasma characterization and identification. Phytoplasma surveys were done in Australia using these molecular-based approaches and a large diversity of phytoplasma taxa affecting grains, legumes, fruit and vegetable crops, ornamentals, native plants, weeds and putative phytoplasma vectors were uncovered [12–18]. To date, twelve phytoplasma 16 Sr groups and five 'Candidatus Phytoplasma' representatives have been described from Australia [19–25]. Members within the 16SrII group are the most commonly detected phytoplasma in Australia, affecting a large number of host species [12, 13, 16, 26]. However, members within the 16SrXII group have been described as the most economically important based on their association with yield-reducing diseases in high-value crops such as grapevines and strawberries [26–28]. Competent vector species remain to be confirmed for many of the phytoplasma taxa present in Australia, although *Orosius argentatus* has been shown to transmit diseases associated with 16SrII phytoplasmas [8, 9, 29, 30].

Phytoplasma diversity and taxonomic analyses, including studies in Australia, have largely relied on the 16S rRNA gene sequence, which only offers low-resolution analyses of inter- and intraspecies diversity [1, 12, 31, 32]. Diversity assessments and species delimitation studies have also involved higher-resolution analyses of three to nine additional housekeeping genes for multilocus sequence typing (MLST) and analysis (MLSA) [24, 33, 34]. With the decreasing cost of high-throughput sequencing (HTS) and increasing sequence data outputs, as well as the advancements in bioinformatic tools for metagenomic data assessments, phytoplasma genomes have increasingly been used to understand their taxonomy [31] and biology [35]. Obtaining draft or complete phytoplasma genomes allows for higher-resolution analyses than those based on one or a few genes for diversity analyses [36]. Additionally, by applying genome-based species delimitation thresholds and criteria specified for culturable bacteria, taxonomic boundaries between phytoplasma strains can be identified [36, 37].

The aim of this study was to assess and update the species and genetic diversity of phytoplasmas in vegetable growing regions in the various states and territories of Australia using whole-genome-based approaches. To this end, (i) plants displaying phytoplasma associated symptoms and insects were collected from vegetable growing regions in Australia and screened by PCR for the presence of phytoplasma infections, (ii) DNA from key historical phytoplasma strains from previous phytoplasma surveys

in Australia were obtained, and (iii) a metagenomic approach was used to obtain draft phytoplasma genome assemblies to be used for genome-based investigations into the phytoplasma taxa infecting the samples, and to identify the host when the plant or insect host identity was inconclusive based on morphology. The results of this study demonstrate the applications, benefits, and challenges of applying metagenomic sequencing to phytoplasma diversity analyses.

METHODS

Sample collection, total nucleic acid extraction, plant tissue preservation, and DNA quantification

Plant samples with phytoplasma-associated symptoms, including little leaf, yellowing, phyllody, and/or stunting that were collected Australia-wide between March 2019 and June 2022 were sent to the laboratory in Melbourne, Victoria (VIC) for analysis (Table S1). Total nucleic acid was extracted from samples using an iodixanol-based phytoplasma enrichment procedure [38] or a modified CTAB-DNeasy protocol without RNase treatment [39] (Table S1 and S2). Petioles, whole leaves or leaf veins were used in total nucleic acid extractions for most samples. Phloem scrapings were sampled for woody material (e.g. *Melaleuca* spp. and *Vitaceae* spp.). When possible, a subsample of the plant material was freeze dried for at least 72 h at −50°C in individually labelled screw cap tubes using the FreeZone 2.5 Liter Benchtop Freeze Dry System (Labconco, MO, USA) and deposited in the Victorian plant pathology herbarium (VPRI) (Table S1). Insect samples were supplied as DNA extracts (Table S1) and had been collected by suction trapping and sweep netting in Jennings, New South Wales (NSW) and Palmerston, Northern Territory (NT), respectively. Insect collections were done in these areas as plants displaying phytoplasma-associated symptoms were present nearby. Phytoplasma-positive samples collected prior to 2019 that are held at The Northern Territory Department of Industry, Tourism and Trade (NT DITT), Darwin, NT, Australia and Department of Agriculture and Fisheries, Mareeba, Queensland (QLD), Australia were also supplied as total nucleic acid extracts (Table S1).

DNA quantity was estimated using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, MA, USA) with the Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, MA, USA). All DNA samples were stored at −20°C.

Phytoplasma screening and preliminary identification by universal phytoplasma 16Sr PCR, and Sanger sequencing

Screening for PCR inhibitors was done using PCR primers for the generic amplification of the bacterial 16S rRNA gene [40]. A nested PCR assay using P1/P7 and R16F2n/m23sr primer pairs were used to screen recent samples for phytoplasma infection and to confirm phytoplasma presence in total nucleic acid extracts from samples collected prior to 2019 [41]. These primers all bind to the phytoplasma 16S rRNA gene, apart from the P7 primer, which binds to the 5′ region of the 23S rRNA gene. The R16F2n/m23sr amplicon of some samples were cloned and screened according to [24] when a poor Sanger sequencing quality was observed in the forward and/or reverse read (Table S1) [24]. All PCR amplicons were visualized by electrophoresis through 1% agarose gels stained with SYBR Safe DNA gel stain (Thermo Fisher Scientific, MA, USA). PCR amplicons of the expected size were purified and directly Sanger sequenced (Macrogen, Seoul, South Korea). The identities of the Sanger sequenced PCR amplicons were determined by BLASTn analysis [42] at the NCBI (https://www.ncbi.nlm.nih.gov/, last accessed 30 July 2022). During the BLASTn analyses, the top hit with the 16S rRNA gene of a 'Ca. Phytoplasma' species reference strain was used to determine the identity of the sample investigated based on the top bit score, percent identity, and e-value, as well as considering the query coverage.

Library preparation and sequencing

Libraries were prepared with fragment sizes between 300 and 500 bp by following the manufacturer's protocols of the NEXTFLEX Rapid XP DNA-Seq Kit (PerkinElmer, MA, USA) with the Unique Dual Index (UDI) barcodes (PerkinElmer, MA, USA) or the Nextera DNA Flex Library Preparation Kit (Illumina, CA, USA) with the IDT for Illumina Nextera DNA Unique Dual (UD) Indexes (Illumina, CA, USA) (Table S1). The libraries were pooled and size-selected prior to sequencing according to [38] and sequenced with Illumina platforms including the MiSeq (2×250 bp), HiSeq2000 (2×150 bp), NovaSeq 6000 on an SP flow cell (2×250 bp), or NovaSeq 6000 on an S1 flow cell (2×150 bp). All the library pools sequenced on the NovaSeq 6000 platform were treated with the Illumina Free Adapter Blocking Reagent (Illumina, CA, USA) according to the manufacturer protocols prior to HTS to mitigate aberrant sequencing results caused by the presence of free adapters.

The phytoplasma genome sequences of 25 phytoplasma-positive samples have been used in previous studies [24, 37, 38]. The genome sequences and metadata associated with these samples were included in this study, however, they were gathered during the sample collection period of this study (Table S2).

Sequence data analyses

Read quality filtering, metagenomic assembly, and identification of phytoplasma-derived contigs and gene annotations

Illumina read filtering and adapter trimming for each sample was done using FastP [43], removing reads shorter than 50 bp and with a Phred quality score (Q score) below Q20. The trimmed reads were used in a metagenomic assembly pipeline according to [38], which implements metaSPAdes version 3.15.2 [44, 45]. Phytoplasma-derived contigs were identified and retrieved using

BLAST+v2.11.0 [42] and a custom *grep script, respectively. Contigs shorter than 500 bp were removed using the reformat.sh* script implemented in the BBMap v.38.61b software suite [46]. The phytoplasma genomes were analysed in metaQUAST [47] to estimate the genome N50 values. Protein coding, tRNA, and rRNA genes were annotated and counted using Prokka [48], specifying RNAmmer for 5S, 16S, and 23S rRNA gene annotations [49].

Selection of phytoplasma genomes for comparative genomic assessments
All phytoplasma genomes available on the NCBI (https://www.ncbi.nlm.nih.gov/) were downloaded, apart from the '*Ca*. Phytoplasma citri' (synonym '*Ca*. Phytoplasma aurantifolia') genome under the RefSeq accession no. GCF_002009625.1 (Table S2). Instead, the re-sequenced assembly of this strain described in [37] was used (Table S2).

The phytoplasma genomes obtained in this study were analysed differently based on the total genome size recovered to mitigate spurious results related to highly incomplete and fragmented genomes. The number of tRNA gene sequences annotated were used to estimate the completeness of the phytoplasma assemblies and, therefore, their suitability for the downstream genome-based analyses. Phytoplasma genomes that were larger than 300,000 bp and which encoded 13 or more tRNA genes were used in further whole-genome-based assessments to classify the phytoplasma strains (Table S1). For genomes smaller than 300,000 bp and/or encoding fewer than 13 tRNA genes, only the 16S rRNA gene was used to classify the phytoplasma strain within a 16 Sr group (Table S1).

Whole-genome analyses
Whole-genome comparisons were performed for all the phytoplasma genomes by average nucleotide identity (ANI) analysis using the ANI with MUMmer (ANIm) algorithm in pyani version 0.2.10 [50]. The pyani heatmap output was manually overlaid with BLASTn results to characterize clusters, which did not contain a representative genome.

For further investigation into phytoplasma samples that did not clearly cluster with reference sequences in the initial ANI analysis, the coverage of aligned genomic segments in each pairwise comparison was analysed along with publicly available phytoplasma reference genomes using pyani version 0.20.10. These values are referred to as the alignment fraction (AF).

In cases where a historic strain was previously classified using *in vitro RFLP but appeared to be a different species based on the maximum-likelihood tree,* the recognition sites for 14 of the 17 restriction enzymes used in phytoplasma subgroup classification were visualized in Geneious Prime version 2022.2.2 (https://www.geneious.com/prime/). The three restriction enzymes that were not used as they are unavailable on Geneious Prime version 2022.2.2 are: *Bfa*I, *Bst*UI (*Tha*I), and *Ssp*I.

Identification of mixed phytoplasma infections using whole-genome analyses
Samples which demonstrated high ANI values (>90% ANI) with more than one representative phytoplasma genome were considered to have a mixed phytoplasma infection of the phytoplasma taxa for which genome sequence data was available. The 16S rRNA gene sequence analyses, including those of the cloned PCR amplicons, as well as the number of annotated tRNA genes were revisited for the mixed infection samples identified during the BLASTn analyses.

Identification of unknown host species
The host family, genus, or species of some plant hosts and two insect hosts were unconfirmed based on external morphology prior to total nucleic acid extraction (Table S1). To confirm the identities of the insects to the species-level, contigs encoding *cytochrome c oxidase subunit 1* (*coI*) were analysed [51]. To confirm the family, genus, and/or species of the unknown plant hosts, the contigs encoding *maturase K* (*matK*) and *ribulose-bisphosphate carboxylase* (*rbcL*) genes were analysed [52, 53]. These barcoding genes were obtained from the metagenome assemblies using a custom grep script to identify the genes and their identities from the BLASTn results. A publicly available accession was listed as the top BLASTn hit when it produced the highest bit-score, lowest e-value, and highest percentage identity. In cases where multiple species in the same genus were considered the top BLASTn hit, only the genus was recorded for the sample and the species names were considered undetermined. In cases where multiple genera were listed as top BLASTn hits, the family of these samples were recorded, and the genus and species names were considered undetermined. The detection location of host species listed as top hits were determined based on searches performed at the Australasian Virtual Herbarium website (https://avh.chah.org.au/, last accessed March 2023), with hosts that are not known to be present in Australia removed from the list.

RESULTS AND DISCUSSION
Phytoplasma-positive sample information
A total of 195 samples were collected between 1998 and 2022 and were either confirmed or suspected to be infected with phytoplasma based on previous PCR-based analyses (sequence similarity and RFLP) or disease symptoms. These samples were subsequently confirmed to be positive for phytoplasma by PCR and direct Sanger sequencing of the amplicon generated using the

universal phytoplasma 16S rRNA primers. Most of the samples were collected from QLD (*n*=78), followed by Western Australia (WA, *n*=46), NT (*n*=32), NSW (*n*=29), and VIC (*n*=10) (Tables 1 and S1).

Phytoplasma samples analysed in this study were collected from all the states and territories of Australia, apart from the Australian Capital Territory (ACT), South Australia (SA), and Tasmania (TAS) (Tables 1 and S1). No samples were collected between 2019 and 2022 in SA due to Covid-19 travel restrictions. There was also an absence of plants showing typical phytoplasma symptoms in TAS during the 2019–2022 collection period (Callum R. Wilson, personal communication), which corresponds to previous observations of low phytoplasma prevalence for the state [54].

An asterisk (*) in the sample name(s) column indicates samples for which the host identity was determined using DNA barcode analysis.

BLASTn of the 16S rRNA gene for phytoplasma-positive samples

Based on BLASTn analysis of the 16S rRNA gene PCR amplicons (Table 2 and S1), samples had single top hits with ten different phytoplasma taxa, including '*Ca.* Phytoplasma australasiaticum subspecies australasiaticum' (*n*=91), '*Ca.* Phytoplasma australasiaticum subspecies ipomoeae' (*n*=69), '*Ca.* Phytoplasma fabacearum' (*n*=15), '*Ca.* Phytoplasma bonamiae' (*n*=2), '*Ca.* Phytoplasma planchoniae' (*n*=1, BAWM-156b), '*Ca.* Phytoplasma stylosanthis' (*n*=6), '*Ca.* Phytoplasma melaleucae' (*n*=2), '*Ca.* Phytoplasma australiense' (*n*=1, BAWM-189), ViLL phytoplasma (*n*=3, BAWM-245 shared a lower nucleotide identity than BAWM-336 and BAWM-337 with the strain at the NCBI), and *Waltheria* Little Leaf phytoplasma (WaLL, *n*=1, BAWM-227). Cloning and Sanger sequencing of the 16S rRNA gene PCR amplicon done for samples BAWM-252, BAWM-253, BAWM-255, and BAWM-257 identified the presence of mixed phytoplasma infections of '*Ca.* Phytoplasma australasiaticum subspecies ipomoeae' and '*Ca.* Phytoplasma stylosanthis' for each sample (Tables 2 and S1). All of these phytoplasma taxa identified based on BLASTn analyses in this study have been previously detected and reported in Australia based on 16S rRNA-based sequence analysis [12–15, 18, 26, 55, 56]

Metagenomic sequence data outputs and metagenome-based host identifications

Metagenomic sequencing output and phytoplasma genome information

After metagenomic HTS, 178 of the 195 total samples passed the phytoplasma genome criteria for further comparative genomic analyses in this study (Table S1). The Illumina sequence data output for these 178 samples ranged from 0.31 Gb (sample BAWM-193a-F1) to 32.60 Gb (sample BAWM-354A) with an average output of 4.76 Gb per metagenomic library (Table S1). The phytoplasma genome sizes of these 178 samples ranged between 321 651 bp (sample BAWM-201) and 1 488 020 bp (sample BAWM-255), with an average genome size of 632,634 bp for of all 178 phytoplasma samples (Table S1). Of the 178 draft genomes, an average of 28 tRNA gene sequences were recovered per genome and ranged between 13 tRNA genes (sample BAWM-198) to 61 tRNA genes (sample BAWM-255) (Table S1). The most tRNA genes recovered from a complete phytoplasma genome to date is 35 from the genome of '*Ca.* Phytoplasma australiense' strain NZSb11 [57], indicating that the phytoplasma genome data of 13 of the 178 samples that had more than 35 tRNA genes annotated may represent tRNAs of more than one phytoplasma species in the sample (Table S1). An average of two phytoplasma rRNA genes could be annotated from these 13 genomes, with a max of five rRNA genes (sample BAWM-307) and none obtained from sample BAWM-350 (Table S1). To date, two identical or nearly identical 16S rRNA genes are known to be encoded per phytoplasma genome [1, 36] indicating that sample BAWM-307 potentially harbours a mixed phytoplasma infection.

