Molecular Phylogenetics and Evolution

Exploring the Genomic Landscape of PWB Phytoplasmas: The Dominance of Selfcatalytic, High-copy Group II Introns in a complete genome of 'Candidatus Phytoplasma partheni' sp. nov.

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	Phytoplasma; Peanut Witches' broom 16SrII group; taxogenomics; group II intron; reverse transcriptase domain containing proteins; reductive genomes
Corresponding Author:	Amit Yadav, Ph. D. National Centre for Cell Science Pune, Maharshtra INDIA
First Author:	Kiran Kirdat, M. Sc.
Order of Authors:	Kiran Kirdat, M. Sc.
	Bhavesh Tiwarekar, M. Sc.
	Shivaji Sathe, Ph. D.
	Amit Yadav, Ph. D.
Abstract:	The Peanut Witches' Broom (PWB) or 16SrII group phytoplasmas are the most notorious bacterial pathogens associated with various diseases in legumes, horticultural crops, and weed species worldwide. Two PWB isolates PR08 and PR34, associated with phyllody and witches' broom disease in common Indian weed Parthenium hysterophorus, were sequenced to understand their genome structure. The single contig circular genomes of PR08 (588,746 bp) and PR34 (614,574 bp) obtained through hybrid assembly revealed the plethora of pathogenesis-related proteins. These phytoplasmas were discovered to have high copy numbered self-catalytic group II introns occupying more than 4% of the total genome size. The G+C content of these introns deviates from the rest of the genome, suggesting that Indian isolates of PWB phytoplasmas have acquired these genes through horizontal transfer. No typical PMU regions were detected in these genomes, although a few scattered PMU-related genes were found. The genomes contained one complete homolog of sodA gene and multiple homologs of truncated hlyB gene, whose mode of action is unknown without accessory genes. The Genome analysis of isolate PR34 showed its unique phylogenetic position separate from 'Ca. P. aurantifolia' and 'Ca. P. australasia', indicating distinct evolution, with orthoANI and dDDH values of 95.45 and 62.1% compared to 'Ca. P. aurantifolia' isolate WBDL, respectively. The phylogenetic analysis, including 16S rRNA gene (>1500 bp), five marker genes deduced based on high densities of informative sites (>1100 aa), and core genome phylogeny (>10,000 aa) shows that isolate PR34 forms a distinct clade among PWB group of phytoplasmas with strong bootstrap support. Based on obtained OGRI values, we propose 'Candidatus Phytoplasma partheni' sp. nov. with reference isolate PR34 and genome sequence CP097206.
Suggested Reviewers:	Chih-Horng Kuo, Ph. D. Asst Director, Academia Sinica chk@gate.sinica.edu.tw Chair, IRPCM-PAM Expert, Phytoplasma Genomics and Evolotion
	Fiona Constable, Ph. D. Professor, Victorian Department of Environment and Primary Industries Fiona.Constable@agriculture.vic.gov.au Expert, Phytoplasma Genomics and Taxonomy
	Suman Lakhanpaul, Ph. D. Professor, University of Delhi sumanlp2001@yahoo.com Phytoplasma biologist

	Isil Tulum, Ph. D. Asst Professor, University of Istambul isil.tulum@istanbul.edu.tr Phytoplasma biologist
	Xavier Foissac, Ph. D. Ex Director, French National Institute for Agricultural Research INRAE xavier.foissac@inrae.fr Senior Phytoplasma biologist
	Kamiya Shigeru, Ph. D. Kyorin University Faculty of Medicine School of Medicine skamiya@ks.kyorin-u.ac.jp Senior Phytoplasma biologist
	Kenro Oshima, Ph. D. Professor, Hosei University kenro@hosei.ac.jp Senior Phytoplasma biologist
Opposed Reviewers:	

June 25, 2023

To,
Editor-in-Chief,
Molecular Phylogeny and Evolution

Subject: Request for Consideration of Manuscript for Publication.

On the behalf all authors, I am submitting a research paper titled ‘Exploring the Genomic Landscape of PWB Phytoplasmas: The Dominance of Self-catalytic, High-copy Group II Introns in a complete genome of 'Candidatus Phytoplasma partheni' sp. nov.’ for consideration of publication in Molecular Phylogeny and Evolution Journal. This manuscript provides a comprehensive analysis of the genome structures and evolutionary characteristics of two isolate of Peanut Witches' Broom (PWB) phytoplasmas.

PWB phytoplasmas are highly impactful bacterial pathogens affecting legumes, crops, and weed species worldwide. Our study highlights their unique characteristics, including abundant pathogenesis-related proteins and a high number of self-catalytic group II introns, constituting a significant portion of the genome. We also discovered a complete homolog of the sodA gene and multiple truncated hlyB gene homologs, with unclear functions in the absence of accessory genes. Furthermore, through genome analysis of PR34, we identified its distinct phylogenetic position as a separate species, 'Candidatus Phytoplasma partheni,' separate from other related phytoplasmas. Phylogenetic analyses using multiple markers and core genome phylogeny supported the formation of a distinct clade by PR34 within the PWB phytoplasma group.

In conclusion, this research enhances our understanding of PWB phytoplasma genomes, their evolution, and their interaction with the common Indian weed host *Parthenium hysterophorus*. The comprehensive analysis of genomic features, effector proteins, and the absence of Potential Mobile Units (PMUs) contributes to our knowledge of phytoplasma-host plant interactions and evolutionary dynamics.

Considering the scope and objectives of Molecular Phylogeny and Evolution Journal, we kindly request you to consider our manuscript for publication. We firmly believe that our research will provide valuable insights into the field of molecular phylogeny and evolution.

Thank you for your time and consideration. We eagerly await your response.

Sincerely,

(Amit Yadav, Ph. D.)
Corresponding author

Highlights:

Exploring the Genomic Landscape of PWB Phytoplasmas: The Dominance of Self-catalytic, High-copy Group II Introns in a complete genome of *'Candidatus* Phytoplasma partheni' sp. nov.

Kiran Kirdat[1,2], Bhavesh Tiwarekar[1], Shivaji Sathe[2], and Amit Yadav[1, *]

[1]National Centre for Cell Science, Ganeshkhind, Pune 411 007, India; [2] Department of Microbiology, Tuljaram Chaturchand College, Baramati, Maharashtra 413102, India.

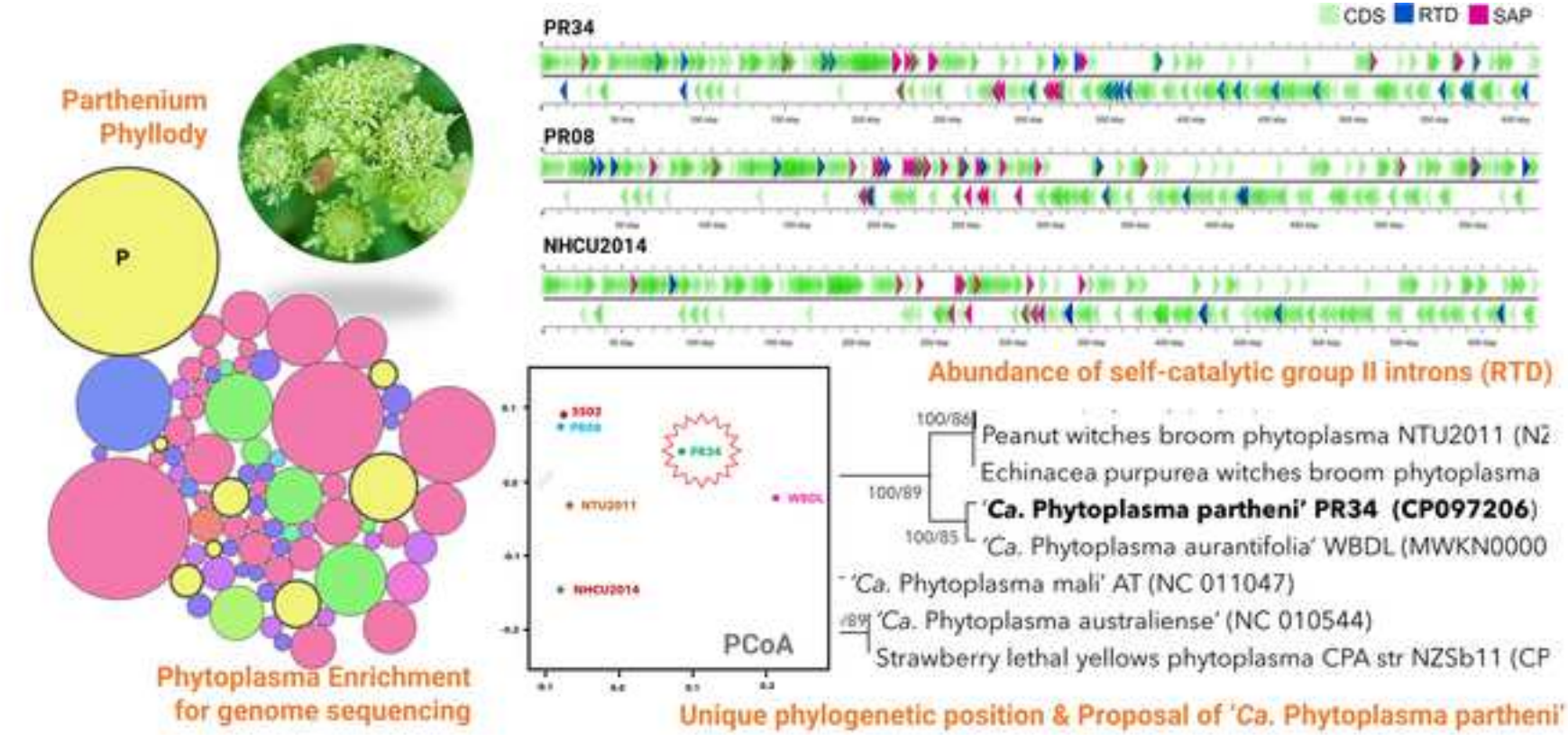
***Corresponding author**: Amit Yadav; E-mail: *amityadav@nccs.res.in*

Keywords: Phytoplasma, Peanut Witches' broom, 16SrII group, OGRI, taxogenomics, group II intron, reverse transcriptase domain containing proteins, reductive genomes.

Highlights

- PWB (16SrII) phytoplasmas are significant pathogens in legumes, crops, and weeds
- Unique sequencing strategies of PWB isolates unveiled circular genomes
- PWB genomes displayed abundant PR proteins & self-catalytic group II introns
- truncated multiple homologs of *hlyB* & a SOD gene were found with unknown function
- PR34's phylogenetic position suggests separate evolutionary lineage and speciation

Graphical Abstract



Exploring the Genomic Landscape of PWB Phytoplasmas: The Dominance of Self-catalytic, High-copy Group II Introns in a complete genome of *'Candidatus* Phytoplasma partheni' sp. nov.

Kiran Kirdat[1,2], Bhavesh Tiwarekar[1], Shivaji Sathe[2], and Amit Yadav[1,*]

[1]National Centre for Cell Science, Ganeshkhind, Pune 411 007, India; [2] Department of Microbiology, Tuljaram Chaturchand College, Baramati, Maharashtra 413102, India.