Phytoplasma genome sequences that were <300 000 bp were recovered for 17 samples in six different host families (Table S1). These genome sequences ranged in size from 1784 bp (sample BAWM-004) to 289 060 bp (sample BAWM-184), with an average size of 114 181 bp (Table S1). The average data output for these 17 samples was 5.07 Gb, with a range of 1.28 Gb (sample BAWM-173) to 22.53 Gb (sample BAWM-189) (Table S1). An average of six tRNA genes could be retrieved from these 17 phytoplasma genomes (range 0 tRNA genes for samples BAWM-003, BAWM-004, BAWM-083, BAWM-216, and BAWM-233 to 21 tRNA genes for BAWM-183). No rRNA genes were retrieved from seven of the 17 phytoplasma genomes. These samples were not used in further genomic-based analyses as these results indicate poor-quality genomes from which limited information can be obtained, including the 16S rRNA gene-based correlation of taxon identification prior to and after metagenomic HTS [58].

Phytoplasma taxon identification using whole-genome comparisons using ANI estimates and 16S rRNA gene analyses

Subspecies of '*Ca.* Phytoplasma australasiaticum'

Based on whole-genome ANI analyses (Fig. 1; Table S1), 160 of the 178 samples (ca. 90%) used in further genome-based analyses clustered at >96% ANI solely with representative genome sequences of 16SrII phytoplasmas. The majority of these samples classified within the 16SrII phytoplasma group were identified as '*Ca.* Phytoplasma australasiaticum' subspecies, including of '*Ca.* Phytoplasma australasiaticum subsp. australasiaticum' (*n*=67), '*Ca.* Phytoplasma australasiaticum subsp. ipomoeae' (*n*=51), and strains identified as a new '*Ca.* Phytoplasma australasiaticum' subspecies (*n*=12, referred to as '*Ca.* Phytoplasma australasiaticum' taxon 1) (Fig. 1).

Table 1. Summary of all host species, genera, and families investigated in this study and associated metadata organized according to their sampling location in Australia (state/territory and closest town/city). Metadata recorded included the sample names, sampling years, and the phytoplasma taxa identified based on BLASTn or ANI analyses performed in this study. Original sample names, if provided, are recorded alongside the corresponding 'BAWM' name in Table S1

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
NSW					
Jennings					
	Cicadellidae	*Orosius argentatus*	'Ca. P. stylosanthis' 'Ca. P. australasiaticum subsp. australasiaticum'	2021 2021	BAWM-342B* BAWM-343A*
	Solanaceae	*Capsicum annuum*	'Ca. P. australasiaticum subsp. australasiaticum' 'Ca. P. australasiaticum subsp. ipomoeae'	2020 2020 2020 2020 2020 2020 2021 2021 2021 2021 2017 2020 2021	BAWM-117 BAWM-118 BAWM-119 BAWM-124 BAWM-126 BAWM-131a BAWM-192 BAWM-194 BAWM-195 BAWM-203 BAWM-054 BAWM-135 BAWM-193a-F1
	Solanaceae	*Solanum lycopersicum*	'Ca. P. australasiaticum subsp. australasiaticum'	2017 2017 2020 2020 2020 2020 2020 2020 2020 2020 2020 2021 2021	BAWM-049 BAWM-053 BAWM-113 BAWM-114 BAWM-115 BAWM-116 BAWM-125a BAWM-127 BAWM-128 BAWM-134a BAWM-134b BAWM-204 BAWM-205
Mudgee					
	Vitaceae	*Vitis vinifera cv.* Riesling	'Ca. P. australasiaticum subsp. australasiaticum'	2021	BAWM-233
Wallicia					
	Solanaceae	*Capsicum annuum*	'Ca. P. australasiaticum subsp. australasiaticum'	2020	BAWM-131a
NT					
Berry Springs					
	Fabaceae	*Vigna unguiculata* ssp. *sesquipedalis*	'Ca. P. australasiaticum subsp. ipomoeae'	2019 2019	BAWM-082 BAWM-083
Darwin					
	Asteraceae	*Gynura crepioides* (Okinawa Spinach)	'Ca. P. australasiaticum' taxon 1	2021	BAWM-338
	Convolvulaceae	*Ipomoea* sp.	New 16SrII species	2021	BAWM-339
	Solanaceae	*Solanum melongena*	'Ca. P. australasiaticum' taxon 1 'Ca. P. australasiaticum subsp. australasiaticum'	2004 2005 2005 2005 2005	BAWM-293 BAWM-295 BAWM-296 BAWM-298 BAWM-297
Girraween					
	Solanaceae	*Solanum melongena*	'Ca. P. australasiaticum' taxon 1	2004 2004 2004	BAWM-289 BAWM-290 BAWM-291
Howard Island					

Continued

Table 1. Continued

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
	Convolvulaceae	*Ipomoea batatas*	'*Ca.* P. australasiaticum subsp. australasiaticum'	1998	BAWM-259
Humpty Doo					
	Cucurbitaceae	*Luffa acutangula*	'*Ca.* P. australasiaticum subsp. ipomoeae'	2018 2018	BAWM-300 BAWM-301
	Cucurbitaceae	*Momordica charantia*	'*Ca.* P. australasiaticum subsp. ipomoeae'	2018	BAWM-303
	Cucurbitaceae	*Trichosanthes cucumerina*	'*Ca.* P. australasiaticum' taxon 3	2018	BAWM-306
	Fabaceae	*Vigna unguiculata* ssp. *sesquipedalis*	'*Ca.* P. fabacearum'	2004	BAWM-286
Katherine					
	Caricaceae	*Carica papaya*	'*Ca.* P. stylosanthis'	1998	BAWM-249
	Convolvulaceae	*Bonamia pannosa*	'*Ca.* P. bonamiae' '*Ca.* P. bonamiae'	1998 1998	BAWM-225 BAWM-226
	Malvaceae	*Waltheria* sp.	*Waltheria* Little Leaf (WaLL) phytoplasma	1998	BAWM-227
Lambells lagoon					
	Cucurbitaceae	*Luffa acutangula*	'*Ca.* P. australasiaticum subsp. ipomoeae'	2019	BAWM-081
Marrakai					
	Cucurbitaceae	*Luffa acutangula*	'*Ca.* P. australasiaticum subsp. ipomoeae'	2016	BAWM-299
	Cucurbitaceae	*Momordica charantia*	*Vigna* Little Leaf (ViLL) phytoplasma	2021 2021	BAWM-336 BAWM-337
Palmerston					
	Cicadellidae	*Orosius orientalis*	'*Ca.* P. australasiaticum subsp. ipomoeae'	2020	BAWM-232
	Fabaceae	*Arachis hypogaea*	'*Ca.* P. australasiaticum subsp. ipomoeae'	2019	BAWM-080
	Fabaceae	*Crotalaria* sp.	'*Ca.* P. australasiaticum subsp. ipomoeae'	2020	BAWM-164
	Solanaceae	*Solanum* sp.	Mixed infection ('*Ca.* P. stylosanthis and '*Ca.* P. australasiaticum' taxon 1)	2019	BAWM-079*
Unknown					
	Solanaceae	*Solanum melongena*	'*Ca.* P. australasiaticum subsp. australasiaticum'	2004	BAWM-287
	Solanaceae	*Solanum* sp.	'*Ca.* P. australasiaticum' taxon 1	2018	BAWM-302*
QLD					
Atherton					
	Fabaceae	*Stylosanthes scabra*	'*Ca.* P. stylosanthis' Mixed infection ('*Ca.* P. stylosanthis and '*Ca.* P. australasiaticum subsp. ipomoeae')	1998 1998 1998 1998 1998	BAWM-254 BAWM-256 BAWM-258 BAWM-253 BAWM-257
Bowen					
	Solanaceae	*Capsicum annuum*	'*Ca.* P. fabacearum'	2020	BAWM-168

Table 1. Continued

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
Brisbane					
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2021 2021	BAWM-217 BAWM-218
Bundaberg					
	Malvaceae	*Sida* sp.	'*Ca*. P. australasiaticum subsp. australasiaticum'	2019	BAWM-037*
Cairns					
	Lecythidaceae	*Planchonia careya*	'*Ca*. P. planchoniae'	2020	BAWM-156b
	Myrtaceae	*Melaleuca* sp.	'*Ca*. P. melaleucae'	2020	BAWM-155a
Croydon					
	Fabaceae	*Glycine max*	'*Ca*. P. fabacearum'	2017	BAWM-350
Dalby					
	Fabaceae	*Vigna radiata*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2017	BAWM-349
Davies Creek					
	Pedaliaceae	*Phyllanthus fuernrohrii*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2018	BAWM-313
Dimbulah					
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum''*Ca*. P. australasiaticum subsp. ipomoeae'	2002 2002 2002 2002 2002 2002 2002 2002 2002	BAWM-260 BAWM-261 BAWM-264 BAWM-265 BAWM-266 BAWM-269 BAWM-262 BAWM-263 BAWM-267
Emerald					
	Fabaceae	*Cajanus cajan*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020	BAWM-351
Gatton					
	Solanaceae	*Capsicum annuum*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2002 2002	BAWM-276 BAWM-277
Glen Aplin					
	Solanaceae	*Capsicum annuum*	'*Ca*. P. australasiaticum subsp. australasiaticum' '*Ca*. P. fabacearum'	2021 2021 2021	BAWM-196 BAWM-198 BAWM-197
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2021 2021	BAWM-206 BAWM-207
Gordonvale					
	Fabaceae	*Arachis hypogaea*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020	BAWM-352
Helidon					
	Solanaceae	*Solanum lycopersicum*	 '*Ca*. P. fabacearum' '*Ca*. P. australasiaticum subsp. australasiaticum'	2021 2021	BAWM-234 BAWM-235
Lyra					

Table 1. Continued

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
	Solanaceae	*Capsicum annuum*		2021	BAWM-201
			'*Ca*. P. australasiaticum subsp. australasiaticum'	2021	BAWM-211a-F3
				2017	BAWM-060
			'*Ca*. P. australasiaticum subsp. ipomoeae'	2021	BAWM-238
			'*Ca*. P. australasiaticum' taxon 1	2021	BAWM-237
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2017	BAWM-055
				2021	BAWM-208
				2021	BAWM-209
				2021	BAWM-210
Mareeba					
	Asteraceae	*Peripleura diffusa*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2015	BAWM-310
	Asteraceae	*Praxelis clematidea*		2018	BAWM-314
			'*Ca*. P. australasiaticum subsp. ipomoeae'	2018	BAWM-320
			Mixed infection ('*Ca*. P. fabacearum' and '*Ca*. P. australasiaticum subsp. australasiaticum)		
	Asteraceae	*Chromolaena odorata*	Mixed infection ('*Ca*. P. fabacearum' and '*Ca*. P. australasiaticum subsp. ipomoeae)	2018	BAWM-316*
	Asteraceae	*Asteraceae* sp.	'*Ca*. P. fabacearum'	2018	BAWM-321*
	Caricaceae	*Carica papaya*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2017	BAWM-309
	Fabaceae	*Cajanus cajan*	'*Ca*. P. fabacearum'	2017	BAWM-323
	Fabaceae	*Crotalaria* sp.	'*Ca*. P. australasiaticum subsp. ipomoeae'	2018	BAWM-315*
	Fabaceae	*Stylosanthes scabra*		2018	BAWM-324
			'*Ca*. P. australasiaticum' taxon 5	2017	BAWM-322
			'*Ca*. P. fabacearum'	1998	BAWM-252
			Mixed infection ('*Ca*. P. stylosanthis and '*Ca*. P. australasiaticum subsp. ipomoeae')	1998	BAWM-255
				2018	BAWM-326
	Malvaceae	*Sida* sp.	'*Ca*. P. australasiaticum' taxon 4	2018	BAWM-319*
Severnlea					
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2017	BAWM-056
	Solanaceae	*Solanum* sp.	'*Ca*. P. australasiaticum' taxon 2	2017	BAWM-057*
Stanthorpe					
	Apiaceae	*Apium graveolens*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2002	BAWM-273
				2002	BAWM-282
				2002	BAWM-283
	Apiaceae	*Petroselinum crispum*		2021	BAWM-236-TSE
			'*Ca*. P. australasiaticum subsp. australasiaticum'		
	Solanaceae	*Capsicum annuum*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2021	7C
Thorndale					
	Asteraceae sp.	*Bidens pilosa*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2017	BAWM-063

Table 1. Continued

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2017	BAWM-050
	Solanaceae	*Solanum melongena*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2017	BAWM-058
Walkamin					
	Fabaceae	*Glycine max*	'*Ca*. P. australasiaticum' taxon 1 '*Ca*. P. fabacearum'	2020 2020	BAWM-150b BAWM-151
	Fabaceae	*Cajanus cajan*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2017	BAWM-311
	Fabaceae	*Crotalaria juncea*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2018	BAWM-312
Wyberba					
	Asteraceae	*Osteospermum* sp.	'*Ca*. P. australasiaticum subsp. australasiaticum'	2021	BAWM-215
	Solanaceae	*Petunia* sp.	'*Ca*. P. australasiaticum subsp. australasiaticum'	2021	BAWM-216
Unknown					
	Solanaceae	*Capsicum annuum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2002 2020	BAWM-132 BAWM-133 BAWM-270
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. ipomoeae'	1998	BAWM-250 BAWM-251
VIC					
Healesville					
	Rosaceae	*Fragaria ×Ananassa Duch. Cv. Pajaro*	'*Ca*. P. australasiaticum' taxon 1	2017	BAWM-073
Kallista					
	Geraniaceae	*Geranium* sp.	'*Ca*. P. australasiaticum subsp. australasiaticum'	2021	BAWM-246
Little Hampton					
	Asteraceae	*Echinacea purpurea*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2020	BAWM-123a
Melbourne					
	Solanaceae	*Solanum tuberosum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2019	4P
Mildura					
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2019	BAWM-001
	Solanaceae	*Solanum melongena*	'*Ca*. P. australasiaticum subsp. australasiaticum' '*Ca*. P. australasiaticum subsp. ipomoeae'	2019 2019	BAWM-003 BAWM-004
	Solanaceae	*Solanum tuberosum*	'*Ca*. P. stylosanthis'	2019	VPRI 43683
	Vitaceae	*Vitis vinifera cv.* Chardonnay	'*Ca*. P. australiense'	2020	BAWM-189

Table 1. Continued

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
Research					
	Geraniaceae	*Pelargonium* sp.	'*Ca*. P. australasiaticum subsp. ipomoeae'	2021	BAWM-243
WA					
Carnavon					
	Convolvulaceae	*Convolvulus clementii*	'*Ca*. P. australasiaticum subsp. australasiaticum'	2020 2020	BAWM-176 BAWM-177
	Cucurbitaceae	*Citrullus lanatus*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2021	BAWM-340
	Cucurbitaceae	*Cucumis sativus*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020	BAWM-184
	Cucurbitaceae	*Luffa* sp.	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020	BAWM-174
	Fabaceae	*Vigna unguiculata* ssp. *sesquipedalis*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020	BAWM-179
	Solanaceae	*Capsicum annuum*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020 2020 2021	BAWM-173 BAWM-175 BAWM-331
	Solanaceae	*Solanum lycopersicum*	'*Ca*. P. australasiaticum subsp. ipomoeae' '*Ca. P*. australasiaticum subsp. australasiaticum'	2020 2020 2020 2020	BAWM-159 BAWM-172 BAWM-180 BAWM-181
	Solanaceae	*Solanum melongena*	'*Ca*. P. australasiaticum subsp. australasiaticum' '*Ca*. P. australasiaticum subsp. ipomoeae'	2020 2020 2021 2021	BAWM-182 BAWM-183 BAWM-332 BAWM-330
	Solanaceae	*Solanum nigrum*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020	BAWM-178
Dandaragan					
	Fabaceae	*Medicago sativa*	'*Ca*. P. fabacearum' Potentially new 16SrII species	2020 2020 2020	BAWM-165 BAWM-166 BAWM-167
	Fabaceae	*Bituminaria bituminosa*	'*Ca*. P. fabacearum'	2019 2019 2019 2019	BAWM-025 BAWM-026 BAWM-027 BAWM-028
Kununurra					
	Apocynaceae	*Catharanthus roseus*	Vigna Little Leaf phytoplasma Mixed infection ('*Ca*. P. stylosanthis' and '*Ca*. P. australasiaticum subsp. ipomoeae)	2021 2021	BAWM-245 BAWM-308
	Caricaceae	*Carica papaya*	'*Ca*. P. australasiaticum subsp. ipomoeae'	2019 2019	BAWM-041 BAWM-042
	Cucurbitaceae	*Cucurbita pepo var. giromontiina* (Zucchini)	'*Ca*. P. australasiaticum subsp. ipomoeae'	2020 2020	BAWM-187 BAWM-188a
	Cucurbitaceae	*Cucurbita maxima* (Pumpkin)	'*Ca*. P. australasiaticum subsp. ipomoeae' Mixed infection ('*Ca*. P. fabacearum' and '*Ca*. P. australasiaticum subsp. australasiaticum)	2020 2021	BAWM-333 BAWM-186

Table 1. Continued

Location	Host family	Host species/genus	Phytoplasma taxa identified by ANI or 16S rRNA BLASTn	Sampling year(s)	Sample name(s)
	Fabaceae	*Cicer arietinum*	'Ca. P. fabacearum'	2019	BAWM-043
	Fabaceae	*Crotalaria juncea*	'Ca. P. australasiaticum subsp. ipomoeae'	2019	BAWM-046
	Fabaceae	*Glycine max*	Mixed infection ('Ca. P. fabacearum' and 'Ca. P. australasiaticum subsp. australasiaticum)	2019	BAWM-044
	Fabaceae	*Vigna radiata*	'Ca. P. australasiaticum subsp. ipomoeae'	2019	BAWM-045
	Fabaceae	*Cajanus cajan*	'Ca. P. australasiaticum subsp. ipomoeae'	2020	BAWM-185
	Goodeniaceae	*Goodenia scaevolina*	Mixed infection ('Ca. P. stylosanthis' and 'Ca. Phytoplasma australasiaticum subsp. australasiaticum)	2021	BAWM-307
	Myrtaceae	*Melaleuca* sp.	'Ca. P. melaleucae'	2022	BAWM-354A
	Solanaceae	*Solanum lycopersicum*	'Ca. P. australasiaticum subsp. ipomoeae'	2003	BAWM-284
	Solanaceae	*Solanum* sp.	'Ca. P. australasiaticum subsp. australasiaticum'	2004	BAWM-285
Lancelin					
	Solanaceae	*Solanum nigrum*	'Ca. P. australasiaticum subsp. ipomoeae'	2021	BAWM-328
Serpentine					
	Malvaceae	*Gomphocarpus fruticosus*	'Ca. P. australasiaticum subsp. australasiaticum'	2021 2021 2021	BAWM-212 BAWM-213 BAWM-214

*An asterisk (*) in the sample name(s) column indicates samples for which the host identity was determined using DNA barcode analysis.

When the 12 16S rRNA sequences of the 'Ca. Phytoplasma australasiaticum' taxon 1 samples extracted from the genomic sequences were queried further, the historic samples were classified as 'TBB' (i.e. 'Ca. Phytoplasma australasiaticum') in the NT DITT phytoplasma database and their 16S rRNA genes shared the highest nucleotide sequence similarity and coverage with the 'Ca. Phytoplasma australasiaticum subsp. australasiaticum' representative strain PR08 (99.92% nucleotide sequence identity, 100% coverage) in BLASTn analyses (Table 1). Additionally, these sequences could only be differentiated from 'Ca. Phytoplasma australasiaticum subsp. australasiaticum' and 'Ca. Phytoplasma australasiaticum subsp. ipomoeae' during *in silico* RFLP analysis by the *Hae*III restriction enzyme (Fig. S1). Therefore, the identification and subsequent characterization of this 'Ca. Phytoplasma australasiaticum' subspecies was likely missed in previous analyses as the *Hae*III restriction enzyme was infrequently used during *in vitro* RFLP analyses [13, 14, 16, 59]. These results illustrate the low resolution of the RFLP of the 16S rRNA gene sequence to delimit separate phytoplasma taxa compared to species and subspecies characterization that is possible using the genome ANI, which has also been emphasized in previous studies [36].

The phytoplasma genomes obtained from samples BAWM-057, BAWM-306, and BAWM-319 formed their own clusters in the ANI analyses and do not share 100% ANI with any other sample. However, these three genomes all clustered with the 'Ca. Phytoplasma australasiaticum' species at <100% ANI (Fig. 1b) with approximately 80% AF (Fig. 1c). These ANI results suggest that each of these samples might represent three novel subspecies (referred to as 'Ca. Phytoplasma australasiaticum' taxon 2, 3, and 4, respectively). Both BAWM-057 and BAWM-319 were indistinguishable from 'Ca. Phytoplasma australasiaticum subsp. australasiaticum' based on nucleotide identities of their 16S rRNA gene in the BLASTn analyses (Table S1). The 16S rRNA gene sequence of BAWM-306 shared the highest nucleotide identity with 'Ca. Phytoplasma australasiaticum subsp. australasiaticum' (99.75% identity, 100% coverage, 4404 total score) (Table S1).

The phytoplasma genome obtained from sample BAWM-324 shares the lowest ANI and AF values with other with 'Ca. Phytoplasma australasiaticum' genomes analysed in this study (ca. 96% ANI and <80% AF) but is still grouped closely with species in the 'Ca. Phytoplasma australasiaticum' cluster. These results suggest that BAWM-324 might represent a single new 'Ca. Phytoplasma

Table 2. Summary of BLASTn top hits of the 16S rRNA gene sequences obtained for the samples investigated in this study, including the putative phytoplasma taxon identified, the number of samples with this result, and the range of percent of nucleotide identities shared with the top hit

'*Candidatus* Phytoplasma' top hits	No. samples	Range of nucleotide identities with the top hit (%)
'*Ca.* Phytoplasma australasiaticum subsp. australasiaticum'	91	98.88–100%
'*Ca.* Phytoplasma australasiaticum subspecies ipomoeae'	69	98.95–100%
'*Ca.* Phytoplasma fabacearum'	15	100%
'*Ca.* Phytoplasma bonamiae'	2	100%
'*Ca.* Phytoplasma planchoniae'	1	100%
'*Ca.* Phytoplasma stylosanthis'	6	100%
'*Ca.* Phytoplasma melaleucae'	2	100%
'*Ca.* Phytoplasma australiense'	1	99.90%
Vigna Little Leaf (ViLL) phytoplasma	3	99.83–99.92%
Waltheria Little Leaf (WaLL) phytoplasma*	1	NA
Mixed phytoplasma infection: '*Ca.* Phytoplasma australasiaticum subspecies ipomoeae' and '*Ca.* Phytoplasma stylosanthis'†	4	100% 100%

*The previous identification of BAWM-227 as a WaLL phytoplasma based on RFLP in the NT DITT phytoplasma database was used as no full-length 16S rRNA gene exists for this taxon [13]. However, the 16S rRNA gene of this phytoplasma shared 99.65% nucleotide identity with that of '*Ca.* Phytoplasma asiaticum'.
†Mixed infection was identified for these samples by Sanger sequencing cloned 16S rRNA PCR amplicons.

australasiaticum' subspecies (Fig. 1b, c; referred to as '*Ca.* Phytoplasma australasiaticum' taxon 5). Sequence divergence between the 16S rRNA gene sequences of BAWM-324 and its top hit, '*Ca.* Phytoplasma australasiaticum subsp. ipomoeae', was also observed (<100% nucleotide identity, 100% coverage, 2207 total score; Table S1), which supports these ANI results.

Together, these results suggest that the previously unrecorded taxa, '*Ca.* Phytoplasma australasiaticum' taxon 1 to taxon 5, are endemic to Australia. This is supported by '*Ca.* Phytoplasma australasiaticum' taxon 1 being present, but misclassified, in a historic sample collected in 2004 (Table S1) but also because these phytoplasma taxa have not been detected in any other country to date based on the 16S rRNA sequences and the limited number of 16SrII phytoplasma genomes that are publicly available. Using the genome-sequence data obtained for these strains in this study, further analyses are required and could be done to confirm whether these five new taxa are truly distinct subspecies of '*Ca.* Phytoplasma australasiaticum' and not artefacts generated during metagenomic sequencing and analyses [36, 37].

A putatively new species within the 16SrII phytoplasma group

The phytoplasma strains obtained from samples BAWM-167 and BAWM-339 shared 100% ANI and >80% AF with each other and approximately 94% ANI and <80% AF with any phytoplasma genomes used in this study, including the closely related subspecies of '*Ca.* Phytoplasma australasiaticum' (Fig. 1b, c). These results indicate that the phytoplasma strains from samples BAWM-167 and BAWM-339 may represent a novel '*Ca.* Phytoplasma' species within the 16SrII group. The 16S rRNA gene analyses support the divergence of these two strains compared to other previously described 16SrII phytoplasmas. The 16S rRNA genes of BAWM-167 and BAWM-339 had 99.92% sequence similarity to the reference sequence of '*Ca.* Phytoplasma australasiaticum subsp. australasiaticum' (100% coverage, 2180–2211 total score) (Table S1). Further phylogenomic analyses are required to confirm whether these three taxa are truly distinct subspecies and not due to artefacts generated during metagenomic sequencing and analyses. However, the identification of this putative novel species from two distinct hosts (*Medicago sativa* and *Ipomoea* sp.) across a large geographic separation may provide positive support for the existence of this species.

This species could be endemic to Australia because these two detections are the first time this species has been detected globally based on the 16S rRNA sequences and they were detected in two geographically distinct areas in Australia (Darwin, NT and Dandaragan, WA) in different hosts and years.

Other phytoplasma taxa identified in this study, which have previously been recorded in Australia

Strains of the 16SrII species, '*Ca.* Phytoplasma fabacearum', formed the third largest cluster (*n*=16) in the ANI analyses, while the cluster containing the group 16SrXXXVIII species '*Ca.* Phytoplasma stylosanthis' was the fifth largest (*n*=6) (Fig. 1). The single WaLL phytoplasma (group 16SrII) strain from BAWM-227, showed between 90 and 95% ANI with four 16SrII phytoplasma
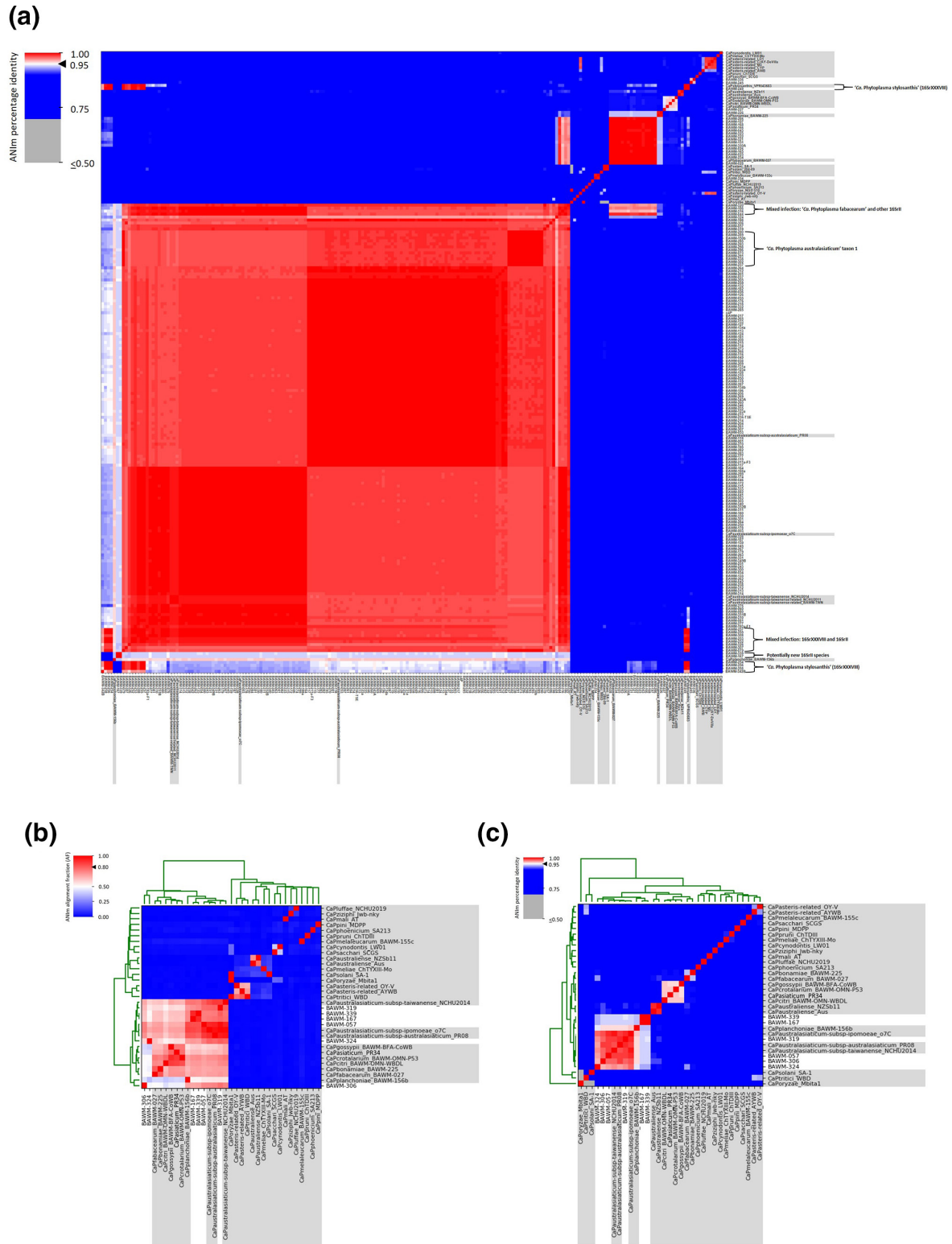
Fig. 1. Whole-genome comparisons for phytoplasma genome sequence data obtained for 178 samples. (a) ANI heatmap, generated by pyani version 0.2.10 using the ANIm algorithm, for all strains sequenced in this study alongside representative and publicly available genomes. Some clusters are highlighted using brackets. (b) ANI percentages and (c) alignment fractions (AF) in each pairwise comparison of samples that did not cluster with representative genomes in Fig. 1a. The genomes of representative strains and publicly available are shaded in grey. See colour gradient representing the percent identities in the heatmaps of (a) and (b) or the AF per genome in (c).

species previously determined to be closely related to each other [37], namely '*Ca*. Phytoplasma citri', '*Ca*. Phytoplasma asiaticum', '*Ca*. Phytoplasma gossypii', and '*Ca*. Phytoplasma crotalariae' (Fig. 1). The two '*Ca*. Phytoplasma bonamiae' strains (group 16SrII) formed their own cluster with 100% ANI between them in the pairwise analysis and shared the next highest ANI with strains of '*Ca*. Phytoplasma fabacearum' (ANI of <94%, Fig. 1). The two '*Ca*. Phytoplasma melaleucae' strains (group 16SrXV) formed their own cluster with 100% ANI between them in the pairwise analysis (Fig. 1). The ViLL phytoplasma strains (16 Sr group unassigned) clustered with each other, but with a pairwise ANI of ca. 97% to each other.