***Corresponding author**: Amit Yadav; E-mail: *amityadav@nccs.res.in*

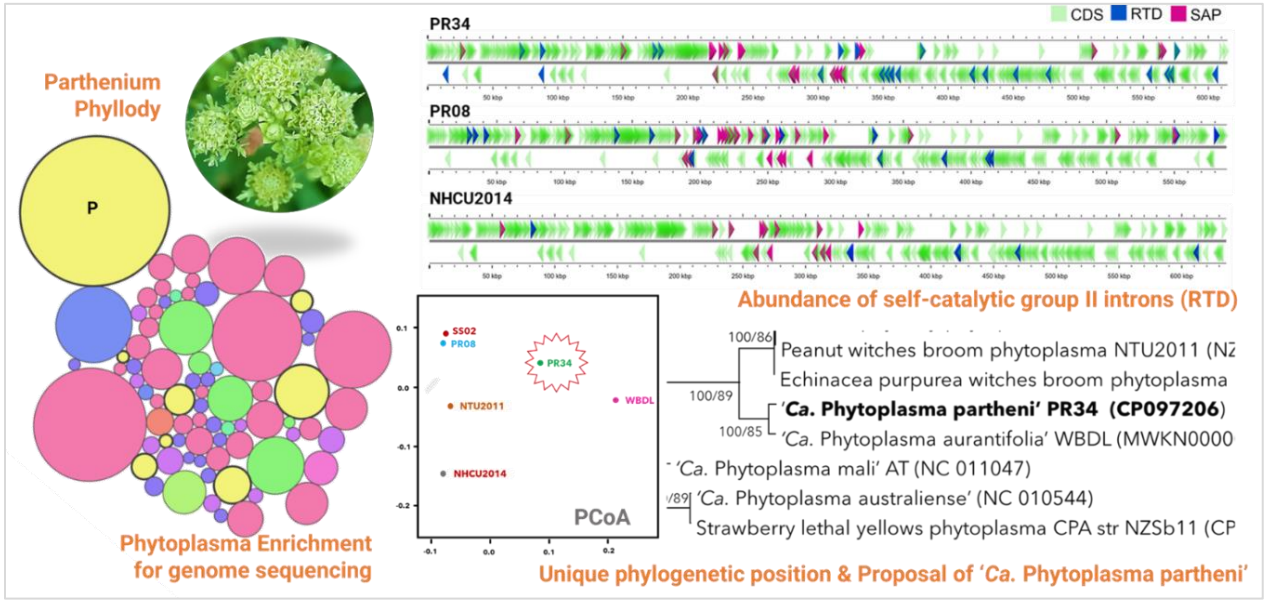
Keywords: Phytoplasma, Peanut Witches' broom, 16SrII group, OGRI, taxogenomics, group II intron, reverse transcriptase domain containing proteins, reductive genomes.

Abbreviations: PWB, Peanut Witches' Broom; *Ca., Candidatus*; aa, amino acid; bp, base pair; PMU, Potential Mobile Units; ANI, Average Nucleotide Identity; OGRI, Overall Genome Relatedness Index; dDDH, Digital DNA-DNA Hybridization; BPGA, Bacterial Pan Genome Analysis; UBCG, Up-to-date Bacterial Core Gene; CDS, coding sequence; MEGA, Molecular Evolutionary Genetics Analysis; PR proteins, pathogenesis related proteins; IEP, Intron encoded protein; RT, Reverse transcriptase; RTD, Reverse Transcriptase Domain containing protein; *hylB*, Alpha-hemolysin translocation ATP-binding protein.

Highlights

- PWB (16SrII) phytoplasmas are significant pathogens in legumes, crops, and weeds
- Unique sequencing strategies of PWB isolates unveiled circular genomes
- PWB genomes displayed abundant PR proteins & self-catalytic group II introns
- truncated multiple homologs of *hlyB* & a SOD gene were found with unknown function
- PR34's phylogenetic position suggests separate evolutionary lineage and speciation

Graphical Abstract



Abstract

The Peanut Witches' Broom (PWB) or 16SrII group phytoplasmas are the most notorious bacterial pathogens associated with various diseases in legumes, horticultural crops, and weed species worldwide. Two PWB isolates PR08 and PR34, associated with phyllody and witches' broom disease in common Indian weed *Parthenium hysterophorus,* were sequenced to understand their genome structure. The single contig circular genomes of PR08 (588,746 bp) and PR34 (614,574 bp) obtained through hybrid assembly revealed the plethora of pathogenesis-related proteins. These phytoplasmas were discovered to have high copy numbered self-catalytic group II introns occupying more than 4% of the total genome size. The G+C content of these introns deviates from the rest of the genome, suggesting that Indian isolates of PWB phytoplasmas have acquired these genes through horizontal transfer. No typical PMU regions were detected in these genomes, although a few scattered PMU-related genes were found. The genomes contained one complete homolog of *sodA* gene and multiple homologs of truncated *hlyB* gene, whose mode of action is unknown without accessory genes. The Genome analysis of isolate PR34 showed its unique phylogenetic position separate from *'Ca*. P. aurantifolia' and '*Ca*. P. australasia', indicating distinct evolution, with orthoANI and dDDH values of 95.45 and 62.1% compared to *'Ca*. P. aurantifolia' isolate WBDL, respectively. The phylogenetic analysis, including 16S rRNA gene (>1500 bp), five marker genes deduced based on high densities of informative sites (>1100 aa), and core genome phylogeny (>10,000 aa) shows that isolate PR34 forms a distinct clade among PWB group of phytoplasmas with strong bootstrap support. Based on obtained OGRI values, we propose '*Candidatus* Phytoplasma partheni' sp. nov. with reference isolate PR34 and genome sequence CP097206.

1. Introduction

Phytoplasmas (Phylum, Mycoplasmatota; class, Mollicutes; Order: Acholeplasmatales; family: *Acholeplasmataceae*; genus, *'Candidatus* Phytoplasma') are bacterial pathogens that lack cell walls and reside in the phloem of plants. They are obligate endophytes and have been associated with diseases in numerous plant species worldwide. These phytoplasma-related diseases pose a significant threat to global agriculture and horticulture, leading to substantial yield losses (Lee et al., 2000; Rao et al., 2017; Tiwari et al., 2023a; Tiwari et al., 2023b). Phytoplasma infected plants exhibit various symptoms, including reduced leaf size, witches' broom, phyllody, virescence, and sterility, affecting a wide range of plant species (Tully, 1993; Gundersen et al., 1994). Among the major groups of phytoplasmas, the Peanut Witches' Broom (PWB) group is particularly significant, causing diseases in pulse crops and weeds worldwide (Rao et al., 2017; Duduk et al., 2018; Mall et al., 2023). The PWB or 16SrII

group comprises two phytoplasma species, namely *'Ca*. P. aurantifolia' and *'Ca*. P. australasia' (White et al., 1998; Zreik et al., 1995).

India, the largest global producer of pulses with 23.3 million tons in 2020, is also the 10th largest exporter of dried legumes, worth $276M in 2020 (Observatory of Economic Complexity World). PWB phytoplasma infestation causes phyllody symptoms and 'no pod disease' in legume crops, leading to significant yield losses (Tiwari et al., 2023b; Mall et al., 2023). These phytoplasmas also infect alfalfa, Napier hybrid grass, bell pepper, pearl millet, cowpea, pigeon pea, lentil, soybean, and mung bean (Thorat et al., 2016a; 2016b; Thorat et al., 2017; Rao et al., 2018).

Weeds act as important secondary hosts for PWB phytoplasmas, aiding their transmission to crop plants through insect vectors (Thorat et al., 2016a; Yadav et al., 2015a; Thorat et al., 2017). Common invasive weeds like *Cleome viscosa, Trichodesma zeylanicum*, Ban Tulsi (*Croton bonplandianus*), and *Tephrosia purpurea* have been found infected with PWB phytoplasmas, which have also been detected in crops such as sesame, soybean, cowpea, and French bean (Yadav et al., 2015a, 2014; Thorat et al., 2016a, 2017; Kirdat et al., 2020b). Notably, the phyllody and witches' broom disease of the notorious weed *Parthenium hysterophorus* is associated with PWB phytoplasmas in India (Yadav et al., 2015a; Thorat et al., 2016a).

This study aimed to sequence the genomes of PWB phytoplasma isolates PR08 and PR34, associated with phyllody and witches' broom disease in *Parthenium hysterophorus*, to investigate the putative effectors and other pathogenesis-related genomic features. The investigation considered the broad host range, geographic region, and unique genomic characteristics of these phytoplasmas, highlighting the importance of understanding their multi-host pathogenicity and evolutionary patterns.

Whole-genome sequencing provides valuable insights into pathogen virulence factors and enables comparative taxogenomics using Overall Genome Relatedness Index (OGRI) data. However, the unculturable nature of phytoplasmas poses challenges in obtaining sufficient genomic DNA for sequencing, and the repetitive nature of their genome hampers full-length assembly with short-read data. To date, only 20 complete genomes have been reported out of 65 Phytoplasma assemblies (Bertaccini et al., 2022; Wei and Zhao, 2022; Kirdat et al., 2023). Previous enrichment methods included cesium chloride density gradient centrifugation and pulse-field gel electrophoresis (PFGE) (Bai et al., 2006; Hogenhout and Segura, 2009; Kube et al., 2012; Marcone, 2014), while current methods involved host DNA removal and immunoprecipitation-based enrichment (Kirdat et al., 2021, 2020a; Tan et al.,

2021). This study focused on enrichment attempts of PWB phytoplasma genomic DNA and the employed sequencing strategies for obtaining single scaffold genome sequences.

The study uncovers the abundance of mobile group II introns in PWB phytoplasma isolates, revealing unique genome features. The presence, copy number, and sequence variation of group II introns indicate potential horizontal transfer events. Additionally, the distinct taxonomic position of isolate PR34 is confirmed through genome-derived PCoA coordinates and OGRI values. This leads to the proposal of a new taxon, *'Candidatus* Phytoplasma partheni,' associated with phyllody disease in *Parthenium hysterophorus*.

2. Materials and Methods

2.1. Sample Collection

The *Parthenium hysterophorus* plants surrounding the soybean and other agriculture fields exhibiting typical symptoms of phyllody and witches' broom were collected from multiple places in the western Maharashtra state of India. The sample collection was carried out from 2014 to 2020 in multiple surveys of legume crops for the incidence of phytoplasma-related diseases. All samples were collected in multiple numbers, cleaned when fresh, and stored at -80 °C until further use.

2.2. Phytoplasma Identification and Quantification

The genomic DNA was isolated from 100 mg of infected leaf tissue using the CTAB method (Doyle and Doyle, 1987). The plant genomic DNA was screened for phytoplasma DNA using phytoplasma 16S rRNA gene-specific primers P1 (Deng and Hiruki, 1991) and P7 (Schneider, 1995), followed by a nested PCR using primers R16.100F and R16.1386R (Kirdat et al., 2022). The 1.28 kb PCR amplicons were purified by the PEG-NaCl method (Green and Sambrook, 2017) and were sequenced on ABI 3730XL DNA Analyzer (Applied Biosystems, USA) using bacterial universal primers 343R, 704F, 907F, 1028F (Baker et al., 2003). The sequences were assembled and curated manually. The phylogenetically closest relatives were searched on the EzBioCloud database (Yoon et al., 2017). The phytoplasma titer in all the samples was quantified by TaqMan-based qPCR assays utilizing phytoplasma-specific primers and probes described earlier (Christensen et al., 2004).

2.3. Enrichment of Phytoplasma DNA and Genome Sequencing

The single, contiguous, complete genome sequences of parthenium phyllody phytoplasma isolates PR08 and PR34 were obtained in multiple DNA enrichment and sequencing attempts. The genomic DNA (1µg) isolated from phytoplasma infected PR08 and PR34 leaf samples were processed for the enrichment of prokaryotic DNA using the NEBNext

microbiome enrichment kit (Cat. No. E2612L; New England BioLabs, USA). Additionally, the obtained enriched DNA was amplified using Ready-To-Go GenomiPhi V3 DNA Amplification Kits (GE25-6601-96; Merck, Germany). The amplified DNA was sequenced on Illumina HiSeq and Oxford Nanopore Technology platforms. Alternatively, the genomic DNA was processed using Long Fragment buffer ONT Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies, UK) and then enriched using the NEBNext microbiome enrichment kit. This enriched DNA was sequenced on the Illumina HiSeq platform. Additionally, the genomic DNA was purified using Genomic-tips 20/G (Cat. No.10223, QIAGEN, Germany), enriched using NEBNext microbiome enrichment kit, and sequenced on the ONT platform. The prokaryotic enrichment was verified at each step by quantifying the reduction in plant 18S rRNA gene using TaqMan-based qPCR assay. The 16S rRNA gene of phytoplasma was quantified simultaneously using phytoplasma-specific primers and probes (Christensen et al., 2004).