The BLASTn analyses of the 16S rRNA genes of all these taxa supported the ANI results (Fig. S1A), and the WaLL phytoplasma was characterized based on RFLP in previous analyses (NT DITT record; Table 2) [13]. Further, the ANI results of WaLL and ViLL phytoplasmas suggest that these taxa could be described as two novel '*Ca*. Phytoplasma' species (ANI <95% with any other genome available for described phytoplasma species). Future work is required to determine whether the WaLL and ViLL phytoplasmas meet the updated requirements for the description of novel '*Ca*. Phytoplasma' species [31] and whether the two ViLL strains represent two individual subspecies. Additionally, the competent insect vector species of WaLL and ViLL taxon remain to be determined.

The identification of mixed phytoplasma infections

Close analysis of the ANI heat map revealed evidence of mixed phytoplasma infections (Fig. 1), where several samples showed a high ANI with two representative genomes. Samples BAWM-044, BAWM-186, and BAWM-316 had mixed infections comprising of '*Ca*. Phytoplasma fabacearum' and '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae'. BLASTn analyses of the 16S rRNA genes obtained for these samples identified the presence of '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' but failed to identify '*Ca*. Phytoplasma fabacearum' (Table S1).

Sample BAWM-320 had a mixed infection of '*Ca*. Phytoplasma fabacearum' and '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' based on the ANI analyses (Fig. 1). However, only '*Ca*. Phytoplasma fabacearum' was identified from this sample based on BLASTn of the 16S rRNA gene obtained (Table S1).

Six samples had mixed infections of '*Ca*. Phytoplasma stylosanthis' and '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' based on the ANI analysis (samples BAWM-252, BAWM-253, BAWM-255, BAWM-257, BAWM-308, BAWM-326; Fig. 1). The 16S rRNA gene of '*Ca*. Phytoplasma stylosanthis' was obtained for all of these samples upon Sanger sequencing of the PCR amplicon apart from BAWM-308, which was identified as '*Ca*. Phytoplasma ipomoeae' when the PCR amplicon was Sanger sequenced (Table S1). Sanger sequencing of the cloned 16S rRNA gene PCR amplicon of samples BAWM-252, BAWM-253, BAWM-255, and BAWM-257 confirmed these mixed infections (data not shown).

'*Ca*. Phytoplasma stylosanthis' was identified in a mixed infection with either '*Ca*. Phytoplasma australasiaticum' taxon 1 (sample BAWM-079) or with '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' (sample BAWM-307). After PCR amplification and Sanger sequencing, only the 16SrII taxon was detected for these samples (Table S1).

When revisiting the number of tRNA gene sequences annotated from these 12 samples identified to contain mixed phytoplasma infections, more than 35 tRNA gene sequences were identified from all of these samples apart from sample BAWM-307 from which 32 tRNA genes were obtained (Table S1). Further, only one sample for which no mixed phytoplasma infection was identified using the ANI approach encoded more than 35 tRNA genes (BAWM-311, *n*=40 tRNA genes). These results highlight that the metagenomic sequencing, assembly, and tRNA annotation approach used in this study can sufficiently resolve the distinct tRNA genes encoded by each phytoplasma species and, thus, support the utility of using the tRNA count as an indicator of mixed phytoplasma infections in a sample in addition to genome completeness criteria that has been proposed previously [60, 61].

The results of species identification based on the 16S rRNA gene and whole-genome comparisons emphasize several important implications of these approaches to phytoplasma identification. Firstly, the 16S rRNA gene sequences obtained by either method often only represented one of the phytoplasma taxa involved in the mixed infection. It is likely that this arose due to differences in the titres of the multiple phytoplasma taxa in the sample, with only the one gene sequence being obtained by direct Sanger sequencing or more genomic sequence data obtained from the phytoplasma present at the higher titre. Alternatively, the multiple 16S rRNA gene sequences of closely related phytoplasma taxa obtained in a sample may have been missed during the process of obtaining the consensus sequence from both Sanger sequencing and metagenomic HTS data.

Phytoplasma identifications made for samples that could not be used in ANI analyses

For the samples for which insufficient phytoplasma data was obtained for ANI analyses, the 16S rRNA sequences shared high sequence similarity with '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' (*n*=9 sequences), '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' (*n*=7 sequences), '*Ca*. Phytoplasma australiense' (*n*=1 sequence, BAWM-189), and the ViLL phytoplasma (*n*=1, BAWM-337) (Table S1). These results highlight the pitfall of applying metagenomic-based approaches to identify the species diversity of phytoplasmas obtained from a diversity of host species. Specifically, it may be difficult to get sufficient data from hosts that harbour low titre infections for comparison with other taxa from metagenome sequencing. Whereas PCR of the 16S rRNA gene enriches for these regions of interest that provide taxonomically informative information [62], albeit at a low

Table 3. Summary of samples for which initial host identifications were unresolved to the family-, genus- or species-level based on visual inspections and for which additional gene regions obtained from metagenomic data were used to determine the host identity. The genes used for plant host identifications included the *maturase K* (*mat*K) and *ribulose-bisphosphate carboxylase* (*rbc*L) genes, and the *cytochrome C oxidase subunit 1* (*co*I) gene was used for insect identification. The e-value, percent identity, and bitscores of the top BLASTn hit(s) for the sample are provided to illustrate the support for the gene-based host identification. NA=Not applicable

Sample name	Host identification based on morphology	Host identification based on barcode analysis	*mat*K	*rbc*L	*co*I
			Top hit name (accession no.), e-value, nucleotide percent identity (%), bitscore per barcode		
BAWM-037	Unknown weed	Family: *Malvaceae* Genus: *Sida* Species: *Sida* sp.	*Sida rhombifolia* (MN006709); *Sida acuta* (MN006706) 1.98E-38, 100%, 22	*Sida cordifolia* (MH588540) 8.18E-47, 100%, 28	NA
BAWM-057	Unknown weed	Family: *Solanaceae* Genus: *Solanum* Species: *Solanum* sp.	*Solanum chenopodioides* (MH464378); *Solanum americanum* (MK244345) 1.72E-39, 100%, 32	*Solanum tripartitum* (KF546075); *Solanum palitans* (KF546070) 4.63E-75, 97.1%, 57	NA
BAWM-079	Unknown weed	Family: *Solanaceae* Genus: *Solanum* Species: *Solanum* sp.	*Solanum nigrum* (KC535797) 1.59E-35, 91.0%, 22	*Solanum rostratum* (NC_057245) 1.00E-69, 100%, 345	NA
BAWM-302	Unknown weed	Family: *Solanaceae* Genus: *Solanum* Species: *Solanum* sp.	*Solanum elaeagnifolium* (EU983576); *Solanum melanospermum* (EU983565) 0, 96.4%, 100	*Solanum rostratum* (MK526696) 0, 95.4%, 100	NA
BAWM-315	Family: *Fabaceae*	Family: *Fabaceae* Genus: *Crotalaria* Species: *Crotalaria* sp.	*Crotalaria juncea* (JQ619982.1) 3.6E-41, 98.90%, 163	*Crotalaria sagittalis* (KY584333); *Crotalaria spectabilis* (KJ773415); *Crotalaria pallida* (KJ773413) 7.6E-29, 100%, 13	NA
BAWM-316	Family: *Lamiaceae*	Family: *Asteraceae* Species: *Chromolaena odorata*	*Chromolaena odorata* (MN558588) 5.6E-57, 100%, 45	*Chromolaena odorata* (MH767490) 0, 91.9%, 58	NA
BAWM-319	Family: *Lamiaceae*	Family: *Malvaceae* Genus: *Sida* Species: *Sida* sp.	*Sida rhombifolia* (KT966997); *Sida fallax* (MF350256) 3.00E-46, 94.78%, 180	*Sida rhombifolia* (MH549993) 1.03E-50, 100%, 30	NA
BAWM-321	Family: *Lamiaceae*	Family: *Asteraceae* Genus: Undetermined Species: Undetermined	*Praxelis clematidea* (KX526581) 2.58E-176, 94.8%, 100 *Chromolaena odorata* (MN558588) 5.59E-173, 94.3%, 100	*Chromolaena odorata* (KY986097) 1.83E-62, 96.7%, 14	NA
BAWM-342B	Genus: *Orosius*	Species: *Orosius argentatus*	NA	NA	*Orosius argentatus* (KR030333) 0, 99.6%, 5
BAWM-343A	Genus: *Orosius*	Species: *Orosius argentatus*	NA	NA	*Orosius argentatus* (KR030333) 0, 99.9%, 5

taxonomic resolution [37]. This highlights the need to sequence additional genomic regions in diversity studies, and/or to have pre-sequencing enrichment tools for phytoplasma cells or DNA to improve genome sequence retrieval and assembly for genomic analysis relevant for applications such as taxonomy [38, 63–65]

Host and geographic ranges of the phytoplasma taxa identified in this study

Identification of unknown hosts
Eight weed samples were either unknown or not confidently identified to family, genus or species based on visual identification. The combination of two barcodes, *mat*K and *rbc*L, extracted from metagenomic data as well as occurrences of the identified species in the geographic region of collection, as recorded in the Australasian Virtual Herbarium, were used to indicate the host plant species (Table 3).

Two plants visually identified as members of the *Lamiaceae* were identified as *Asteraceae* members, including *Chromolaena odorata* (BAWM-316) and an undetermined species likely in the *Chromolaena* or *Praxelis* genera (BAWM-321). (Tables 3 and S1). One weed species (sample BAWM-319) was also visually identified as a member of the *Lamiaceae* but was identified as a member of the Malvaceae instead based on the DNA barcodes (*Sida* sp., BAWM-319) (Table S1). The host species of BAWM-319 is likely *Sida rhombifolia* as this species is present in QLD where the sample was collected, whereas *Sida fallax* reports were made from

WA in the Australasian Virtual Herbarium records. Species in the *Sida* genus have been reported as a host in Australia previously [12]. The results for these three samples are consistent with other phytoplasma detections made in Australia, where members in the *Lamiaceae* are not known to host phytoplasmas in Australia but where several *Asteraceae* and *Malvaceae* species have been recorded [27].

Samples BAWM-037, BAWM-057, BAWM-079, and BAWM-302 were all recorded as 'unknown weed' species based on morphological observations (Tables 3 and S1). Samples BAWM-057 and BAWM-302 were subsequently identified to the genus-level based on BLASTn of the *mat*K and *rbc*L gene sequences (both *Solanum* spp., Solanaceae). The host of BAWM-079 was also identified as a *Solanum* sp. (Solanaceae) by the host DNA barcode analyses (Table 3, either *Solanum nigrum* or *Solanum rostratum*). However, it is likely that the host species of BAWM-079 is *Solanum nigrum* as this plant species has a wide geographic distribution in Australia based on the Australasian Virtual Herbarium, including in the NT where BAWM-079 was sampled, and due to the high BLASTn results for this gene with *Solanum nigrum* (228 total score, 5E-60 e-value, 99.2% identity; GenBank accession number: M588530). *Solanum rostratum* is not present in the NT according to the Australasian Virtual Herbarium. However, the *rbc*L gene sequence had higher BLASTn scores with *Solanum rostratum* and it was, therefore, recorded in Table 3. Based on BLASTn of the *mat*K and *rbc*L genes obtained from the metagenomic data, the host of BAWM-037 was identified as a *Sida* sp. (Malvaceae). The host species of BAWM-037 could not be determined due to the inconsistencies between the top hit species listed for the two DNA barcodes (Table 3).

Two insect hosts were identified as *Orosius* sp. based on their external morphology (samples BAWM-342B and BAWM-343A). Using sequence analyses of their *coI* gene, both samples were identified as *Orosius argentatus*. Based on several studies, *Orosius argentatus* is a known phytoplasma vector in Australia and is detected across a broad geographic range in Australia [9, 29, 30, 66].

These sequence-supported identifications of plant or insect hosts at the species-, genus-, or family-level when they were not known based on visual inspection highlight the added benefit of a metagenomic-based approach to investigating phytoplasma diversity and their host associations. However, the host species listed using this approach are considered preliminary indications of the host taxa sampled, especially when (i) the nucleotide identities of the DNA barcodes were not identical to those of voucher specimens on the NCBI, despite the nucleotide identities being above 90% in all cases in this study (Table 3), (ii) recording species-level identifications, and (iii) considering that some barcodes may be missing for the species under investigation but for which they are available for a closely related species [67, 68]. This is due to the limitations of the available and well-validated plant DNA barcodes in the public databases.

Summary of phytoplasma plant and insect hosts for Australia

Over 40 different insect or plant genera were reported as phytoplasma hosts in this study, representing 16 plant families and one insect family (*Cicadellidae*) (Table 4) (Tables 1, 4 and S1). A total of 56 unique species were sampled. Of the 195 phytoplasma-infected samples, 158 were classed as crop species, 24 as weed species, seven as ornamental plants, six were native plants, and three were individual insect samples (Tables 4 and S1). Of the 158 crop species, ca. 59% were in the *Solanaceae* (n=91), ca. 17% were in the *Fabaceae* (n=34), ca. 8% were in the *Cucurbitaceae* (n=15), with the remaining samples from the *Apiaceae* (n=4), *Caricaceae* (n=4), *Convolvulaceae* (n=3), *Vitaceae* (n=2), and *Asteraceae* (n=1) families. The crop species *Solanum lycopersicum* (n=42), *Capsicum annuum* (n=33), and *Solanum melanogena* (n=16), and *Stylosanthes scabra* (n=10) were collected in high numbers (Tables 1 and S1).

Of the 24 phytoplasma-positive weed samples, 16 were unidentified to the genus- or species-level based on morphology or DNA barcode analysis (Tables 3 and 4). These 16 samples included species in the plant families *Asteraceae* (*Praxelis clematidea*, n=2; *Bidens pilosa*, n=1), *Convolvulaceae* (*Bonamia pannosa*, n=2; *Ipomoeae* sp., n=1), *Fabaceae* (*Crotalaria juncea*, n=2; *Crotalaria* sp., n=1), *Malvaceae* (*Gomphocarpus fruticosus*, n=3; *Waltheria* sp., n=1), and the *Solanaceae* (*Solanum nigrum*, n=2; *Solanum* sp., n=1). Ornamental hosts included species in the families *Apocynaceae* (*Catharanthus roseus*, n=2), *Asteraceae* (*Echinaceae purpurea*, n=1; *Osteospermum* sp., n=1), *Geraniaceae* (*Geranium* sp., n=1; *Pelargonium* sp., n=1), and *Solanaceae* (*Petunia* sp., n=1) (Table 4).

Native plants included those in the *Asteraceae* (*Peripleura diffusa*, n=1), *Goodeniaceae* (*Goodenia scaevolina*, n=1), *Lecythidaceae* (*Planchonia careya*, n=1), *Myrtaceae* (*Melaleuca* sp., n=2), and *Pedaliaceae* (*Phyllanthus fuernrohrii*, n=1) families (Table 4).

All three insects in this study from which phytoplasmas were detected were species of the *Cicadellidae* family, with one identified as *Orosius orientalis* (BAWM-232), and the other two identified as *Orosius argentatus* (BAWM-342B and BAWM-343A) (Table 4). '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum', '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae', and '*Ca*. Phytoplasma stylosanthis' were detected in *Orosius* species (family: Cicadellidae) analysed in this study (Tables 1–3 and S1). *Orosius* species, including *Orosius argentatus*, have been identified as confirmed or putative vectors of diseases thought to be associated with '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' and '*Ca*. Phytoplasma stylosanthis', such as tomato big bud, tobacco little leaf, and legume little leaf diseases [15, 27, 29]. The detection of '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' and '*Ca*. Phytoplasma stylosanthis' in *Orosius argentatus* in this study (Tables 1 and S1) is therefore consistent with the results of previous collection efforts in Australia [15] and it may be that this leafhopper species is a vector of several phytoplasma taxa. However, the detection of phytoplasma strains from the total nucleic extracts from insect whole bodies or

Table 4. List of phytoplasma hosts investigated in this study, characterized based on whether the host was recorded previously as a phytoplasma host/putative vector or not, and whether these hosts were classified as crop (C), insects (I), native plant (NP), ornamental (O), or weed (W) in this study. A list of the phytoplasma taxa that were identified are listed for the respective host

Host family	Host genus/species name	Host classification	Phytoplasma taxa detected
Previously recorded phytoplasma hosts for Australia collected in this study			
Apiaceae	*Apium graveolens*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'
Apocynaceae	*Catharanthus roseus*	O	*Vigna* Little Leaf phytoplasma; Mixed infection ('*Ca*. P. australasiaticum subsp. ipomoeae' and '*Ca*. P. stylosanthis')
Asteraceae sp.	*Asteraceae* sp.*	W	'*Ca*. P. fabacearum'
Caricaceae	*Carica papaya*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'; '*Ca*. P. stylosanthis'
Cicadellidae	*Orosius argentatus***	I	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. stylosanthis'
Cicadellidae	*Orosius orientalis*	I	'*Ca*. P. australasiaticum subsp. ipomoeae' (BLASTn)
Convolvulaceae	*Bonamia pannosa*	W	'*Ca*. P. bonamiae'
Convolvulaceae	*Ipomoea batatas*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'
Convolvulaceae	*Ipomoea* sp.	W	New 16SrII species
Cucurbitaceae	*Cucurbita maxima* (Pumpkin)	C	'*Ca*. P. australasiaticum subsp. ipomoeae'; Mixed infection ('*Ca*. P. australasiaticum subsp. ipomoeae' and '*Ca*. P. fabacearum')
Fabaceae	*Arachis hypogaea*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Fabaceae	*Cajanus cajan*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Fabaceae	*Cicer arietinum*	C	'*Ca*. P. fabacearum'
Fabaceae	*Crotalaria* sp.	W	'*Ca*. P. australasiaticum subsp. ipomoeae'
Fabaceae	*Crotalaria* sp.‡	W	'*Ca*. P. australasiaticum subsp. ipomoeae'
Fabaceae	*Glycine max*	C	'*Ca*. P. australasiaticum' taxon 1; '*Ca*. P. fabacearum'; Mixed infection ('*Ca*. P. australasiaticum subsp. ipomoeae' and '*Ca*. P. fabacearum')
Fabaceae	*Medicago sativa*	C	'*Ca*. P. fabacearum'; New 16SrII species
Fabaceae	*Stylosanthes scabra*	C	'*Ca*. P. stylosanthis'; '*Ca*. P. australasiaticum' taxon 5; Mixed infection ('*Ca*. P. australasiaticum subsp. ipomoeae' and '*Ca*. P. stylosanthis')
Fabaceae	*Vigna radiata*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Fabaceae	*Vigna unguiculata* ssp. *sesquipedalis*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'; '*Ca*. P. fabacearum'
Lecythidaceae	*Planchonia careya*	NP	'*Ca*. P. planchoniae'
Malvaceae	*Sida* sp.	W	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. australasiaticum' taxon 4
Malvaceae	*Waltheria* sp.	W	*Waltheria* Little Leaf phytoplasma
Myrtaceae	*Melaleuca* sp.	NP	'*Ca*. P. melaleucae'
Rosaceae	*Fragaria* ×*Ananassa Duch. Cv. Pajaro*	C	'*Ca*. P. australasiaticum' taxon 1
Solanaceae	*Capsicum annuum*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. australasiaticum subsp. ipomoeae'; '*Ca*. P. australasiaticum' taxon 1; '*Ca*. P. fabacearum'
Solanaceae	*Petunia* sp.	O	'*Ca*. P. australasiaticum subsp. australasiaticum'
Solanaceae	*Solanum lycopersicum*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. australasiaticum subsp. ipomoeae'; '*Ca*. P. fabacearum'

Continued

Table 4. Continued

Host family	Host genus/species name	Host classification	Phytoplasma taxa detected
Solanaceae	*Solanum melongena*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. australasiaticum subsp. ipomoeae'; '*Ca*. P. australasiaticum' taxon 1
Solanaceae	*Solanum nigrum*	W	'*Ca*. P. australasiaticum subsp. ipomoeae'
Solanaceae	*Solanum* sp.	W	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. australasiaticum' taxon 1; '*Ca*. P. australasiaticum' taxon 2; Mixed infection ('*Ca*. P. australasiaticum' taxon one and '*Ca*. P. stylosanthis')
Solanaceae	*Solanum tuberosum*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'; '*Ca*. P. stylosanthis'
Vitaceae	*Vitis vinifera cv.* Chardonnay/*Riesling*	C	'*Ca*. P. australiense' (16SrXII) (BLASTn); 16SrII Alfalfa phytoplasma (Sudan) (BLASTn)
New phytoplasma host records for Australia from this study			
Apiaceae	*Petroselinum crispum*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'
Asteraceae	*Bidens pilosa*	W	'*Ca*. P. australasiaticum subsp. ipomoeae'
Asteraceae†	*Chromolaena odorata*	W	Mixed infection ('*Ca*. P. australasiaticum subsp. ipomoeae' and '*Ca*. P. fabacearum')
Asteraceae	*Gynura crepioides* (Okinawa Spinach)	C	'*Ca*. P. australasiaticum' taxon 1
Asteraceae	*Osteospermum* sp.	O	'*Ca*. P. australasiaticum subsp. australasiaticum'
Asteraceae	*Peripleura diffusa*	NP	'*Ca*. P. australasiaticum subsp. ipomoeae'
Asteraceae	*Praxelis clematidea*	W	'*Ca*. P. australasiaticum subsp. ipomoeae'; Mixed infection ('*Ca*. P. australasiaticum subsp. australasiaticum' and '*Ca*. P. fabacearum')
Convolvulaceae	*Convolvulus clementii*	C	'*Ca*. P. australasiaticum subsp. australasiaticum'
Cucurbitaceae	*Citrullus lanatus*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Cucurbitaceae	*Cucumis sativus*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Cucurbitaceae	*Luffa acutangula*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Cucurbitaceae	*Luffa* sp.	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Cucurbitaceae	*Momordica charantia*	C	'*Ca*. P. australasiaticum subsp. ipomoeae'; ViLL
Cucurbitaceae	*Trichosanthes cucumerina*	C	'*Ca*. P. australasiaticum' taxon 3
Fabaceae	*Bituminaria bituminosa*	C	'*Ca*. P. fabacearum'
Fabaceae	*Crotalaria juncea*	W	'*Ca*. P. australasiaticum subsp. ipomoeae'
Geraniaceae	*Geranium* sp.	O	'*Ca*. P. australasiaticum subsp. australasiaticum'
Geraniaceae	*Pelargonium* sp.	O	'*Ca*. P. australasiaticum subsp. ipomoeae'
Asteraceae	*Echinacea purpurea*††	O	'*Ca*. P. australasiaticum subsp. australasiaticum'
Cucurbitaceae	*Cucurbita pepo var. giromontiina* (Zucchini)††	C	'*Ca*. P. australasiaticum subsp. ipomoeae'
Goodeniaceae	*Goodenia scaevolina*††	NP	Mixed infection ('*Ca*. P. australasiaticum subsp. australasiaticum' and '*Ca*. P. stylosanthis')
Malvaceae	*Gomphocarpus fruticosus*††	wW	'*Ca*. P. australasiaticum subsp. australasiaticum'
Pedaliaceae	*Phyllanthus fuernrohrii*††	NP	'*Ca*. P. australasiaticum subsp. ipomoeae'

*Indicates a sample for which the initial host family was unknown based on morphology but was subsequently identified to be a species within the *Asteraceae* using plant DNA barcodes.
†Indicates a sample where the host was initially identified as a *Lamiaceae* sp. but was subsequently identified as an *Asteraceae* species.
‡Indicates a sample for which the host was initially identified as a *Fabaceae* species but was subsequently identified to be a species within the genus *Crotalaria*.
§Indicates a sample where the host was initially identified to be a *Lamiaceae* species but for which the plant barcodes identified the host as a *Malvaceae* species (genus: *Sida*).
¶Indicates a sample for which the host species was unknown but subsequently identified as a species within the *Solanaceae* family.
**Indicates samples where an insect host was identified to the genus *Orosius* but which was subsequently identified as *Orosius argentatus* based on analysis of the COI gene sequence.
††Indicates a new host species for phytoplasmas in Australia but not a new host genus record.

subsampled body sections, as done in this study, does not provide definitive evidence of vector competence and further transmission trials are required. A comprehensive list of insect species that serve as competent vectors of phytoplasma diseases of vegetable crops in Australia remains to be determined.

Summary of new phytoplasma host records for Australia

Of the 56 unique species sampled in this study, 23 were recorded as new phytoplasma hosts for Australia (Tables 4 and S1). Eleven different plant genera were determined to be new phytoplasma hosts for Australia (Tables 4 and S1), and included the species *Bidens pilosa*, *Citrullus lanatus*, *Convolvulus clementii*, *Geranium* sp., *Pelargonium* sp., *Gynura crepioides* (Okinawa spinach), *Momordica charantia*, *Osteospermum* sp., *Peripleura diffusa*, *Praxelis clematidea*, and *Trichosanthes cucumerina*. Plant species identified as new phytoplasma host reports but for which the genus was already known to be a host in Australia (*n*=5), included *Crotalaria juncea* (previous detection reported as *Crotalaria goreensis* or *Crotalaria novae-hollandiae* [12]; or as *Crotalaria* sp. [8, 12]), *Echinacea purpurea* (previous detection in *Echinacea pallida* [69]), *Gomphocarpus fruticosus* (previous detection in *Gomphocarpus physocarpus* [70]), *Goodenia scaevolina* (previous detection reported as *Goodenia* sp. [13]), and *Phyllanthus fuernrohrii* (previous detection in *Phyllanthus amarus* [71]) (Tables 4 and S1).

Group 16SrII phytoplasma taxa

Species classified within the 16SrII group were detected in a broad range of host species and were the most frequently detected from the geographic regions and vegetable crops investigated in this study (Figs 1 and 2) (Tables 4 and S1). A total of 12 16SrII taxa were identified in the samples analysed and included: '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum', '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae', '*Ca*. Phytoplasma australasiaticum' taxon 1 to taxon 5, '*Ca*. Phytoplasma bonamiae', '*Ca*. Phytoplasma fabacearum', '*Ca*. Phytoplasma planchoniae', the WaLL phytoplasma, and the potentially new 16SrII species (Fig. 1).

'*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' and '*Ca*. Phytoplasma ustralasiaticum subsp. pomoeae' (16SrII)

'*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' and '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' were detected in all five states/territories investigated in this study (Table 1; Fig. 2). These two subspecies were also the most abundant taxa sampled in each state (Fig. 2). '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum', was detected in seven host families (Fig. 3a, b). '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' was identified in single infections in nine of the 34 new phytoplasma host species or genera identified in this study. New phytoplasma host records for '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' included the new phytoplasma hosts for Australia in the plant families *Apiaceae* (*Petroselinum crispum*), *Asteraceae* (*Echinaceae purpurea*, *Osteospermum* sp.), *Convolvulaceae* (*Convolvulus clementii*), *Geraniaceae* (*Geranium* sp.), and *Malvaceae* (*Gomphocarpus fruticosus*) (Table 4). This phytoplasma was associated with 11 symptom types, including little leaf, witches'-broom, yellowing, big bud, phyllody, and stunting (Fig. 3c and Table S1).

'*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' was detected from samples classified into seven host families (Fig. 3b) and was associated with symptoms of little leaves, witches'-broom, leaf discolouration (yellowing, purpling/reddening), big bud, phyllody, aborted flower/fruit development, fruit distortions and leaf curl/crinkle (Fig. 3c and Table S1). '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' was identified in single infections in 12 of the 34 new phytoplasma host species or genera identified in Australia (Table 4), including those in the *Asteraceae* (*Peripleura diffusa*, *Praxelis clematidea*), *Cucurbitaceae* (*Citrullus lanatus; Momordica charantia*, *Luffa acutangular*, a *Luffa* sp.), *Fabaceae* (*Crotalaria juncea*), *Geraniaceae* (*Pelargonium* sp.), and *Pedaliaceae* (*Phyllanthus fuernrohrii*).

The observation that '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' and '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' were the most frequently detected, the most widespread geographically, and detected from the most host families and species in this study is not surprising (Figs 1–3). These two species have historically been detected from a broad range of hosts in Australia, including many crop species [27]. '*Ca*. Phytoplasma australasiaticum subsp. australasiaticum' and '*Ca*. Phytoplasma australasiaticum subsp. ipomoeae' are major threats to crop production in Australia as they are commonly associated with crop species in many vegetable growing areas around the country, and disease incidences can be high [72]. Additionally, the symptoms they are associated with directly affect crop yield (e.g. phyllody). Further research is required to determine whether these two subspecies have distinct host ranges, symptomologies, or vector species for a better understanding of their biology and how to mitigate outbreaks of diseases associated with them.