2.4. Assembly, Annotation, and OGRI values

The metagenomic raw reads obtained from all Illumina HiSeq and ONT sequencing runs were processed for taxonomic classification using the Kaiju web server to check the nature of enrichment and abundance of sequence reads (Menzel et al., 2016). All QC- passed Illumina reads were processed for assembly using MEGAHIT v1.1.3 (Li et al., 2016). The obtained assembly was subjected to metagenomic binning using MetaBAT2 v2.12.1 (Kang et al., 2015). The taxonomic assignment of bins was done using CheckM v1.0.14 (Parks et al., 2015). The Illumina reads were mapped to available PWB phytoplasma genomes and CheckM generated phytoplasma bins using Bowtie2 v2.3.5.1 (Langmead and Salzberg, 2012). These genomes were Peanut Witches' broom NTU2011 (NZ_AMWZ00000000) (Chung et al., 2013), Echinacea purpurea witches' -broom phytoplasma NCHU2014 (CP040925) (Tan et al., 2021), and '*Ca*. P. aurantifolia' WBDL (MWKN00000000) (El-Sisi et al., 2018). The ONT data was base-called using GUPPY with quality filtering of Q7. The reads were further filtered using Nanofilt v2.8.0 (De Coster et al. 2018) to remove reads shorter than 1kb and mapped to the available PWB phytoplasma genomes using minimap2 v2.22-r1101 (Li, 2018).

The mapped reads from both the platforms were used to obtain hybrid assembly in Unicycler v0.4.8 (Wick et al., 2017). The order and orientation of contigs were determined, and scaffolding was done by MeDuSa v1.6 (Bosi et al., 2015) using the complete chromosome of Echinacea purpurea witches'-broom phytoplasma (CP040925) (Tan et al., 2021) as the reference genome. To validate the final version of the assembly, the Illumina raw reads and ONT reads were mapped to the assembly using Bowtie v2.3.5.1 and

minimap2 v2.22-r1101, respectively. The resulting alignments were visually inspected using IGV v2.3.57 (Thorvaldsdóttir et al., 2013). The genome coverage was estimated using BBMap (Bushnell, 2014). The assembly was annotated using Prokka v1.14.6 (Seemann, 2014) and the coding sequences (CDS) obtained from Prokka were used as a query for BLASTX search against the NCBI nr database to identify and remove possible sequence contamination from the plant host or other microbes.

Finally curated assemblies of parthenium phyllody isolates PR08 and PR34 were submitted to DDBJ/ENA/GenBank database and were annotated using PGAP (Zhao et al., 2012). The orthoANI values were calculated using EzBioCloud orthoANI calculator (Yoon et al., 2017). Intergenomic distances were derived digitally using Genome-to-Genome Distance Calculator (GGDC) (Auch et al., 2010). The functional annotation was carried out using eggNOG-mapper v2 (Cantalapiedra et al., 2021). The comparative analysis of orthologous gene clusters among the genomes of PWB phytoplasmas was performed using OrthoVenn2 (Xu et al. 2019).

2.5. Genome Structure and Effector Protein Analysis

Secretary proteins of PWB phytoplasma isolates PR08 and PR34 were predicted by SignalP V4.1 (Nielsen, 2017) with default cut-off values of 0.57 and V5.0 (Armenteros et al., 2019) using the Gram-positive bacteria model. The presence of transmembrane domain and nuclear localization signal in secretary proteins was predicted by TMHMM V2.0 (Krogh et al., 2001) and NLS mapper (Kosugi et al., 2009b), respectively. The virulence factors were identified based on sequence homology in the core and curated dataset of the Virulence Factors of Pathogenic Bacteria (VFDB) database (Chen et al., 2005) using the BlastP program. The unique homologs were scanned for domain matches against the InterPro protein signature databases using the InterProScan tool (Blum et al., 2021). Homologs of known effector proteins were also searched in genomes of isolate PR08 and PR34.

2.6. Phylogenetic Analyses

The phylogenetic analysis was initially carried out based on the partial 16S rRNA gene sequences using the Neighbour-Joining (NJ), Maximum-Likelihood (ML) and Maximum-Parsimony (MP) methods in MEGA7 (Kumar et al., 2016). Additionally, the phylogenetic position of isolates PR08 and PR34 was assessed by the pan-genome phylogenetic tree, constructed using BPGA v1.3.0 (Chaudhari et al., 2016) and UBCG v3.0 (Na et al., 2018) tools. Briefly, phylogenetically related genomes were downloaded for construction of whole-genome phylogeny and annotated using Blast2GO (Conesa and Götz, 2008) and were searched for core genes using the BPGA tool. The phylogenetic tree was constructed using

the BPGA concatenated file by the Neighbour-Joining method in MEGA7. The UBCG phylogenetic tree was constructed consisting of HMMER (Potter et al., 2018) generated 89 core markers and visualized in MEGA 7 as described earlier (Kirdat et al., 2021, 2020a).

The obtained phylogenetic positions of PR08 and PR34 were further confirmed by the combined phylogenetic analysis using five marker genes proposed by Cho et al., 2020. These marker genes are, Replication initiation protein DnaD (dnaD), DegV family protein (degV), TIGR00282 family metallophosphoesterase, Preprotein translocase SecY (secY), and RluA family pseudo uridine synthase (rluA) genes. These genes were deduced based on high densities of informative sites and relatively high substitution rates compared to single-copy genes reported earlier (Cho et al., 2020).

2.7. PCoA Analysis

The gene content of PWB genomes was compared by Principal Coordinates Analysis (PCoA) (Gower, 2014). The genome sequences of Sesame phyllody SS02 (JAHBAJ000000000.2) (Ranebennur et al., 2022), Peanut Witches broom NTU2011 (NZ_AMWZ00000000) (Chung et al., 2013), Echinacea purpurea witches -broom phytoplasma NCHU2014 (CP040925) (Tan et al., 2021) closely related to '*Ca*. P. australasia' and WBDL (MWKN00000000), the reference isolate of '*Ca*. P. aurantifolia' (El-Sisi et al., 2018; Zreik et al., 1995) were selected for this analysis. The gene cluster obtained using OrthoMCL was converted into a matrix of 6 genomes containing 476 genes clusters (Fischer et al., 2011). This matrix was then converted into a Jaccard distance matrix among genomes using the VEGAN package in R (Dixon, 2003), then processed using the PCoA function in the APE package (Paradis and Schliep, 2019).

3. Results and Discussion

3.1. Phytoplasma Identification and Quantification

The partial 16Sr rRNA gene sequences obtained from 36 symptomatic parthenium samples were analyzed, revealing that they are belonging to PWB or 16SrII group of phytoplasmas. Among them, 16 parthenium phyllody samples showed 99.02 % to 99.40 % sequence similarity to '*Ca.* P. aurantifolia' isolate WBDL (U15442) and 20 showed 99.28% to 100% sequence similarity with '*Ca.* P. australasia' isolate Carica papaya (Y10097). Interestingly, no differences in symptoms were observed between these groups, regardless of geographic area or phytoplasma isolate. The partial 16S rRNA gene sequences of symptomatic parthenium phyllody isolate PR08 (LN879443) matched 100 % similar to '*Ca.* P. australasia' isolate Carica papaya (Y10097) and of PR34 (MZ724173) matched 99.38% to '*Ca.* P.

aurantifolia' isolate WBDL (U15442). These isolates were selected for genome sequencing based on their phytoplasma titer compared to other samples (data not shown).

3.2. Prokaryotic DNA Enrichment and Genome Sequencing

The whole-genome sequencing of NEBNext enriched and Illustra amplified DNA of isolate PR34 resulted in 54,720,081 Illumina sequencing reads (~12 GB data, 150x2 chemistry). The NEBNext Microbiome Enrichment kit yielded less than 40 ng (<1ng/µl in 50 µl) enriched DNA with an input of 1µg of genomic DNA; the Illustra amplification was therefore included to aid the library preparation with a sufficient quantity of DNA. However, the de novo assembly generated using Illustra-generated sequencing data was highly fragmented with 134 contigs (GCA_015100165.1), most likely due to biased amplification of a section of the phytoplasma genome. The amplification strategy using Illustra-generated sequences did not cover the entire phytoplasma genome. Also, although the DNA was enriched, it was still contaminated with significant amount of host genomic DNA. The raw reads from each sequencing run were initially analyzed on the Kaiju server to assess the level of 'prokaryotic' DNA enrichment (Figure 1).

For the second iteration, the genomic DNA strands of 3 kb or longer were size selected using Long Fragment Buffer (LFB) and subjected to prokaryotic DNA enrichment. This approach led to increased enrichment efficiency and generated 54,094,799 Illumina sequencing reads (~15 GB data), substantially reducing the number of contigs (33 contigs, GCA_015100165.2) in better assembly generated subsequently. In a third attempt, genomic DNA purified using Qiagen Genomic-tips 20/G and processed for DNA enrichment showed a significant reduction in plant genomic DNA compared to the earlier two methods (Supplementary Figure S1). The ONT sequencing of this DNA yielded 601,709 raw reads (~1.24 GB data). A single scaffold generated using MeDuSa was inspected manually and corrected for Ns and low coverage bases by mapping raw reads and visualization in IGV. For the final assembly, the data generated from illustra amplified DNA was omitted due to its repetitive nature. The single circular chromosome of size 614,574 (CP097206.1) obtained was rotated to start with the dnaA gene using mauve. The genome coverage for Illumina reads was 5700X, and for ONT, it was 180X.

In its first attempt, the parthenium phyllody isolate PR08 generated a single circular genome using enriched DNA run on the ONT platform (CP060385). However, the PROKKA annotation predicted over 1100 coding regions from the generated assembly. The PGAP annotation predicted over 600 coding regions and 255 pseudogenes, with 77 CDS having internal stops. The genome CP060385 was 52.68 % complete when analyzed using CheckM and therefore

suspected erroneous. In the first Illumina sequencing run, the Long Fragment Buffer (LFB) purified, and NEBNext enriched genomic DNA of isolate PR08 generated 18,519,377 reads (~6.7 GB data). This enriched DNA resulted in 389,506 reads (~1.62 GB of data) on the ONT sequencing platform. The hybrid assembly generated five contigs (JAGXLX010000000), and gaps could not be sealed with available data. The second Illumina run for isolate PR08 resulted in 39,875,638 reads (~13.7 GB data), where DNA was purified using Qiagen Genomic-tips 20/G tips and enriched later. The manually curated Unicycler hybrid assembly generated by combining Illumina and ONT reads resulted in a single circular contig (CP097207). The genome coverage for Illumina reads was 2758X, and for ONT, it was 73X. Supplementary Figure 1 presents a comparison of the enrichment yield based on the input genomic DNA type for parthenium phyllody isolates PR34 and PR08. The Kaiju taxonomic analysis of the raw reads indicated that the phylum Tenericutes (now Mycoplasmatota) exhibited the highest enrichment in all cases, providing insight into the extent of phytoplasma enrichment (Figure 1).

Phytoplasma genomes are characterized by their repetitive nature and nucleotide composition biased towards AT. In the PR34 and PR08 genomes, there are 23 and 17 copies, respectively, of the 'reverse transcriptase domain-containing protein' (RTD). The presence of these highly repetitive regions posed challenges in assembling the PR34 and PR08 genomes into a single scaffold, resulting in multiple contigs and limiting the mapping of raw reads to only a few identical regions. To address this, the contig gaps were manually inspected using IGV and carefully curated to obtain the final single, circular scaffolds for PR34 and PR08. Subsequently, the raw reads were remapped to the single contig to assess the coverage of the final assembly.

Based on the obtained results, Qiagen Genomic-tips 20/G purified and NEBNext enriched DNA samples demonstrated the highest suitability for Illumina sequencing, enabling the acquisition of complete phytoplasma genome sequences. Conversely, the NEBNext prokaryotic enrichment kit's output DNA was found unsuitable for ONT sequencing, resulting in significantly fewer sequences than anticipated, approximately 10 to 15-fold lower. Although the expected DNA quantity was obtained at each step of the ONT library preparation, the diminished sequencing output is speculated to be attributed to an unknown contaminant from the NEBNext enrichment kit, adversely affecting ONT sequencing efficiency. However, the long reads obtained through the ONT sequencing platform played a crucial role in facilitating hybrid assembly and generating a single scaffold circular genome, despite the reduced number of reads.

3.3. Genome Statistics and OGRI Values

The phytoplasma isolate PR34, designated CP097206, has a single scaffold circular genome with a size of 614,574 base pairs and a G+C content of 24.65%. The genome contains two rRNA operons, 474 protein-coding sequences, 28 transfer RNA (tRNA) genes, and 18 pseudogenes. In comparison, the parthenium phyllody phytoplasma isolate PR08 (CP097207) has the smallest reported circular genome among the 16SrII phytoplasmas, measuring 588,746 base pairs in size. It contains two operons, 27 tRNA genes, 468 protein-coding sequences (CDSs), and 15 pseudogenes, with a G+C content of 24.36% (Table 1).

The gene content analysis of parthenium phyllody phytoplasmas confirmed that their genome has reduced content similar to that of other phytoplasmas. Among the protein-coding genes, 304 (64.1%) for PR34 and 352 (75.2%) for PR08 were assigned to 19 functional categories of Orthologous Groups of Proteins (COGs). The most abundant genes in isolate PR34 and PR08 were those responsible for translation, ribosomal structure, and biogenesis, with 104 and 109, respectively. This was followed by genes responsible for transport and metabolism (70 and 71, respectively) (data not shown). The coding density of isolate PR34 is the lowest among the 16SrII phytoplasmas and other complete phytoplasma genomes. Additionally, the genome contains the highest number of tRNA among the 16SrII or PWB phytoplasma genomes (Table 1). The comparison of orthologous gene clusters generated using OrthoVenn2 among PWB phytoplasma genomes revealed distinct patterns of shared clusters and variations in protein content. The OrthoVenn2 occurrence table highlights the presence of specific species within each cluster, while the cluster count and protein count offer quantitative measures of shared clusters and proteins, respectively (Supplementary figure S2).

In addition, the orthoANI values for isolate PR34 were compared to those of *'Ca*. P. aurantifolia' isolate WBDL and *'Ca*. P. australasia' isolate NCHU2014, resulting in values of 95.45% and 87.52 %, respectively. The isolate PR34 shares 65.17 % of its genomic segments with reference isolate WBDL of *'Ca*. P. aurantifolia' and over 75% with *'Ca*. P. australasia'. Using the recommended formula 2 (Auch et al., 2010), digital DNA–DNA hybridization values for PR34 were calculated to be 62.1 and 31.6 against isolate WBDL and NCHU2014, respectively (Table 2).

The OGRI values serve as valuable tools for tracing pathogen dissemination and discerning their potential origins and evolutionary trajectories. By comparing pathogen genomes from diverse sources or regions, their level of relatedness can be determined. Higher OGRI values indicate close relationships among pathogens, suggesting a shared ancestry, whereas lower

values imply less relatedness and potential independent origins. In the case of isolate PR34, the obtained OGRI values distinctly indicate its independent evolution compared to closely related phytoplasma species. The geographic location and plant host range of the phytoplasma isolated -related to isolate PR34 differ from those of *'Ca*. P. aurantifolia' isolate WBDL and *'Ca*. P. australasia' isolate NCHU2014 (Supplementary Table S1).

3.4. Phylogenetic and PCoA Analysis

The phylogenetic analysis based on the 16S rRNA gene sequence reveals that isolate PR34 belongs to a separate clade that is distinct from closely related reference PWB (16SrII) phytoplasmas, *'Ca*. P. aurantifolia' and *'Ca*. P. australasia' (Figure 2). This finding is supported by the pan-genome phylogenetic tree (Figure 3) and the phylogenies inferred using five marker genes (Supplementary Figure S3). These phylogenetic analyses suggest that isolate PR34 has undergone a distinct evolutionary trajectory from the other PWB phytoplasmas and has acquired unique genetic characteristics.

The multivariate analysis method (Principal Coordinate Analysis, PCoA) derived from a similarity matrix of over 476 genes supported the phylogenetic analysis placing the isolate PR34 distantly from both '*Ca.* P. aurantifolia' and '*Ca.* P. australasia' (Figure 4). The PCoA analysis is based on the pairwise comparison of genomes, with the goal of identifying patterns and trends that can reveal the evolutionary relationships between organisms. In this case, the PCoA plot identified 5 clusters of phytoplasma genomes among the 16SrII phytoplasma genomes. The PCoA plot shows the distances between the genomes in a two - dimensional space clearly demonstrating the unique evolutionary position of the isolate PR34. The analysis further showed that the member isolates of '*Ca.* P. australasia' exhibits differences at the genomic level placing the stains of Indian origin (PR08 and SS02) in a distant cluster suggesting a separate evolutionary lineage.

The results of the comprehensive analysis suggest that isolate PR34 exhibits unique genetic features and therefore should be designated as a new species. We propose a novel phytoplasma taxon *'Candidatus* Phytoplasma partheni,' referring to its plant host (*Parthenium hysteroporous* L.); insect vector [*Orosius albicinctus* Distant, (Yadav et al., 2015a)]; its phylogenetic position determined by the 16S rRNA gene, orthologous protein sequences, and whole genome sequence; its distant position according to PCoA coordinates and comparative OGRI values obtained from its genome. The proposed reference isolate PR34 represents the group of PWB phytoplasma isolates mentioned in the Supplementary Table S1.

3.5. Description of *'Candidatus* Phytoplasma partheni'

'*Candidatus* Phytoplasma partheni' (par' the. ni N. L. gen. n. partheni of *Parthenium*) is associated with phyllody disease of *Parthenium hysterophorus* L. The epithet 'partheni' comes from *Parthenium hysterophorus,* referring to the plant host. The reference isolate PR34 is associated with *Parthenium hysterophorus* showing phyllody, witches' broom, little leaf, and stunting symptoms. The reference genome sequence of isolate PR34 is CP097206 and reference 16S rRNA gene sequence is MZ724173. The orthoANI and the digital DNA–DNA hybridization values for isolate PR34 against its closest known relative, *'Ca*. P. aurantifolia' isolate WBDL (NZ_MWKN00000000.1) is 95.45 and 62.1 %; respectively, tested using recommended formula 2.

[Mollicutes] NC; NA; O, wall-less; NAS (CP097206); G+C content 24.65 %; oligonucleotide sequences of unique regions of the 16S rRNA gene are A (353), A (183), A (1006), C (1151), T (1416); P [*Parthenium hysterophorus*., phloem]; M].

3.6. Genome Features

The genomes of the parthenium phyllody phytoplasmas (PR08 and PR34) exhibit unique features that distinguish them from other phytoplasmas within the PWB (16SrII) group as well as outside the group. This includes number and diversity of effector proteins, absence of Potential Mobile Units (PMU) segments, presence of *sodA* gene, presence of truncated hemolysin gene and characteristic presence of high copy number of group II introns.

3.6.1. Presence of wide array of Effector Proteins

Phytoplasmas produce distinct effector proteins that manipulate host plant development, facilitating phytoplasma growth and dissemination (Oshima et al., 2013; Singh et al., 2019). These effectors are transported to the plant cytoplasm through the Sec secretion system, where they interact with host proteins (Sugio et al., 2011; Oshima et al., 2013; Mittelberger et al., 2019). Certain effectors contain nuclear localization signals (NLS) that enable their entry into the nucleus, where they engage with host DNA to modulate host development (Bai et al., 2009; MacLean et al., 2011; Strohmayer et al., 2021).

SignalP 4.1 and SignalP 5.0 predicted 23 secretory proteins for isolate PR34, while SignalP 4.1 predicted 10 secretory proteins for isolate PR08, including products of the *rpsL* and *rpsR* genes encoding 30S ribosomal subunits. On the other hand, SignalP 5.0 predicted 27 secretory proteins for isolate PR08, including those predicted by SignalP 4.1 but excluding the rpsL and *rpsR* gene products. Despite the difference in predicted

protein numbers, SignalP 4.1 is recommended for phytoplasma secretory protein prediction due to its consistent and comprehensive ability (Garcion et al., 2021). The signal peptide associated with these secretory proteins was identified as a phytoplasma-specific terminal sequence-variable mosaic (SVM) protein signal sequence (Pfam entry: PF12113). Nuclear localization signal (NLS) analysis was performed for all secretory proteins in isolate PR34, except for four proteins with a transmembrane domain. Two proteins in PR34, namely SAP11 homolog (UQV27401) with an NLS score of 7.9 and SAP_08 homolog (UQV27394) with an NLS score of 7.5, displayed NLS signals above 7 (Kosugi et al., 2009b, 2009a). In PR08, one protein with an NLS score above 7 was identified as a homolog of SAP_08 from AY-WB (UQV26961; NLS score = 7.3). The homolog of SAP11 (UQV26586) in PR08 exhibited an NLS score of 5.6.

Further, the proteins UQV27401 (PR34) and UQV26586 (PR08) showed 40.50 % and 41.2% sequence similarity to SAP11 of the isolate AYWB, respectively. The SAP11-secretary protein first reported in AYWB phytoplasma is known to destabilize class II TB/CYC-TCP transcription factors, implicating their roles in witches' broom symptoms (Pecher et al., 2019). The functional SAP11 protein is known to have three domains viz signal peptide domain, Nuclear Localization domain, and the TCP binding domain (Sugio et al., 2014). In the case of SAP54, it has been hypothesized that binding to MADS-domain TFs is a characteristic of the hydrophobic pattern achieved through convergent sequence evolution rather than the amino acid sequence in k- the domain (Rümpler et al., 2015). The hydrophobicity pattern of UQV27401 (PR34) and UQV26586 (PR08) proteins was comparable to SAP11 owing to the conservative substitutions implying they are likely to be active and responsible for the witches' broom symptoms (Figure 5A).

Effector proteins evolve faster than the housekeeping genes to adapt to the host environment (Ma and Guttman, 2008). The phylogenetic analysis SAP11 homologs of PWB phytoplasmas reveals that the SAP11 homolog of parthenium phyllody phytoplasma PR34 has evolved differently from the SAP11 homologs of *'Ca*. P. aurantifolia' and *'Ca*. P. australasia' (Figure 5B).

The homolog of SAP_08 in isolate PR34 and PR08 showed the presence of NLS with a strong score (7.5 and 7.3), indicating the protein might travel to the nucleus and interact with DNA leading to symptom development. Additionally, the isolate PR34 and PR08 showed the presence of homologs of ATP_189, Eff115, Eff197, Eff211, PHYL1, SAP05, SAP15, SAP19, SAP21-like, SAP22, SAP40, SAP41, SAP42, SAP49, SAP69 with amino acid sequence similarity 34.7% to 96.8%.