'*Ca*. Phytoplasma australasiaticum' taxon 1 (16SrII)

'*Ca*. Phytoplasma australasiaticum' taxon 1 was detected in the NT (*n*=9), QLD (*n*=2), and VIC (*n*=1) (Fig. 2). This taxon was detected for hosts in the families *Solanaceae* (*Solanum melanogena, n*=7 samples; *Solanum* sp. weed, *n*=1 sample), *Asteraceae* (*Gynura crepioides*, *n*=1 sample), *Fabaceae* (*Glycine max*, *n*=1 sample), and *Rosaceae* (*Fragaria×Ananassa* Duch. Cv. Pajaro, *n*=1 sample) (Fig. 3b). The detection of '*Ca*. Phytoplasma australasiaticum' taxon 1 in *Gynura crepioides* (family: *Asteraceae*) represents a new phytoplasma host record for Australia (Table S1). The symptoms of the *Fragaria×Ananassa* Duch. Cv. Pajaro and *Solanum* sp. hosts were present upon sampling but were not recorded (Fig. 3c and Table S1). Six symptom types were recorded for the 10
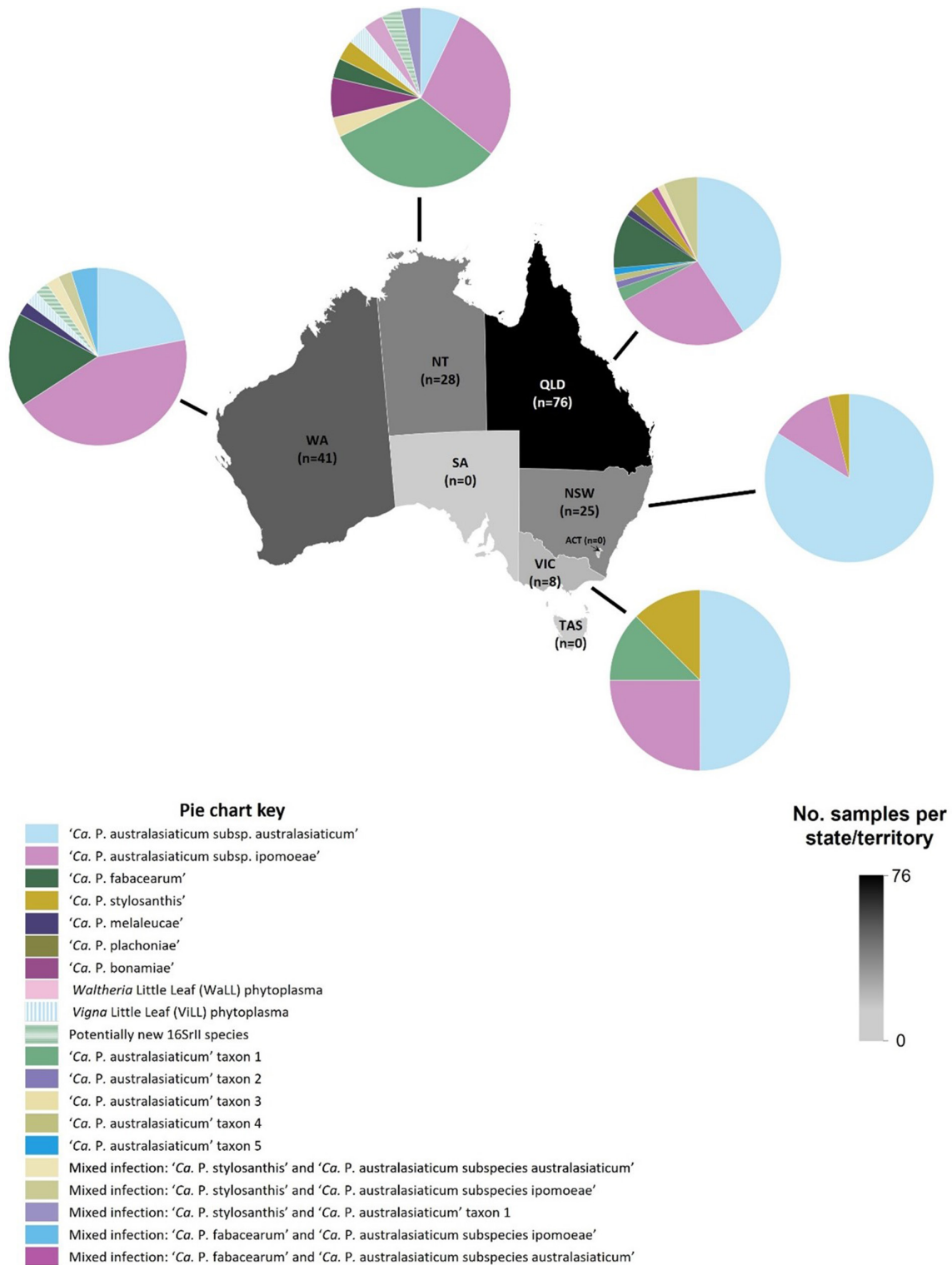
Fig. 2. Map of Australia showing the number of phytoplasma-positive samples collected per state or territory, with pie charts illustrating the proportions of ANI identified phytoplasma taxa identified per state or territory (see key below for descriptions of colour-coding). The scale on the right indicates the number of samples collected for each state or territory, with the number in brackets indicating the total number of ANI-identified samples per location within the map area. Abbreviations: ACT, Australian Capital Territory; NSW, New South Wales; NT, Northern Territory; QLD, Queensland; TAS, Tasmania; VIC, Victoria; WA, Western Australia.
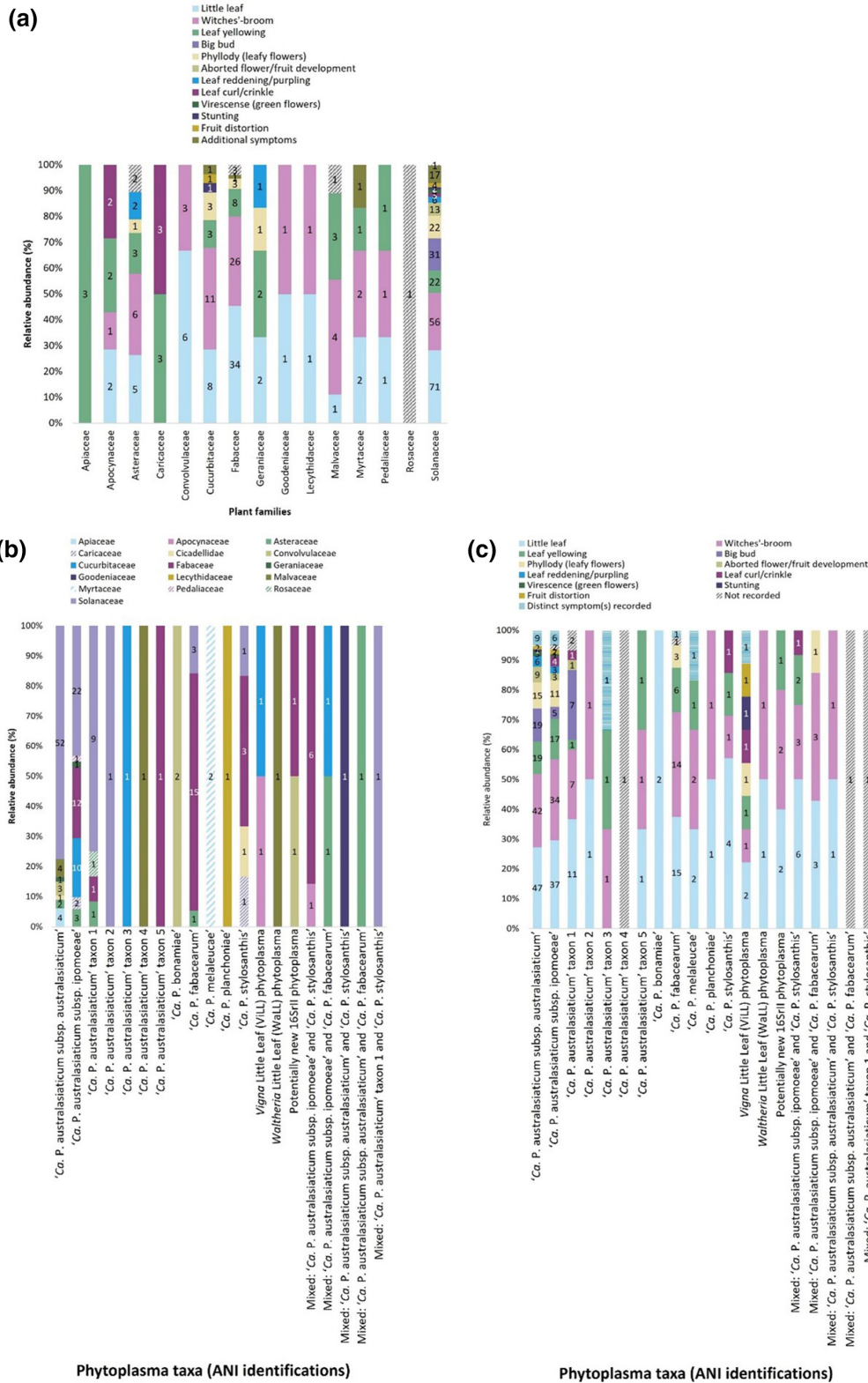
Fig. 3. Bar graphs indicating the relative abundances of (a) symptom types recorded for each plant host family analysed in this study (*n*=176 samples); (b) the ANI-identified phytoplasma taxa per plant or insect host family analysed in this study (*n*=178 samples); and (c) the symptom types recorded for each ANI-identified phytoplasma taxon analysed in this study (*n*=176 samples). Numbers in the bar graphs indicate the total number of samples. Colour legends are shown above each graph.

remaining hosts, including little leaf (*n*=10), big bud (*n*=7, all *Solanum melanogena* samples), witches'-broom (*n*=5), as well as leaf yellowing, aborted flower/fruit development and leaf curl/crinkle (*n*=1 each, *Capsicum annuum*).

This taxon may present a moderate threat to crop production in Australia, with the potential to affect Solanaceae hosts in particular. This is due to this taxon being detected from several crop hosts displaying symptoms that directly affect crop yield but also due to the large geographic range of the detections made in this study (from the NT, QLD, and VIC). Additionally, the detection of '*Ca*. Phytoplasma australasiaticum' taxon 1 was likely missed in previous RFLP-based analyses done to assess taxon diversity in Australia and may, therefore, have a broader host and geographic range than what is reported in this study. Further research is required to investigate the prevalence and vector(s) of this taxon.

'*Ca.* Phytoplasma australasiaticum' taxa 2 to 5 (16SrII)

'*Ca*. Phytoplasma australasiaticum' taxon 2 (*n*=1, a Solanaceae weed in the genus *Solanum,* BAWM-057), '*Ca*. Phytoplasma australasiaticum' taxon 4 (*n*=1, Malvaceae weed in the genus *Sida*, BAWM-319), and '*Ca*. Phytoplasma australasiaticum' taxon 5 (*n*=1, *Stylosanthes scabra*, Fabaceae, BAWM-324) were all detected in QLD (Figs 2 and 3b and Table S1). No symptoms were recorded for '*Ca*. Phytoplasma australasiaticum' taxon 4, whereas '*Ca*. Phytoplasma australasiaticum' taxon 2 and 5 both displayed witches'-broom symptoms (Fig. 3c and Table S1). Additional symptoms for these hosts included little leaf symptoms associated with the *Solanum* sp. infected with '*Ca*. Phytoplasma australasiaticum' taxon 2, as well as both little leaf and yellowing for the *Stylosanthes scabra* host infected with '*Ca*. Phytoplasma australasiaticum' taxon 5 (Fig. 3c and Table S1). '*Ca*. Phytoplasma australasiaticum' taxon 3 (BAWM-306) was detected in the NT from a *Trichosanthes cucumerina* (Cucurbitaceae) sample (Figs 2 and 3b; Table S1). The phytoplasma was associated with witches'-broom and yellowing, as well as 'shoe stringing' of the leaves (Fig. 3c and Table S1).

The ANI results indicated that these taxa were distinct and are likely novel subspecies of '*Ca*. Phytoplasma australasiaticum' (Fig. 1), representing the first time these taxa have been identified in Australia and likely globally. These novel detections are potentially due to these identifications being missed previously due to the high sequence similarity of their 16S rRNA gene to other described 16SrII phytoplasma taxa (>99% nucleotide sequence similarity; Table S1). Additionally, these novel detections may not have been made in the past as four of these novel phytoplasma taxa were detected from either weed species or from phytoplasma crop host species that are first records of phytoplasma hosts for Australia (Table 4).

The potential threat of these phytoplasma taxa remains to be determined as only a few detections in plant hosts were made in this study and their detections were likely missed in previous low resolution (RFLP-based) analyses (Table S1). Additionally, no vector species have been identified for these taxa. These results highlight the importance of sampling weed species in and around cropping areas, as well as collecting diverse species of symptomatic hosts in an area. These taxa need to be assessed further to determine whether they are truly distinct subspecies, which can be done using further comparative and phylogenomic assessments in future [36, 37].

'*Ca.* Phytoplasma fabacearum' (16SrII)

'*Ca*. Phytoplasma fabacearum' was detected in the NT (*n*=1), QLD (*n*=8), and WA (*n*=7) (Fig. 2). '*Ca*. Phytoplasma fabacearum' was detected in nine distinct plant species and was most frequently detected in *Fabaceae* species (*n*=15 samples, *Fabaceae* species: *Cajanus cajan*, *Cicer arietinum*, *Glycine max*, *Medicago sativa*, *Stylosanthes scabra*, *Bituminaria bituminosa*, and *Vigna unguiculata* ssp. *sesquipedalis*; Fig. 2b and Table S1). '*Ca*. Phytoplasma fabacearum' was also detected in three Solanaceae samples (Solanaceae species: *Capsicum annuum* and *Solanum lycopersicum*), and one Asteraceae sp. sample (Fig. 3b and Table S1). In these hosts, '*Ca*. Phytoplasma fabacearum' was associated with symptoms of little leaf, witches'-broom, yellowing, and phyllody, but also the distinct symptom of clearing of new leaf growth for one *Capsicum annuum* sample (Fig. 3c and Table S1).

These results highlight that *Fabaceae* crops across a broad geographic range in Australia are at a high risk of losses due to infection by '*Ca*. Phytoplasma fabacearum', although some *Asteraceae, Cucurbitaceae*, and *Solanaceae* hosts might also be at risk. This is also supported by reports of high incidence of phytoplasma diseases in Australia likely attributed, in part, to '*Ca*. Phytoplasma fabacearum' [72].

Historic samples: '*Ca.* Phytoplasma bonamiae' and the *Waltheria* Little leaf phytoplasma (16SrII)

Strains of '*Ca*. Phytoplasma bonamiae' (*n*=2) identified from *Bonnamia pannosa* and the WaLL phytoplasma (*n*=1) identified from a *Waltheria* sp. were only identified from the historically collected samples analysed in this study from QLD and the NT, respectively (Figs 2 and 3b, c; Table S1). No new host species or geographic range expansions are, therefore, reported for these taxa. '*Ca*. Phytoplasma bonamiae' was associated with little leaf symptoms in both samples, while the WaLL phytoplasma was associated with both little leaf and witches'-broom symptoms. No insect vector species have been identified for these phytoplasma taxa.

This study provides a full-length sequence of the 16S rRNA gene as well as genomic data for the WaLL phytoplasma for the first time. This sequence data identified this phytoplasma as a member of the 16SrII group (Figs 1 and S1; Table S1), which confirms previous reports based on nucleotide analysis of regions within the 16S rRNA gene [13]. Additionally, 16Sr rRNA and ANI sequence analysis showed that the '*Ca*. Phytoplasma bonamiae' and WaLL phytoplasmas were close relatives of the '*Ca*.