3.6.2. Absence of Potential Mobile Units (PMUs)

The genomes of phytoplasmas are distinguished by their potential mobile units (PMUs), which have played a significant role in chromosomal rearrangement (Dickinson, 2010a; Ku et al., 2013b). Many phytoplasma genomes contain numerous PMUs that are often organized in clusters that resemble composite transposons. The longest and most complete PMU in the AYWB phytoplasma, which is about 20 kb in length, exists in both linear chromosomal and circular extrachromosomal forms. This suggests that PMUs can transpose through circular intermediates. This PMU is bordered by the tra5 gene and contains ORFs for genes involved in DNA replication (*ssb, dnaB, and dnaG*), synthesis (*tmk*), and recombination (*himA*) (Toruño et al., 2010). Notably, no continuous region containing this set of genes was found in the genomes of isolate PR34 and PR08. However, some of the core PMU genes were found scattered throughout both genomes.

The genome of the PR34 isolate has four copies of the sigma-70 family RNA polymerase sigma factor, which is a putative member of the PMU. However, this gene is mutated and has become a pseudogene. Additionally, the genome contains three copies of the single-stranded DNA-binding protein (ssb), one of which is also a pseudogene. The genome also contains another pseudogene, IS3 family transposase (*ISErh1* gene named tra5). The genome has two copies of both the ATP-dependent zinc metalloprotease *ftsH* (*hflB*) gene and the thymidylate kinase (tmk) gene. There is one copy each of the DNA-binding protein HU (*himA*), DNA-directed RNA polymerase subunit alpha, RNA polymerase subunit sigma, replicative DNA helicase (*dnaB*), and DNA primase (*dnaG*). These genes are scattered throughout the genome with different orientations and do not form a cluster as observed in the AYWB genome (NC_007716). However, there is a single instance where the single-stranded DNA-binding protein and the sigma-70 family RNA polymerase sigma factor are located upstream of the pseudogene tra5 with undetermined significance.

The phylogenetic analyses on homologs of three PMU signature genes (*hflB*, *dnaB*, and *dnaG*) from the PMU of PWB phytoplasma NTU2011 suggest horizontal gene transfer (HGT) of PMU (Chung et al., 2013). This observation also explains the coexistence of drastically different PMUs in '*Ca*. P. australiense' (Ku et al., 2013b; Tran-Nguyen et al., 2008). Moreover, PMUs may be involved in HGT with other bacteria in the same ecological niche, as indicated by the horizontal transfer of a gene in AYWB PMU3 (*mgs1*) and spiroplasmas (Ku et al., 2013a). The transfer of PMUs facilitates the dissemination of effector genes and broadens the host range of recipient phytoplasma isolates. However, integrated PMUs undergo gradual degradation over time, primarily driven by

random mutations, deletional bias, and genetic drift in obligate pathogens. Consequently, PMUs may undergo substantial changes, rendering them unrecognizable, while selectively advantageous effector genes are retained, as evidenced by the genomes of PR08 and PR34 analyzed in this study (Ku et al., 2013b; Kuo et al., 2009; Kuo and Ochman, 2009). AYWB PMUs (PMU2-4) and other PMU-like regions in the AYWB genome (NC_007716) contain fewer or truncated ORFs and are degenerated versions of PMU1.

3.6.3. Presence of Superoxide dismutase genes

Superoxide dismutase (SODs) is a primary cellular antioxidant defense system for the survival of intracellular pathogens against reactive oxygen species (ROS) produced by the plant defense (Tripathy & Oelmüller, 2012; Maurya & Namdeo, 2021; Dumanović et al., 2021). Production of ROS in response to phytoplasma infection by plants has been reported earlier (Musetti et al., 2004, 2005). Here we identified the *sodA* gene encoding Fe-Mn SOD in isolate PR34 (UQV27294) and PR08 (UQV26792). Homologs of the *sodA* gene are found in previously reported phytoplasma genomes of '*Ca*. Phytoplasma asteris' isolate OY-M (Bai et al., 2006), '*Ca*. Phytoplasma mali' (Seemuller and Schneider, 2004), '*Ca*. Phytoplasma australiense' (Tran-Nguyen et al., 2008) and *'Ca*. Phytoplasma prunorum' (Seemuller and Schneider, 2004). The integrity of the *sodA* gene has been maintained in phytoplasma genomes irrespective of rearrangement and reductive evolution, indicating a functional role of SOD enzymes. The antioxidant enzymatic activity of SOD was confirmed in OY phytoplasma (Miura et al., 2012).

However, the absence of peroxidases (POX) or catalases (CAT) in both sequenced phytoplasma genomes raises the possibility of alternative, yet uncharacterized enzymes downstream of the superoxide dismutase (SOD) enzyme. In the major ROS scavenging pathway, SOD converts superoxide radicals (H_2O_2) into hydrogen peroxide, which is typically detoxified by POX or CAT enzymes (Michiels et al., 1994). According to previous reports, two mycoplasmal genes, MGA1142 of *Mycoplasma gallisepticum* and MG_454 of *M. genitalium,* encode an antioxidant protein that functions as an organic hydroperoxide reductase (Jenkins et al., 2008; Saikolappan et al., 2009). It has also been reported that a peroxiredoxin that detoxifies H_2O_2, MhPrx, was present in *M. hyopneumoniae* (Machado et al., 2009). Other bacteria, such as *Streptococcus mutans*, have been reported to defend against ROS stress with alkyl hydroperoxide reductase (*ahpR*) as an alternative to CAT (Higuchi et al., 2000). However, the homologous genes for these antioxidant proteins are not found in phytoplasma genomes earlier (Bai et al., 2006; Oshima et al., 2013) or in PR08 and PR34 genomes. The SOD protein of

Mycobacterium tuberculosis is known to be released in the extracellular milieu (Dussurget et al., 2001; Tullius et al., 2001). But the Fe-Mn SOD of parthenium phyllody phytoplasma did not show the presence of terminal sequence-variable mosaic (SVM) protein signal sequence (Pfam entry: PF12113), which is described as a detectable feature associated with secretion (Jomantiene et al., 2007).

Further, the thioredoxin reductase (TrxR) system found in mycoplasmas has been proposed to play a protective role against reactive oxygen compounds (Ben-Menachem et al., 1997). The TrxR system defends against oxidative stress by reducing oxidized TrxR, which, in turn, can activate the antioxidant functions of SOD. The reduced form of TrxR can interact with and activate various antioxidant enzymes, including SOD, thereby increasing the cellular defense against ROS-induced oxidative stress (Espinosa-Diez et al., 2015). Overall, the TrxR system and the SOD gene are two crucial components of the cellular defense against oxidative stress. The presence of thioredoxin family proteins in isolate PR34 (UQV27055) and PR08 (UQV26564) suggests that phytoplasma may use either or both systems in response to oxidative stress within the cell.

3.6.4. Presence of Multiple and Truncated Hemolysin Genes

The presence of multiple and truncated hemolysin (*hlyB*) genes was discovered through the analysis of PR08 and PR34 genomes using the Virulence Factor database (VFDB). Hemolysin genes encode cytolytic toxins that target and disrupt host cell membranes, benefiting bacteria by providing nutrients and facilitating infection spread. Hemolysins play various roles in cellular processes, such as quorum sensing and stress responses. The expression of hemolysin genes is regulated by a complex network of factors and environmental conditions (Bhakdi et al., 1988). The T1SS system found in gram-negative organisms consists of crucial genes involved in the synthesis, modification, and secretion of hemolysin. These genes include *hlyA*, which codes for a polypeptide that requires modification by the *hlyC* gene to become active, and *hlyB* and *hlyD*, which encode proteins involved in exporting the hemolysin. The *tolC* gene, not part of the hly operon, is responsible for hemolysin secretion (Thomas et al., 2014; Wandersman, 1992). Both these genomes lack homologs of *hlyA, hlyC, hlyD*, and *TolC* genes.

The genomes of PR34 and PR08 found to have multiple copies (7 and 13, respectively) of the *hlyB* gene, with varying lengths (246 to 614 aa in PR34 and 212 to 499 aa in PR08), containing an ABC transporter-like ATP-binding domain. The *hlyB* genes in isolate PR34 are arranged consecutively as (UQV27256-57 and UQV27384-85). ABC transporters are membrane proteins that use ATP as energy to transport different substances in and out

of cells where hlyB is a smaller ABC transporter that forms a functional pair when ATP binds to it (Thomas et al., 2014). Two hlyB homologs in isolate PR34 (577 to 614 aa) exhibited ABC transporter type 1 transmembrane and ATP-binding domains. The identified proteins lack a peptidase domain with an active site, unlike in *E. coli*.

Some phytoplasma isolates possess hemolysin III (HlyIII) genes belonging to the RTX family of pore-forming toxins. HlyIII forms pores in the host cell membrane, leading to lysis and tissue damage. hlyA and HlyIII, although important virulence factors in their respective bacterial species, have different molecular structures and mechanisms of action. However, these genes were not detected in the genomes of PR08 and PR34.

The genomes of parthenium phyllody phytoplasma exhibit the presence of a septation protein from the SpoVG family. SpoVG is a DNA/RNA-binding protein that plays a crucial role in various essential bacterial processes, including cell division, sporulation, biofilm formation, and virulence (Burke et al., 2016; Huang et al., 2021; Benthien et al., 2022). It has been reported to be involved in hemolysis caused by *Bacillus subtilis* (Pan et al., 2014). However, the specific mechanism and targets of SpoVG in phytoplasmas are under investigation, and further research is necessary to fully comprehend its role in this process.

3.6.5. Presence of Self-catalytic, High-copy Group II Introns

Multiple copies of full-length and truncated Group II introns were identified in the genome of isolates PR08 and PR34. These are mobile genetic elements found in all domains of life, exhibiting diversity within the bacterial kingdom. These introns exist in ORF-less and ORF-containing forms, with the RNA component accounting for 600 to 900 bp and the Intron encoded proteins (IEPs) ORF being 1800 bp, approximately (Zimmerly and Semper, 2015). The splicing process involves intermediate lariat formation and can occur through retro-homing or retro-transposition (Bonen and Vogel, 2001; Lambowitz and Zimmerly, 2011, 2004; Zimmerly and Semper, 2015).

Bacterial group II introns, often surrounded by other mobile genetic elements, are known to transfer horizontally across the species in addition to vertical inheritance (Klein and Dunny, 2002; Rest and Mindell, 2003; Sheveleva and Hallick, 2004). The genomes of PR08 and PR34, Group II introns are spread throughout the genome independent of adjacent genes (Figure 6A). Six full-length IEPs were annotated in the genome of PR34 along with 7 truncated and 9 pseudo-ORFs (23 in total), while in PR08, five were annotated as full-length IEPs, 11 truncated and 3 pseudo-genes according to Zimmerly's

group II intron database (Candales et al., 2012) and sequence based InterProScan (Blum et al., 2021) analysis (Figure 6B).

This analysis confirmed that the introns in both these genomes are complete and putatively functional, containing RT-X-En domains. Group II introns have characteristic secondary structures consisting of six domains, with domain IV enclosing the ORF for IEP. They are classified based on RNA sequence and secondary structure, as well as the amino acid sequence of the IEP (Michel and Lang, 1985). Bacteria contain all group II lineages, while mitochondria and chloroplasts have group II ML and CL introns exclusively. The ML (mitochondrial-like) introns and CL (chloroplast-like) introns are subgroups or subtypes of Group II introns (Simon et al., 2008; Toro et al., 2002; Zimmerly et al., 2001). All ORFs in PR34 and PR08 belong to RNA class IIA1 and IEP class ML, except for one ORF in PR34 belongs to RNA class IIB and IEP class B.

The presence of IEP homologs varies among phytoplasma genomes, with some species having multiple copies, while others lack them. They are present in multiple copies in all PWB (16SrII) phytoplasma genomes and genomes of *'Ca.* Phytoplasma solani' (isolate 231/09), Mulberry dwarf phytoplasma, and Strawberry lethal yellows phytoplasma (CPA). The full-length group II introns with complete ORFs are rare in class Mollicutes, often being truncated. They were found in only one species of Anaeroplasma (*Anaeroplasma bactoclasticum*), one species of Haloplasma (*Haloplasma contractile*), and entirely absent in Entomoplasmatales. The phylogenetic analysis suggests that the IEPs in phytoplasmas are more closely related to *Clostridium difficile* and *Lactococcus lactis* than to Acholeplasma or Anaeroplasma (Figure 6C).

Bacterial group II introns exhibit the ability to transfer horizontally across species, alongside vertical inheritance (Dai and Zimmerly, 2002; Kamikawa et al., 2009; Sheveleva and Hallick, 2004; Simon et al., 2009; Simon and Zimmerly, 2008). Horizontal transfer of group II introns in Wolbachia has been demonstrated through the presence of introns from different phylogenetic classes (Leclercq et al., 2011). Additionally, nearly identical group II introns have been identified in multiple strains of *B. cereus*, suggesting their dissemination through bacteriophages (Tourasse and Kolstø, 2008). The sporadic occurrence of group II introns in certain strains of *E. coli, Salmonella, Klebsiella pneumoniae, Proteus mirabilis, Rhizobium*, and *Bradyrizobium* indicates their acquisition through horizontal transfer rather than vertical inheritance (Ferat et al., 1994; Rodríguez-Martínez et al., 2012). The G+C content of the *itrA* gene in phytoplasma exhibits a significant deviation from the rest of the genome, ranging from 36.6% to 37.4%. This observation, along with the phylogenetic analysis and classification of Group II introns,

provides indirect evidence for a potential recent horizontal gene transfer to phytoplasmas.

We observed a remarkably high abundance of Group II introns in the genomes of PWB or 16SrII phytoplasmas, with notable differences in their copy numbers among phytoplasma species (Figure 6B). These group II introns are widely distributed throughout the genome, occupying more than 4% of the total genome length. They represent selfish mobile elements that generate multiple copies within genomes, thereby contributing to genome expansion and rearrangement. The growing reports based on whole genome sequencing of endosymbionts suggest that their interrelation with mobile genetic elements is much more complex, as exemplified by Rickettsiales (Darby et al., 2007) and Wolbachia (Leclercq et al., 2011). The high abundance and distribution of group II introns in phytoplasma genomes significantly influenced the genomic arrangement of PWB phytoplasmas. Future studies on group II introns hold promise in unravelling the potential coevolution of these genetic elements and the phytoplasma genome.

4. Conclusion

In conclusion, the study focused on genome sequencing of two phytoplasma isolate associated with parthenium phyllody, a disease affecting legumes and the common weed *Parthenium hysterophorus* in India. The sequencing process involved various methods to enrich DNA and improve assembly quality inferring that the Qiagen Genomic-tips purification were found to be effective for NEBNext enrichment followed by next generation sequencing. The isolate PR34 exhibited a circular genome with distinct genetic features, suggesting it as a new species, '*Ca.* Phytoplamsa partheni'. The isolate PR08 had a smaller circular genome with reduced gene content compared to similar phytoplasmas. Both isolate possessed diverse effector proteins and lacked Potential Mobile Units (PMUs) but had multiple copies of Group II introns. The presence of defense-related genes and unidentified truncated hemolysin genes added to the understanding of their unique genomic characteristics. These findings contribute to our knowledge of phytoplasma evolution and their interaction with host plants.

The GenBank/EMBL/DDBJ accession numbers for the reference 16S rRNA gene sequences of phytoplasmas isolates PR34 and PR08 are MZ724173 and LN879443. The accession number of complete genomes are CP097206 and CP097207. The versions described in this manuscript are CP097206.1 and CP097207.1. Other partial 16Sr rRNA gene sequences were submitted under accession numbers HG792252, LN878981, 82; LN879437 to 43; LT558766 to 69; LT558783, 84, 89; MG748740 to 45; MT555411, 12; MT940950 to 69; MZ724173 and 74.

Author Contributions: KK and AY conducted the sample collection. KK performed the genome sequencing and analysis. KK and BT conducted the phylogenetic analysis and data compilation. KK drafted the initial version of the manuscript. SS revised and edited the initial draft. AY conceptualized, revised, and finalized the manuscript. All authors have read and approved the final version of the manuscript.

Generative AI usage statement: During the preparation of this work the author(s) used 'GPT-4-OpenAI' in order to include concise and jargon free English language. After using this tool/service, the author(s) thoroughly reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding information: The authors acknowledge the project funding and fellowships to K.K. and B.T. by the Department of Science and Technology (DST), Government of India under grant number SERB/EEQ/2016/000752; the authors also acknowledge the funding by Department of Biotechnology (DBT), Government of India under grant number BT/COORD.II/01/03/2016 (NCMR) used for in-house laboratory facilities. The authors gratefully acknowledge the University Grant Commission (UGC) of the Government of India for providing of CSIR-UGC NET-JRF fellowship to K.K. (Ref. No. 857/CSIR-UGC NET JUNE 2017).

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

5. References

Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423. https://doi.org/10.1038/s41587-019-0036-z.

Auch, A.F., von Jan, M., Klenk, H.-P., Göker, M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic. Sci. 2, 117–134. https://doi.org/10.4056/sigs.531120.

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

Bai, X., Zhang, J., Ewing, A., Miller, S.A., Radek, A.J., Shevchenko, D.V., Tsukerman, K., Walunas, T., Lapidus, A., Campbell, J.W., Hogenhout, S.A. 2006. Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. J. Bacteriol. 188, 3682–3696. https://doi.org/10.1128/JB.188.10.3682-3696.2006.

Baker, G.C., Smith, J.J., Cowan, D.A. 2003. Review and re-analysis of domain-specific 16S primers. J. Microbiol. Methods. 55, 541–555. https://doi.org/10.1016/j.mimet.2003.08.009.

Ben-Menachem, G., Himmelreich, R., Herrmann, R., Aharonowitz, Y., Rottem, S., 1997. The thioredoxin reductase system of mycoplasmas. Microbiol. 143, 1933–1940. https://doi.org/10.1099/00221287-143-6-1933.

Benthien, H., Fresenborg, B., Pätzold, L., Elhawy, M.I., Huc-Brandt, S., Beisswenger, C., Krasteva-Christ, G., Becker, S.L., Molle, V., Knobloch, J.K., 2022. The transcription factor SpoVG is of major importance for biofilm formation of *Staphylococcus* epidermidis under *in vitro* conditions, but dispensable for *in vivo* biofilm formation. Int. J. Mol. Sci. 23, 3255. https://doi.org/10.3390/ijms23063255.

Bertaccini, A., Arocha-Rosete, Y., Contaldo, N., Duduk, B., Fiore, N., Montano, H.G., Kube, M., Kuo, C.H., Martini, M., Oshima, K., Quaglino, F., Schneider, B., Wei, W., Zamorano, A. 2022. Revision of the *'Candidatus* Phytoplasma' species description guidelines. Int. J. Syst. Evol. Microbiol. 72. https://doi.org/10.1099/ijsem.0.005353.

Bhakdi, S., Mackman, N., Menestrina, G., Gray, L., Hugo, F., Seeger, W., & Holland, I. B. 1988. The hemolysin of *Escherichia coli*. Eur. J. Epidemiol., 4, 135–143. https://doi.org/10.1007/BF00144740

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S. 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354. https://doi.org/10.1093/nar/gkaa977.

Bonen, L., Vogel, J. 2001. The ins and outs of group II introns. Trends in Genet. 17, 322–331. https://doi.org/10.1016/s0168-9525(01)02324-1.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R., Fondi, M. 2015. MeDuSa: a multi-draft-based scaffolder. Bioinform. 31, 2443–2451. https://doi.org/10.1093/bioinformatics/btv171.

Burke, T.P., Portnoy, D.A., 2016. SpoVG is a conserved RNA-binding protein that regulates Listeria monocytogenes lysozyme resistance, virulence, and swarming motility. MBio 7, e00240-16. DOI: https://doi.org/10.1128/mBio.00240-16

Bushnell, B. 2014. BBMap: a fast, accurate, splice-aware aligner. Mol. Phylogenet. Evol., Lawrence Berkeley National Lab (LBNL), Berkeley, CA, United States.

Cai, H., Wei, W., Davis, R.E., Chen, H., Zhao, Y., 2008. Genetic diversity among phytoplasmas infecting Opuntia species: virtual RFLP analysis identifies new subgroups in the peanut witches'-broom phytoplasma group. Int. J. Syst. Evol. Microbiol. 58, 1448–1457. https://doi.org/10.1099/ijs.0.65615-0

Candales, M.A., Duong, A., Hood, K.S., Li, T., Neufeld, R.A.E., Sun, R., McNeil, B.A., Wu, L., Jarding, A.M., Zimmerly, S. 2012. Database for bacterial group II introns. Nucleic Acids Res. 40, D187–D190. https://doi.org/10.1093/nar/gkr1043.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. 38, 5825–5829 https://doi.org/10.1093/molbev/msab293.

Chaudhari, N.M., Gupta, V.K., Dutta, C. 2016. BPGA-an ultra-fast pan-genome analysis pipeline. Sci. Rep. 6, 1–10. https://doi.org/10.1038/srep24373.

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., Jin, Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 33, D325–D328. https://doi.org/10.1093/nar/gki008.

Cho, S.-T., Kung, H.-J., Huang, W., Hogenhout, S.A., Kuo, C.-H. 2020. Species boundaries and molecular markers for the classification of 16SrI phytoplasmas inferred by genome analysis. Front. Microbiol. https://doi.org/10.3389/fmicb.2020.01531.

Christensen, N.M., Nicolaisen, M., Hansen, M., Schulz, A. 2004. Distribution of Phytoplasmas in Infected Plants as Revealed by Real-Time PCR and Bioimaging. Mol. Plant-Microbe Interact. 17, 1175–1184. https://doi.org/10.1094/MPMI.2004.17.11.1175.

Chung, W.C., Chen, L.L., Lo, W.S., Lin, C.P., Kuo, C.H. 2013. Comparative Analysis of the Peanut Witches'-Broom Phytoplasma Genome Reveals Horizontal Transfer of Potential Mobile Units and Effectors. PLoS One 8, 1–10. https://doi.org/10.1371/journal.pone.0062770.

Conesa, A., Götz, S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008. https://doi.org/10.1155/2008/619832

Dai, L., Zimmerly, S., 2002. The dispersal of five group II introns among natural populations of Escherichia coli. Rna 8, 1294–1307. https://doi.org/10.1017/s1355838202023014

Darby, A.C., Cho, N.-H., Fuxelius, H.-H., Westberg, J., Andersson, S.G.E., 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. TRENDS Genet. 23, 511–520. https://doi.org/10.1016/j.tig.2007.08.002

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinform. 34, 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Deng, S., Hiruki, C. 1991. Amplification of 16S rRNA genes from culturable and nonculturable mollicutes. J. Microbiol. Methods 14, 53–61. https://doi.org/10.1016/0167-7012(91)90007-D.

Dickinson, M. 2010. Mobile units of DNA in phytoplasma genomes. Mol. Microbiol. 77, 1351–1353. https://doi.org/10.1111/j.1365-2958.2010.07308.x.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14, 927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

Doyle, J.J., Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull., 19, 11-15.