Phytoplasma fabacearum' strains (Figs 1 and 2) [13]. However, since both these phytoplasma taxa are infrequently detected in crop plants, and since 'Ca. Phytoplasma bonamiae' has only been detected from the Australian native plant *Bonamia pannosa* (based on 14 samples in the NT DITT database and [15, 27]), they are both unlikely to pose a major threat to crop production in Australia.

It is likely that the WaLL phytoplasma strain can be described as a novel species as it shares less than 96% ANI with other phytoplasma species (Fig. 1). Eight WaLL strains are in the NT DITT database and future work can investigate whether these WaLL phytoplasma strains further fulfil the updated guidelines to be described as a novel 'Ca. Phytoplasma' species should more sequence data be made available for them [31].

'Ca. Phytoplasma planchoniae' (16SrII)
'Ca. Phytoplasma planchoniae' was detected from a *Planchonia careya* host sampled in QLD (Figs 2 and 3b) that displayed little leaves and witches'-broom symptoms (Fig. 3c). 'Ca. Phytoplasma planchoniae' has previously been detected in Australia and has only been associated with native plant *Planchonia careya* in far north QLD [73]. Due to its narrow host range in a non-crop species and its restricted geographic range, 'Ca. Phytoplasma planchoniae' is unlikely to pose a major threat to crop production in Australia.

Potentially new 16SrII species
Strains of the potentially new 16SrII species were detected from an *Ipomoea* sp. (*Convolvulaceae*, sample BAWM-339) and from a *Medicago sativa* sample (*Fabaceae*, sample BAWM-167) from the NT and WA, respectively (Figs. 2 and 3b; Table S1). Both hosts showed symptoms of little leaf and witches'-broom, but the *Medicago sativa* sample BAWM-167 also showed symptoms of yellowing (Fig. 3c and Table S1). While further investigations are required to determine whether these two strains belong to a novel 'Candidatus Phytoplasma' species within the 16SrII group, support for their delimitation include the observation that (i) they produced ANI and AF values below the within-species threshold (<95% and <80%) with already described 'Candidatus Phytoplasma' species, and (ii) more than one strain of this potential species was identified from distinct hosts from different areas in Australia. It remains to be determined what threat to crop production in Australia this taxon presents as only these two strains were identified in this study with the competent or putative vector species unknown.

Group 16SrXXXVIII Phytoplasma samples – 'Ca. Phytoplasma stylosanthis'
'Ca. Phytoplasma stylosanthis' was detected in four of the five states/territories, including the NT ($n=1$), NSW ($n=1$), QLD ($n=3$), VIC ($n=1$), and WA ($n=1$) (Fig. 2). This phytoplasma was detected as single phytoplasma infection from four host families, namely the *Caricaceae* ($n=1$, *Carica papaya*), *Cicadellidae* ($n=1$, *Orosius argentatus*), *Fabaceae* ($n=3$, *Stylosanthes scabra*), and *Solanaceae* ($n=1$, *Solanum tuberosum*) (Fig. 3b and Table S1). Four of the five 'Ca. Phytoplasma stylosanthis'-infected plant hosts displayed little leaf symptoms, with the one sample, that did not show little leaf symptoms, showing yellowing and leaf curl/crinkle symptoms instead (BAWM-249, *Carica papaya*) (Fig. 3c). The *Solanum tuberosum* sample had the additional symptom of witches'-broom (Fig. 3c).

Since 'Ca. Phytoplasma stylosanthis' had previously only been reported in NT, QLD, and NSW [13, 18, 74], the *Solanum tuberosum* sample represents both a host and geographic range expansion for this phytoplasma (Table S1) as described previously (sample VPRI 43683 [24]). Since this phytoplasma has been identified from a broad range of crop species across a large geographic area, 'Ca. Phytoplasma stylosanthis' has the potential to be associated with reductions of economically important crops such as *Carica papaya* [15, 27].

Group 16SrXXV Phytoplasma samples – 'Ca. Phytoplasma melaleucae'
'Ca. Phytoplasma melaleucae' was detected in QLD ($n=1$; BAWM-155) and WA ($n=1$; BAWM-354) (Fig. 2) and were only detected as single infections from *Melaleuca* spp. (Myrtaceae) in Australian regions above the Tropic of Capricorn (Tables 1 and S1). This is the first report of 'Ca. Phytoplasma melaleucae' for WA and the furthest west occurrence of this phytoplasma. Prior to the present study, this phytoplasma was only reported from far north QLD and the Western Province of Papua New Guinea in *Melaleuca* spp., with one case reported for *Synsepalum dulcificum* (Sapotaceae) [16, 37]. The two samples analysed in this study displayed little leaf and witches'-broom symptoms (Fig. 3c and Table S1), which is consistent with previous detections [16]. At present, this phytoplasma is unlikely to pose a major threat, if any, to vegetable crop production in Australia due to its restricted host range to non-crop hosts.

Phytoplasma 16Sr group unassigned – vigna little leaf phytoplasma
The ViLL phytoplasma was detected in the NT and WA ($n=1$ per state/territory; Fig. 2). Sample BAWM-245 represents a host and geographic range expansion for the ViLL phytoplasma, being detected for the first time in WA and in a *Catharanthus roseus* sample (*Apocynaceae*; Table S1; [27]). A second host expansion for this phytoplasma and new phytoplasma host for Australia was *Momordica charantia* (*Cucurbitaceae*), detected in the NT (sample BAWM-336) where there was a high incidence of disease (70–80% of crop affected, in-field observation by S. Bond). Prior to this study, this phytoplasma taxon was only reported in Australia from within or near Katherine and Darwin in the NT [13, 15, 74]. Both samples showed little leaf symptoms, however

the *Catharanthus roseus* sample (BAWM-245) also showed symptoms of leaf yellowing and leaf curl/crinkle, while the *Momordica charantia* sample (BAWM-336) showed additional symptoms of witches'-broom, phyllody, stunting, and fruit distortions (Fig. 3c and Table S1).

ViLL phytoplasmas have been reported from symptomatic *Tridax procumbents* (family: *Asteraceae*), *Stylosanthes scabra* (family: *Fabaceae*), and *Vigna lanceolata* (family: *Fabaceae*) [13, 15, 74], with *Austroagallia torrida* (Evans) (family: *Cicadellidae*) and a *Batracomorphus* species (family: *Cicadellidae*) identified as possible vectors [15]. Although this phytoplasma taxon has been infrequently detected in Australia, the high incidence of disease observed for the crop *Momordica charantia* in the NT (in-field observation by S. Bond) and the broad host range suggests that this phytoplasma could be a threat to crop production in Australia, particularly in regions above the Tropic of Capricorn.

Group 16SrXII phytoplasma (16S rRNA PCR identification only)
A single 'Ca. Phytoplasma australiense' (16SrXII) detection was made in this study from a *Vitis vinifera cv*. Chardonnay sample (BAWM-189) collected in VIC (Table S1). This sample did not produce sufficient phytoplasma data to be used in the genome-based analyses despite the application of several approaches aimed at increasing the amount of phytoplasma DNA sequence data obtained from the background 'contaminating' DNA. These approaches included sampling different tissue types (e.g. whole leaves and petioles, vs petioles and leaf veins only), the application of the iodixanol-based phytoplasma enrichment method, as well as Illumina sequencing of these samples at a high depth (>22 Gb of data; Table S1). Although only one detection of this phytoplasma species was made in this study, this phytoplasma has previously been detected in a broad range of high-value hosts in Australia, including: *Carica papaya*, *Cucurbita maxima*, *Fragaria×Ananassa* Duch. Cv. Pajaro, and *Vitis vinifera* [12, 13, 17, 75–77] suggesting that this phytoplasma has the potential to impact crop production in Australia. While attempts have been made to identify the vector species of 'Ca. Phytoplasma australiense' in Australia [78], they remain to be determined.

Mixed phytoplasma infections
Mixed infections were identified from a total of 12 plant samples, which were collected in the NT (*n*=1), QLD (*n*=7), and WA (*n*=4) (Fig. 2). Mixed infections of 'Ca. Phytoplasma australasiaticum subsp. ipomoeae' and 'Ca. Phytoplasma stylosanthis' were recorded for five samples in QLD (all *Stylosanthes scabra* hosts collected in Mareeba and Atherton in the years 1998 and 2018; Tables 1 and S1; Fig. 2). One mixed infection of 'Ca. Phytoplasma australasiaticum subsp. ipomoeae' and 'Ca. Phytoplasma stylosanthis' was recorded for WA in *Catharanthus roseus*, representing both a host and geographic range expansion for 'Ca. Phytoplasma stylosanthis'. Mixed infections involving 'Ca. Phytoplasma fabacearum' and 'Ca. Phytoplasma australasiaticum subsp. ipomoeae' were also reported from QLD (*n*=1, *Chromolaena odorata*) and WA [*n*=2, *Cucurbita* sp. (Pumpkin) and *Glycine max*] (Tables 1 and S1; Fig. 2).

Three different mixed infection types were recorded for single plant samples collected either in the NT, QLD, or WA (Fig. 2 and Table S1). One of these included a single *Solanum* sp. sample with a mixed infection involving 'Ca. Phytoplasma australasiaticum' taxon 1 and 'Ca. Phytoplasma stylosanthis' collected in the NT (Fig. 2), which represents a new host species for 'Ca. Phytoplasma stylosanthis' (Fig. 1; Tables 3 and S1). The mixed infection involving 'Ca. Phytoplasma fabacearum' and 'Ca. Phytoplasma australasiaticum subsp. australasiaticum' was only detected in a single *Praxelis clematidea* sample collected in QLD (Fig. 2). The mixed infection involving 'Ca. Phytoplasma australasiaticum subsp. australasiaticum' and 'Ca. Phytoplasma stylosanthis' was detected in a single *Goodenia scaevolina* sample collected in WA (Fig. 2). *Goodenia scaevolina* represents a new phytoplasma host species record for Australia (Tables 4 and S1).

Although symptoms were present upon sampling, they were not recorded for the sample with the mixed infection of 'Ca. Phytoplasma australasiaticum' taxon 1 and 'Ca. Phytoplasma stylosanthis', nor for the mixed infection involving 'Ca. Phytoplasma australasiaticum ipomoeae' and 'Ca. Phytoplasma fabacearum' (Fig. 3c; Table S1). The remaining 10 mixed infection samples displayed little leaf symptoms (*n*=10), with witches'-broom symptoms as the next most prevalent symptom (*n*=7), followed by yellowing (*n*=2), phyllody (*n*=1) and leaf crinkle/curl (*n*=1).

Apart from the 16SrII phytoplasmas, 'Ca. Phytoplasma stylosanthis' was found most frequently in mixed infections with another phytoplasma. Similar instances of mixed infections involving 16SrII phytoplasmas and 'Ca. Phytoplasma stylosanthis' have been reported previously from both plant and insect hosts [14, 15, 74], including mixed infections of 'Ca. Phytoplasma australasiaticum subsp. ipomoeae' and 'Ca. Phytoplasma stylosanthis' [74]. These results demonstrate the large geographic and host range of 'Ca. Phytoplasma stylosanthis' in Australia and its overlapping geographic range with other phytoplasma taxa in the 16SrII group.

Summary of associations of phytoplasma and host species on symptoms presented in plants
Most of the plant samples displayed little leaf (*n*=138), proliferation/witches'-broom (*n*=120), and/or yellowing symptoms (*n*=57), followed by phyllody (*n*=37), big bud (*n*=31), aborted flower or fruit development (*n*=13), leaf curl or crinkle (*n*=13), leaf reddening or purpling (*n*=9), fruit distortions (*n*=6), virescence (*n*=4), and stunting (*n*=2) (Fig. 3a; Table S1). Additionally, little leaf, witches'-broom, and yellowing symptoms were observed within 13, 12, and 10 of the 14 plant families, respectively, for which

symptoms were recorded (Fig. 3a). The most symptom types (*n*=12) were recorded for the plants classified within the *Solanaceae*, followed by those in the *Asteraceae* and *Fabaceae* plant families (*n*=5 each). No symptoms were provided for 12 samples in this study (Fig. 3a; Table S1).

Distinct symptoms were observed and noted for 19 plant hosts (Fig. 3a; Table S1), including 'shoe stringing' of the leaves (*Cucurbitaceae* and *Solanaceae* samples BAWM-306 and BAWM-332, respectively), heart-shaped leaves (*Capsicum annuum* samples BAWM-211a-F3 and o7C), and gigantism of the calyx (*Solanum melongena* sample BAWM-003). Nine *Solanum lycopersicum* samples displayed distinct symptoms, including adventitious roots (*n*=5), wilt (*n*=3), and leaf tip necrosis (*n*=1). An interesting observation was the absence of phyllody or virescence for the ViLL phytoplasma-infected *Catharanthus roseus* sample BAWM-245.

No additional symptoms were observed for the mixed infection compared to samples where only a single phytoplasma taxon was observed (Fig. 3c; Table S1). While these associations need to be investigated more thoroughly as only a few host species overlapped between the single and mixed infection samples in this study, these observations have been reported previously [74]. The association of all plant symptoms with phytoplasma infection is, however, difficult to disentangle as phytoplasmas remain to be cultured and, as such, Koch's postulates cannot be fulfilled. Abiotic factors, herbicide treatments, insect damage, the presence of other microbes, viruses, or a combination of these factors could also contribute to the symptoms that are presented by the plant hosts [79, 80].

CONCLUSIONS

In this study, phytoplasma-infected crop and non-crop hosts from historic collections and contemporary collections (2015 to 2022) from vegetable growing regions around Australia were metagenomically sequenced to identify the crop-infecting phytoplasma taxa and potential alternative hosts. A total of 15 distinct phytoplasma taxa were identified from the metagenomic data obtained for these hosts (Figs 1 and 2). '*Ca*. Phytoplasma australasiaticum' subspecies and '*Ca*. Phytoplasma stylosanthis' were two of the most frequently detected taxa identified, and from the broadest range of hosts and locations sampled across Australia (Fig. 2). Additionally, six previously undescribed phytoplasma taxa were identified from the samples analysed in this study, namely: '*Ca*. Phytoplasma australasiaticum' taxon 1 to 5, and a potentially new 16SrII phytoplasma species. A few phytoplasma taxa were infrequently detected in this study, with some only associated with diseases in non-crop plants and, therefore, likely pose a low threat to crop production in Australia (e.g. '*Ca*. Phytoplasma melaleucae'). Five different phytoplasma mixed infections were also identified (Figs 1 and 2). An updated list of phytoplasma 16 Sr groups, species, subspecies, and unclassified taxa present in Australian vegetable growing areas, as well as the prevalence and combinations of mixed phytoplasma infections was therefore provided by this study. A list of symptoms per host (Fig. 3a) and per phytoplasma taxon (Fig. 3c) are also provided and, with the previous literature, will aid in-field detections of phytoplasma associated disease in crop and non-crop plant hosts (Table S1).