Duduk, B., Stepanović, J., Yadav, A., Rao, G.P. 2018. Phytoplasmas in Weeds and Wild Plants, in: Rao, G. P., Bertaccini, A., Fiore, N. and Liefting, L. W. (Eds.) Phytoplasmas: Plant Pathogenic Bacteria. Springer, p. 313. https://doi.org/10.1007/978-981-13-0119-3_11

Dumanović, J., Nepovimova, E., Natić, M., Kuča, K., Jaćević, V. 2021. The significance of reactive oxygen species and antioxidant defense system in plants: A concise overview. Front. Plant. Sci. 11, 552969. https://doi.org/10.3389/fpls.2020.552969.

Dussurget, O., Stewart, G., Neyrolles, O., Pescher, P., Young, D., Marchal, G. 2001. Role of Mycobacterium tuberculosis copper-zinc superoxide dismutase. Infect. Immun. 69, 529–533. https://doi.org/10.1128/IAI.69.1.529-533.2001.

El-Sisi, Y., Omar, A.F., Sidaros, S.A., Elsharkawy, M.M., Foissac, X. 2018. Multilocus sequence analysis supports a low genetic diversity among *'Candidatus* Phytoplasma australasia' related strains infecting vegetable crops and periwinkle in Egypt. Eur. J. Plant Pathol. 150, 779–784. https://doi.org/10.1007/s10658-017-1311-9.

Espinosa-Diez, C., Miguel, V., Mennerich, D., Kietzmann, T., Sánchez-Pérez, P., Cadenas, S., Lamas, S., 2015. Antioxidant responses and cellular adjustments to oxidative stress. Redox Biol. 6, 183–197. https://doi.org/10.1016/j.redox.2015.07.008

Ferat, J.L., Le Gouar, M., Michel, F., 1994. Multiple group II self-splicing introns in mobile DNA from Escherichia coli. C. R. Acad. Sci. III. 317, 141–148. [No DOI available]

Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., Stoeckert Jr, C.J. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr. Protoc. Bioinform. 35, 6–12. https://doi.org/10.1002/0471250953.bi0612s35.

Garcion, C., Beven, L., Foissac, X. 2021. Comparison of current methods for signal peptide prediction in phytoplasmas. Front. Microbiol. 12, 661524. https://doi.org/10.3389/fmicb.2021.661524.

Gower, J.C. 2014. Principal Coordinates Analysis. Wiley StatsRef: Statistics Reference Online 1–7. https://doi.org/10.1002/9781118445112.stat05670.

Green, M.R., Sambrook, J. 2017. Preparation of single-stranded bacteriophage M13 DNA by precipitation with polyethylene glycol. Cold Spring Harb. Protoc. 2017, pdb-prot093419. https://doi.org/10.1101/pdb.prot093419.

Gundersen, D.E., Lee, I.M., Rehner, S.A., Davis, R.E., Kingsbury, D.T. 1994. Phylogeny of mycoplasma-like organisms (phytoplasmas): A basis for their classification. J. Bacteriol. 176, 5244–5254. https://doi.org/10.1128/JB.176.17.5244-5254.1994.

Higuchi, M., Yamamoto, Y., Poole, L.B., Shimada, M., Sato, Y., Utsumi, R., 2000. Functions of two types of NADH oxidases in energy metabolism and oxidative stress of *Streptococcus* mutans. J. Bacteriol. 181, 5940- 5947 https://doi.org/10.1128/JB.181.19.5940-5947.1999.

Hogenhout, S.A., Seruga, M., 2009. Phytoplasma genomics, from sequencing to comparative and functional genomics-what have we learnt? In: Weintraub, P. G. Jones, P. (Eds.) Phytoplasmas: Genomes, Plant Hosts and Vectors. CABI, UK. pp 19–36.

Huang, Q., Zhang, Z., Liu, Q., Liu, F., Liu, Y., Zhang, J., Wang, G., 2021. SpoVG is an important regulator of sporulation and affects biofilm formation by regulating Spo0A transcription in *Bacillus cereus* 0–9. BMC Microbiol. 21, 172. https://doi.org/10.1186/s12866-021-02239-6.

Jenkins, C., Samudrala, R., Geary, S.J., Djordjevic, S.P., 2008. Structural and functional characterization of an organic hydroperoxide resistance protein from Mycoplasma gallisepticum. J. Bacteriol. 190, 2206–2216. https://doi.org/10.1128/JB.01685-07

Jomantiene, R., Zhao, Y., Davis, R.E. 2007. Sequence-variable mosaics: composites of recurrent transposition characterizing the genomes of phylogenetically diverse phytoplasmas. DNA Cell Biol. 26, 557–564. https://doi.org/10.1089/dna.2007.0610.

Kamikawa, R., Masuda, I., Demura, M., Oyama, K., Yoshimatsu, S., Kawachi, M., Sako, Y., 2009. Mitochondrial group II introns in the raphidophycean flagellate Chattonella spp. suggest a diatom-to-Chattonella lateral group II intron transfer. Protist 160, 364–375. https://doi.org/10.1016/j.protis.2009.02.003

Kang, D.D., Froula, J., Egan, R., Wang, Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165. https://doi.org/10.7717/peerj.1165.

Kirdat, K., Tiwarekar, B., Sathe, S., Yadav, A. 2023. From Sequences to Species: Charting the Phytoplasma Classification and Taxonomy in the Era of Taxogenomics. Front. Microbiol. 14, 555. https://doi.org/10.3389/fmicb.2023.1123783.

Kirdat, K., Tiwarekar, B., Swetha, P., Padma, S., Thorat, V., Manjula, K.N., Kavya, N., Sundararaj, R., Yadav, A. 2022. Nested Real-Time PCR Assessment of Vertical Transmission of Sandalwood Spike Phytoplasma ('*Ca*. Phytoplasma asteris'). Biology (Basel) 11, 1494. https://doi.org/10.3390/biology11101494.

Kirdat, K., Tiwarekar, B., Thorat, V., Narawade, N., Dhotre, D., Sathe, S., Shouche, Y., Yadav, A. 2020a. Draft genome sequences of two phytoplasma strains associated with sugarcane grassy shoot (SCGS) and Bermuda grass white leaf (BGWL) diseases. Mol. Plant-Microbe Interac. 33, 715–717. https://doi.org/10.1094/MPMI-01-20-0005-A.

Kirdat, K., Tiwarekar, B., Thorat, V., Sathe, S., Shouche, Y., Yadav, A. 2021. *'Candidatus* Phytoplasma sacchari', a novel taxon-associated with Sugarcane Grassy Shoot (SCGS) disease. Int. J. Syst. Evol. Microbiol. 71, 1–7. https://doi.org/10.1099/ijsem.0.004591.

Kirdat, K., Tiwarekar, B., Thorat, V., Sathe, S., Yadav, A. 2020b. First report of association of a 16SrII group phytoplasma with a witches' broom disease of *Croton bonplandianum*. Phytopathogenic Mollicutes 10, 100–103. https://doi.org/10.5958/2249-4677.2020.00013.4.

Klein, J.R., Dunny, G.M. 2002. Bacterial group II introns and their association with mobile genetic elements. Front Biosci. 7, d1843–56 https://doi.org/10.2741/klein1.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., Yanagawa, H. 2009a. Six classes of nuclear localization signals specific to different binding grooves of importin α. J. Biol. Chem. 284, 478–485. https://doi.org/10.1074/jbc.M807017200.

Kosugi, S., Hasebe, M., Tomita, M., Yanagawa, H. 2009b. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. USA. 106, 10171–10176. https://doi.org/10.1073/pnas.0900604106.

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

Ku, C., Lo, W.-S., Chen, L.-L., Kuo, C.-H. 2013a. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in Spiroplasma. Genome Biol. Evol. 5, 1151–1164. https://doi.org/10.1093/gbe/evt084.

Ku, C., Lo, W.-S., Kuo, C.-H. 2013b. Horizontal transfer of potential mobile units in phytoplasmas. Mob. Genet. Elements 3, e62770. https://doi.org/10.4161/mge.26145.

Kube, M., Mitrovic, J., Duduk, B., Rabus, R., Seemüller, E. 2012. Current view on phytoplasma genomes and encoded metabolism. ScientificWorldJournal. 2012:185942 https://doi.org/10.1100/2012/185942.

Kumar, S., Stecher, G., Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

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

Kuo, C.-H., Ochman, H. 2009. Deletional bias across the three domains of life. Genome Biol. Evol. 1, 145–152. https://doi.org/10.1093/gbe/evp016.

Lambowitz, A.M., Zimmerly, S. 2004. Mobile group II introns. Annu. Rev. Genet. 38, 1–35. https://doi.org/10.1146/annurev.genet.38.072902.091600.

Lambowitz, A.M., Zimmerly, S. 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. 3, a003616. https://doi.org/10.1101/cshperspect.a003616

Langmead, B., Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9, 357. https://doi.org/10.1038/nmeth.1923

Lava Kumar, P., Sharma, K., Boahen, S., Tefera, H., Tamo, M., 2011. First report of soybean witches'-broom disease caused by group 16SrII phytoplasma in soybean in Malawi and Mozambique. Plant Dis. 95, 492. https://doi.org/10.1094/PDIS-01-11-0016

Leclercq, S., Giraud, I., Cordaux, R. 2011. Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol. Biol. Evol. 28, 685–697. https://doi.org/10.1093/molbev/msq238

Lee, I.M., Davis, R.E., Gundersen-Rindal, D.E. 2000. Phytoplasma: Phytopathogenic Mollicutes. Annu. Rev. Microbiol. 54, 221–255. https://doi.org/10.1146/annurev.micro.54.1.221

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., Lam, T.-W. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinform. 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Ma, W., Guttman, D.S. 2008. Evolution of prokaryotic and eukaryotic virulence effectors. Curr. Opin. Plant Biol. 11, 412–419. https://doi.org/10.1016/j.pbi.2008.05.001

Machado, C.X., Pinto, P.M., Zaha, A., Ferreira, H.B., 2009. A peroxiredoxin from *Mycoplasma hyopneumoniae* with a possible role in H2O2 detoxification. Microbiol. 155, 3411–3419. https://doi.org/10.1099/mic.0.030643-0

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

Makarova, O., Contaldo, N., Paltrinieri, S., Kawube, G., Bertaccini, A., and Nicolaisen, M. 2012. DNA barcoding for identification of '*Candidatus* Phytoplasmas' using a fragment of the elongation factor Tu gene. PLoS One 7, e52092. https://doi.org/10.1371/journal.pone.0052092

Mall, S., Kirdat, K., Singh, A., Tiwarekar, B., Sathe, S., Marcone, C., Yadav, A. 2023. Updates on phytoplasma diseases associated with weeds acting as alternate hosts in Asian countries. Phytoplasma Diseases of Major Crops, Trees, and Weeds, (Volume II), Tiwari, A.K., Caglayan, K., Hoat, T.X., Subhi, A., Nejat, N., Reddy, G. (Eds.), Academic Press, London, pp. 347–365. https://doi.org/10.1016/B978-0-323-91897-8.00011-3

Marcone, C. 2014. Molecular biology and pathogenicity of phytoplasmas. Ann. Appl. Biol. 165, 199–221. https://doi.org/10.1111/aab.12151

Martini, M., Lee, I.-M., Bottner, K.D., Zhao, Y., Botti, S., Bertaccini, A., Harrison, N.A., Carraro, L., Marcone, C., Khan, A.J., 2007. Ribosomal protein gene-based phylogeny for finer differentiation and classification of phytoplasmas. Int. J. Syst. Evol. Microbiol. 57, 2037–2051. https://doi.org/10.1099/ijs.0.65013-0

Maurya, R., Namdeo, M. 2021. Superoxide dismutase: A key enzyme for the survival of intracellular pathogens in the host. In: Ahmed, R., (Ed). Reactive Oxygen Species. https://doi.org/10.5772/intechopen.100322