PCR of the 16S rRNA gene using universal nested phytoplasma primers combined with direct Sanger sequencing was sufficient as a triage tool to screen and provide a preliminary identification of the phytoplasma taxon present in every sample analysed in this study. However, it lacked the taxonomic resolution afforded by the ANI analysis of draft metagenomic-assembled phytoplasma genomes (Fig. S1), emphasizing results from other studies [36]. Additionally, the PCR-based approach often failed to accurately identify mixed infections (Table S1), which has been reported previously for Sanger sequencing of the PCR amplicon obtained directly from a sample [74]. The metagenomic-based approach employed in this study based on whole-genome ANI, however, was able to resolve strains to the subspecies-level and could identify the presence of a mixed phytoplasma infection in a single sample. An additional benefit of using the metagenomic approach during phytoplasma collection was that it allowed for host taxa to be identified through the use of genetic barcodes present in the metagenomic dataset when they were unable to be resolved to the family-, genus-, or species-level based on morphology (Table 3). Together, these results provided more informative data with a more precise assessment of the prevalence and host range of phytoplasmas in vegetable growing regions in Australia compared to previous studies, which could only use RFLP or sequence analysis of the 16S rRNA gene [12, 13, 15, 74]. The results presented in this study highlight the benefits of combining metadata (host, location, date, etc.) and metagenomic sequencing for phytoplasma diversity assessments and to understand their epidemiology.

Sufficient phytoplasma genomic data was obtained for 178 (12 mixed infections) of the 195 symptomatic samples for genome based sequence analyses and to be submitted to public sequence repositories (excluding mixed infection samples). The dataset presented here is the largest contribution of phytoplasma genome sequences from a single study to date, increasing the number of publicly available sequences from 47 [81] to a total of 213 (when excluding samples with mixed phytoplasma infections). The incomplete and draft phytoplasma genomes sequenced in this study have significantly increased the taxon sampling of subclade II, which is one of three subclades described in [82] (Fig. S1). The work presented here was possible due to the ever-decreasing cost of HTS and the increased volume of sequence data generated. The phytoplasma genome data obtained in this study can be used in future research to improve phytoplasma taxonomy and diagnostics, and will assist in genomic epidemiology analyses. The reliable genome sequence assemblies will also serve as a resource from which genes involved in symptomology and host/vector interactions can be investigated when combined with the appropriate metadata, as well as comparative and functional analyses.

Together, this genome resource will contribute significantly to the knowledge of phytoplasma biology, ecology, and can be used to inform management practises to help mitigate or prevent losses associated with major phytoplasma outbreaks in Australia.

Author contributions
Conceptualization: B.R.J., F.C., B.R., C.G., L.T.T.-N. Project administration: C.G., F.C. Funding acquisition: B.R., C.G., F.C. Laboratory work: B.R.J., W.M.K. Data curation, formal analysis, investigation, writing original draft: B.R.J. Provision of sample material: B.R.J., F.C., B.R., C.G., L.T.T.-N., M.S., C.W., M.K., R.D., L.J., S.B., N.P., W.M.K., T.C. Supervision: F.C., B.R., C.G., L.T.T.-N. Review and editing of the manuscript: F.C., B.R., C.G., L.T.T.-N., M.S., C.W., M.K., R.D., L.J., S.B., N.P., W.M.K., T.C.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
1. **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.

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. **Duduk B**, **Stepanović J**, **Yadav A**, **Rao GP**. *Phytoplasmas in Weeds and Wild Plants*. 2018.

4. **Kumari S**, **Nagendran K**, **Rai AB**, **Singh B**, **Rao GP**, *et al*. Global status of phytoplasma diseases in vegetable crops. *Front Microbiol* 2019;10:1–15.

5. **Hemmati C**, **Nikooei M**, **Al-Subhi AM**, **Al-Sadi AM**. History and current status of phytoplasma diseases in the middle east. *Biology* 2021;10:1–20.

6. **Cobb NA**. Tomato blights. *Agric Gaz New Wales* 1902;13:410–414.

7. **Helson GAH**. The transmission of witches' broom virus disease of lucerne by the common brown leafhopper, *Orosius argentatus* (Evans). *Aust J Sci Res B* 1951;4:115–124.

8. **Hutton EM**, **Grylls NE**. Legume "Little leaf", a virus disease of subtropical pasture species. *Aust J Agric Res* 1956;7:85.

9. **Bowyer J**, **Atherton J**, **Teakle D**, **A Ahern G**. Myooplasma-like bodies in plants affected by legume little leaf, tomato big bud, and lucerne witches' broom diseases. *Aust Jnl Of Bio Sci* 1969;22:271–274.

10. **Seemüller E**, **Schneider B**, **Mäurer R**, **Ahrens U**, **Daire X**, *et al*. Phylogenetic classification of phytopathogenic mollicutes by sequence analysis of 16S ribosomal DNA. *Int J Syst Bacteriol* 1994;44:440–446.

11. **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. *Phytopathology* 1996;85:169.

12. **Davis RI**, **Schneider B**, **Gibb KS**. Detection and differentiation of phytoplasmas in Australia. *Aust J Agric Res* 1997;48:535.

13. **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.

14. **Tran-Nguyen L**, **Blanche KR**, **Egan B**, **Gibb KS**. Diversity of phytoplasmas in northern Australian sugarcane and other grasses. *Plant Pathol* 2000;49:666–679.

15. **Padovan AC**, **Gibb KS**. Epidemiology of phytoplasma diseases in papaya in Northern Australia. *J Phytopathol* 2001;149:649–658.

16. **Davis RI**, **Jacobson SC**, **Rue SJDL**, **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.

17. **Constable FE**, **Jones J**, **Gibb KS**, **Chalmers YM**, **Symons RH**. The incidence, distribution and expression of Australian grapevine yellows, restricted growth and late season leaf curl diseases in selected Australian vineyards. *Ann App Biol* 2004;144:205–218.

18. **Gopurenko D**, **Fletcher MJ**, **Liu J**, **Gurr GM**. Expanding and exploring the diversity of phytoplasmas from lucerne (*Medicago sativa*). *Sci Rep* 2016;6:37746.

19. **Davis RE**, **Dally EL**, **Gundersen DE**, **Lee IM**, **Habili N**. "*Candidatus* Phytoplasma australiense," a new phytoplasma taxon associated with Australian grapevine yellows. *Int J Syst Bacteriol* 1997;47:262–269.

20. **White DT**, **Blackall LL**, **Scott PT**, **Walsh KB**. Phylogenetic positions of phytoplasmas associated with dieback, yellow crinkle and mosaic diseases of papaya, and their proposed inclusion in "*Candidatus* Phytoplasma australiense" and a new taxon, "*Candidatus* Phytoplasma australasia." *Int J Syst Bacteriol* 1998;48 Pt 3:941–951.

21. **ConstableF**, **WhitingJ**, **SymonsB**, **SymonsRH**. A new grapevine phytoplasma from the ovens valley of Victoria, Australia. *Ext Abstr* 2000;41:147–153.

22. **Marcone C**, **Gibb KS**, **Streten C**, **Schneider B**. "*Candidatus* Phytoplasma spartii", "*Candidatus* Phytoplasma rhamni" and "*Candidatus* Phytoplasma allocasuarinae", respectively associated with spartium witches'-broom, buckthorn witches'-broom and allocasuarina yellows diseases. *Int J Syst Evol Microbiol* 2004;54:1025–1029.

23. **Zhao Y**, **Davis RE**. Criteria for phytoplasma 16Sr group/subgroup delineation and the need of a platform for proper registration of new groups and subgroups. *Int J Syst Evol Microbiol* 2016;66:2121–2123.

24. **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 multi-locus sequence analysis of 16S rRNA, secA, tuf, and rp genes. *Int J Syst Evol Microbiol* 2020;71:1–11.

25. **Jones LM**, **Pease B**, **Perkins SL**, **Constable FE**, **Kinoti WM**, *et al*. '*Candidatus* Phytoplasma dypsidis', a novel taxon associated with

a lethal wilt disease of palms in Australia. *Int J Syst Evol Microbiol* 2021;71.

26. Gibb KS, Constable FE, Moran JR, Padovan AC. Phytoplasmas in Australian grapevines - detection, differentiation and associated diseases. *Vitis* 1999;38:107–114.

27. Streten C, Gibb KS. Phytoplasma diseases in sub-tropical and tropical Australia. *Austral Plant Pathol* 2006;35:129.

28. Dermastia M, Bertaccini A, Constable F, Mehle N. Worldwide distribution and identification of Grapevine yellows diseases. In: Dermastia M, Bertaccini A, Constable F and Mehle N (eds). *Grapevine Yellows Diseases and Their Phytoplasma Agents*. Switzerland: Springer; 2017. pp. 17–46.

29. Helson GHA. Distribution in Australia of three virus diseases and of their common vector *Orosius argentatus* (Evans). *J Aust Institue Agric Sci* 1949;15:160–161.

30. 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.

31. 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.

32. Bradbury JM. Minutes of the interim meeting, 12 and 18 July 1996, Orlando, Florida, U.S.A. *Int J Syst Evol Bacteriol* 1997;47:174–176.

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. Pilet F, Quaicoe RN, Osagie IJ, Freire M, Foissac X. Multilocus sequence analysis reveals three distinct populations of "*Candidatus* Phytoplasma palmicola" with a specific geographical distribution on the African continent. *Appl Environ Microbiol* 2019;85:1–16.

35. Orlovskis Z, Canale MC, Haryono M, Lopes JRS, Kuo C-H, *et al*. A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Ann Bot* 2017;119:869–884.

36. 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.

37. Rodrigues Jardim B, Tran-Nguyen LTT, Gambley C, Al-Sadi AM, Al-Subhi AM, *et al*. The observation of taxonomic boundaries for the 16SrII and 16SrXXV phytoplasmas using genome-based delimitation. *Int J Syst Evol Microbiol* 2023;73.

38. 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.

39. 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* 1999;83:482–485.

40. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 1991;173:697–703.

41. Constable FE, Joyce PA, Rodoni BC. A survey of key Australian pome fruit growing districts for exotic and endemic pathogens. *Austral Plant Pathol* 2007;36:165.

42. 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.

43. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.

44. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics* 2020;70:1–29.

45. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824–834.

46. Bushnell B. *Bbmap: A Fast, Accurate, Splice-Aware Aligner*. 2014.

47. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 2016;32:1088–1090.

48. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

49. 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.

50. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 2016;8:12–24.

51. Marquina D, Andersson AF, Ronquist F. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. *Mol Ecol Resour* 2019;19:90–104.

52. CBOL Plant Working Group. A DNA barcode for land plants. *Proc Natl Acad Sci U S A* 2009;106:12794–12797.

53. Kress WJ. Plant DNA barcodes: applications today and in the future. *J Syst Evol* 2017;55:291–307.

54. Guy PL, Cross PA, Wilson CR. A review of the plant virus and viroid records for Tasmania. *Australasian Plant Pathol* 2020;49:479–492.

55. 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.

56. Blanche KR, Tran-Nguyen LTT, Gibb KS. Detection, identification and significance of phytoplasmas in grasses in northern Australia. *Plant Pathology* 2003;52:505–512.

57. Andersen MT, Liefting LW, Havukkala I, Beever RE. Comparison of the complete genome sequence of two closely related isolates of "*Candidatus* Phytoplasma australiense" reveals genome plasticity. *BMC Genomics* 2013;14:1–15.

58. 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.

59. 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.

60. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, *et al*. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;35:725–731.

61. 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.

62. Gundersen DE, Lee I-M. Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathol Mediterr* 1996;35:144–151.

63. 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.

64. Nijo T, Iwabuchi N, Tokuda R, Suzuki T, Matsumoto O, *et al*. Enrichment of phytoplasma genome DNA through a methyl-CpG binding domain-mediated method for efficient genome sequencing. *J Gen Plant Pathol* 2021;87:154–163.

65. Tran-Nguyen LTT, Schneider B. Cesium chloride-Bisbenzimide gradients for separation of Phytoplasma and plant DNA. In: *Methods in Molecular Biology*. 2012. pp. 381–393.

66. Hill A, Mandryk M. A study of the virus diseases "Big bud" of tomato and "Yellow Dwarf" of tobacco. *Aust J Agric Res* 1954;5:617.

67. Xu C, Dong W, Shi S, Cheng T, Li C, *et al*. Accelerating plant DNA barcode reference library construction using herbarium specimens: improved experimental techniques. *Mol Ecol Resour* 2015;15:1366–1374.

68. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. *PLoS One* 2011;6:e19254.

69. **Pearce TL**, **Scott JB**, **Pethybridge SJ**. First report of a 16SrII-D subgroup phytoplasma associated with pale purple coneflower witches'-broom disease in Australia. *Plant Dis* 2011;95:773.

70. **Streten C**, **Herrington ME**, **Hutton DG**, **Persley DM**, **Waite GK**, *et al*. Plant hosts of the phytoplasmas and rickettsia-like-organisms associated with strawberry lethal yellows and green petal diseases. *Austral Plant Pathol* 2005;34:165.

71. **Tran-Nguyen LTT**, **Smith SH**, **Liberato JR**. Sweet potato little leaf strain V4 phytoplasma associated with snake bean in the Northern Territory, Australia. *Australasian Plant Dis Notes* 2012;7:147–150.

72. **Sharman M**, **Filardo F**, **Brier H**, **Gambley C**, **Steele V**, *et al*. Phytoplasma disease outbreaks in multiple crops in eastern Australia. In: *APPS Science Protecting Plant Health*. Brisbane, 2017. pp. 26–28.

73. **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.

74. **Rue SDL**, **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.

75. **Liefting LW**, **Padovan AC**, **Gibb KS**, **Beever RE**, **Andersen MT**, *et al*. Candidatus Phytoplasma australiense' is the phytoplasma associated with Australian grapevine yellows, papaya dieback and Phormium yellow leaf diseases. *Eur J Plant Pathol* 1998;104:619–623.

76. **Padovan A**, **Gibb K**, **Persley D**. Association of ' *Candidatus* Phytoplasma australiense' with green petal and lethal yellows diseases in strawberry. *Plant Pathol* 2000;49:362–369.

77. **Streten C**, **Conde B**, **Herrington M**, **Moulden J**, **Gibb K**. *Candidatus* Phytoplasma australiense is associated with pumpkin yellow leaf curl disease in Queensland, Western Australia and the Northern Territory. *Austral Plant Pathol* 2005;34:103–105.

78. **Osmelak JA**. Monitoring for potential leafhopper vectors (*Hemiptera*: *Cicadelloidea* and *Fulgoroidea*) of the causal agent of Australian grapevine Yellows. *Plant Prot Q* 1989;4:8–10.

79. **Harris MO**, **Pitzschke A**. Plants make galls to accommodate foreigners: some are friends, most are foes. *New Phytol* 2020;225:1852–1872.

80. **Abou-Khater L**, **Maalouf F**, **Patil SB**, **Balech R**, **Nacouzi D**, *et al*. Identification of tolerance to metribuzin and imazethapyr herbicides in faba bean. *Crop Sci* 2021;61:2593–2611.

81. **Wei W**, **Zhao Y**. Phytoplasma taxonomy: nomenclature, classification, and identification. *Biology* 2022;11:1119.

82. **Zhao Y**, **Davis RE**, **Wei W**, **Shao J**, **Jomantiene R**. Phytoplasma genomes: through mutually complementary mechanisms, gene loss and horizontal acquisition. In: **Gross DC** (eds). *Genomics of Plant-Associated Bacteria*. Berlin Heidelberg: Springer-Verlag; 2014. pp. 1–278.