Menzel, P., Ng, K.L., Krogh, A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7, 1–9. https://doi.org/10.1038/ncomms11257

Michel, F., Lang, B.F. 1985. Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature 316, 641–643. https://doi.org/10.1038/316641a0

Michiels C., Raes M., Toussaint O., Remacle J. 1994. Importance of Se-glutathione peroxidase, catalase, and Cu/Zn-SOD for cell survival against oxidative stress. Free Radic. Biol. Med. 17, 235–248. https://doi.org/10.1016/0891-5849(94)90079-5

Mittelberger, C., Stellmach, H., Hause, B., Kerschbamer, C., Schlink, K., Letschka, T., Janik, K. 2019. A novel effector protein of apple proliferation phytoplasma disrupts cell integrity of Nicotiana spp. protoplasts. Int. J. Mol. Sci. 20, 4613. https://doi.org/10.3390/ijms20184613

Miura, C., Sugawara, K., Neriya, Y., Minato, N., Keima, T., Himeno, M., Maejima, K., Komatsu, K., Yamaji, Y., Oshima, K., 2012. Functional characterization and gene expression profiling of superoxide

dismutase from plant pathogenic phytoplasma. Gene 510, 107–112. https://doi.org/10.1016/j.gene.2012.09.001

Musetti, R., Toppi L., Ermacora, P., Favali, M. A. 2004. Recovery in apple trees infected with the apple proliferation phytoplasma: an ultrastructural and biochemical study. Phytopathol. 94, 203–208. https://doi.org/10.1094/PHYTO.2004.94.2.203

Musetti, R., Toppi L., Martini, M., Ferrini, F., Loschi, A., Favali, M. A., Osler, R. 2005. Hydrogen peroxide localization and antioxidant status in the recovery of apricot plants from European stone fruit yellows. Eur. J. Plant Pathol. 112, 53–61. https://doi.org/10.1007/s10658-004-8233-z

Na, S.-I., Kim, Y.O., Yoon, S.-H., Ha, S., Baek, I., Chun, J. 2018. UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. J. Microbiol. 56, 280–285. https://doi.org/10.1007/s12275-018-8014-6

Nabi, S., Madhupriya, D. D. K., Rao, G. P., Baranwal, V. K., and Sharma, P. 2015. Characterization of phytoplasmas associated with sesame (*Sesamum indicum*) phyllody disease in North India utilizing multi-locus genes and RFLP analysis. Indian Phytopathol. 68, 112–119.

Nielsen, H. 2017. Predicting secretory proteins with SignalP. In: Protein Function Prediction. Springer, pp. 1611, 59–73. https://doi.org/10.1007/978-1-4939-7015-5_6

Observatory of Economic Complexity (OEC) World - India Trade Data. 2023: https://oec.world/en/profile/country/ind (Accessed Jan 12, 2023)

Oshima, K., Maejima, K., Namba, S. 2013. Genomic and evolutionary aspects of phytoplasmas. Front. Microbiol. 4, 1–8. https://doi.org/10.3389/fmicb.2013.00230

Pan, X., Chen, X., Su, X., Feng, Y., Tao, Y., Dong, Z. 2014. Involvement of SpoVG in hemolysis caused by *Bacillus subtilis*. Biochem. Biophys. Res. Commun. 443(3), 899–904. https://doi.org/10.1016/j.bbrc.2013.12.069

Paradis, E., Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformat. 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114.

Pecher, P., Moro, G., Canale, M.C., Capdevielle, S., Singh, A., MacLean, A., Sugio, A., Kuo, C.-H., Lopes, J.R.S., Hogenhout, S.A. 2019. Phytoplasma SAP11 effector destabilization of TCP transcription factors differentially impact development and defense of *Arabidopsis* versus maize. PLoS Pathog.15, e1008035. https://doi.org/10.1371/journal.ppat.1008035

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., Finn, R.D. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46, W200–W204. https://doi.org/10.1093/nar/gky448

Pramesh, D., Prasanna Kumar, M., Buela, P., Yadav, M., Chidanandappa, E., Pushpa, H., et al. 202). First report of '*Candidatus* Phytoplasma aurantifolia' associated with phyllody disease of snake weed in India. Plant Dis. 104, 277. https://doi.org/10.1094/PDIS-04-19-0825-PDN

Ranebennur, H., Kirdat, K., Tiwarekar, B., Rawat, K., Chelam, C., Solanke, A., Yadav, R., Singh, K., Sathe, S., Yadav, A. 2022. Draft Genome Sequence of '*Candidatus* Phytoplasma australasia', strain SS02 associated with sesame phyllody disease. 3Biotech 12, 107. https://doi.org/10.1007/s13205-022-03163-w

Rao, G. P., Bahadur, A., Das, S. C., Ranebennur, H., Mitra, S., Kumar, M., et al. 2020. First report of 16Sr II-C subgroup phytoplasma association with *Acacia mangium* in Tripura, India. For. Pathol. 50, e12573. https://doi.org/10.1111/efp.12573

Rao, G.P., Kumar, M., Tomar, S., Maya, B., Singh, S.K., Johnson, J.M. 2018. Detection and identification of four 16Sr subgroups of phytoplasmas associated with different legume crops in India. Eur. J. Plant Pathol. 150, 507–513. https://doi.org/10.1007/s10658-017-1278-6

Rao, G.P., Madhupriya, T.V., Manimekalai, R., Tiwari, A.K., Yadav, A. 2017. A century progress of research on phytoplasma diseases in India. Phytopathogenic Mollicutes 7, 1–38. https://doi.org/10.5958/2249-4677.2017.00001.9

Ravi, M., Rao, G. P., Meshram, N. M., and Sundararaj, R. 2022. Genetic diversity of phytoplasmas associated with several bamboo species in India. For. Pathol. 52, e12741. https://doi.org/10.1111/efp.12741

Reddy, M. G. 2020. Characterization of pathogens associated with chickpea stunt disease and its epidemiology. 3 Biotech 11, 112. 10.1007/s13205-020-02613-7

Rest, J.S., Mindell, D.P. 2003. Retroids in archaea: phylogeny and lateral origins. Mol Biol Evol. 20, 1134–1142. https://doi.org/10.1093/molbev/msg135

Rodríguez-Martínez, J.-M., Nordmann, P., Poirel, L., 2012. Group IIC intron with an unusual target of integration in Enterobacter cloacae. J. Bacteriol. 194, 150–160. https://doi.org/10.1128/JB.05786-11

Rümpler, F., Gramzow, L., Theißen, G., Melzer, R. 2015. Did convergent protein evolution enable phytoplasmas to generate 'zombie plants'? Trends Plant Sci. 20, 798–806. https://doi.org/10.1016/j.tplants.2015.08.004

Saikolappan, S., Sasindran, S.J., Yu, H.D., Baseman, J.B., Dhandayuthapani, S., 2009. The *Mycoplasma genitalium* MG_454 gene product resists killing by organic hydroperoxides. J. Bacteriol. 191, 6675–6682. https://doi.org/10.1128/JB.01066-08

Schneider, B., Seemüller, E., Smart, C.D. and Kirkpatrick, B.C. (1995) Phylogenetic classification of plant pathogenic mycoplasma-like organisms or Phytoplasmas. In: Razin, S. and Tully, J.G. (Eds.), Molecular and Diagnostic Procedures in Mycoplasmology, Vol. I. Academic Press, San Diego, 369-380. http://dx.doi.org/10.1016/B978-012583805-4/50040-6

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seemuller, E., Schneider, B., 2004. '*Candidatus* Phytoplasma mali', '*Candidatus* Phytoplasma pyri' and '*Candidatus* Phytoplasma prunorum', the causal agents of apple proliferation, pear decline and European stone fruit yellows, respectively. Int. J. Syst. Evol. Microbiol. 1217–1226. https://doi.org/10.1099/ijs.0.02823-0

Sheveleva, E. V, Hallick, R.B., 2004. Recent horizontal intron transfer to a chloroplast genome. Nucleic Acids Res. 32, 803–810. https://doi.org/10.1093/nar/gkh225

Simon, D.M., Clarke, N.A.C., McNeil, B.A., Johnson, I., Pantuso, D., Dai, L., Chai, D., Zimmerly, S. 2008. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA 14, 1704–1713. https://doi.org/10.1261/rna.1056108

Simon, D.M., Kelchner, S.A., Zimmerly, S. 2009. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. Mol. Biol. Evol. 26, 2795–2808. https://doi.org/10.1093/molbev/msp193

Simon, D.M., Zimmerly, S. 2008. A diversity of uncharacterized reverse transcriptases in bacteria. Nucleic Acids Res. 36, 7219–7229. https://doi.org/10.1128/microbiolspec.MDNA3-0058-2014

Singh, A., Verma, P., Lakhanpaul, S. 2019. Phytoplasma effector molecules and their structural aspects: A review. Phytopathogenic Mollicutes 9, 241–251. https://doi.org/10.5958/2249-4677.2019.00121.X

Strohmayer, A., Schwarz, T., Braun, M., Krczal, G., Boonrod, K. 2021. The effect of the anticipated nuclear localization sequence of *'Candidatus* Phytoplasma mali' SAP11-like protein on localization of the protein and destabilization of TCP transcription factor. Microorganisms 9, 1756. https://doi.org/10.3390/microorganisms9081756

Sugio, A., MacLean, A.M., Hogenhout, S.A. 2014. The small phytoplasma virulence effector SAP11 contains distinct domains required for nuclear targeting and CIN-TCP binding and destabilization. New Phytol. 202, 838–848. https://doi.org/10.1111/nph.12721

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

Tan, C.M., Lin, Y.-C., Li, J.-R., Chien, Y.-Y., Wang, C.-J., Chou, L., Wang, C.-W., Chiu, Y.-C., Kuo, C.-H., Yang, J.-Y., 2021. Accelerating complete phytoplasma genome assembly by immunoprecipitation-based enrichment and MinION-based DNA sequencing for comparative analyses. Front. Microbiol. 12, 766221. https://doi.org/10.3389/fmicb.2021.766221

Thomas, S., Holland, I. B., & Schmitt, L. 2014. The type 1 secretion pathway—the hemolysin system and beyond. Biochim. Biophys. Acta. 1843(8), 1629–1641. https://doi.org/10.1016/j.bbamcr.2013.09.017

Thorat, V., Bhale, U., Sawant, V., More, V., Jadhav, P., Mane, S.S., Nandanwar, R., Tripathi, S., Yadav, A., 2016a. Alternative weed hosts harbors 16SrII group phytoplasmas associated with little leaf and witches' broom diseases of various crops in India. Phytopathogenic Mollicutes 6, 50–55. https://doi.org/10.5958/2249-4677.2016.00009.8

Thorat, V., Kirdat, K., Takawale, P., Yadav, A., 2017. First report of 16SrII-D phytoplasmas associated with fodder crops in India. Phytopathogenic Mollicutes 7, 106–110. https://doi.org/10.5958/2249-4677.2017.00015.9

Thorat, V., More, V., Jadhav, P., Mane, S.S., Nandanwar, R.S., Surayavanshi, M., Shouche, Y., Yadav, A., 2016b. First report of a 16SrII-D group phytoplasma associated with witches'-broom disease of soybean (*Glycine max*) in Maharashtra, India. Plant Dis. 100, 2521. https://doi.org/10.1094/PDIS-05-16-0741-PDN

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14, 178–192. https://doi.org/10.1093/bib/bbs017

Tiwari, A.K., Gazel, M., Yadav, A., Al-Sadi, A.M., Abeysinghe, S., Nejat, N., Oshima, K., Bertaccini, A., Rao, G.P., 2023a. Overview of phytoplasma diseases in Asian countries, in: Tiwari, A. K., Caglayan, K., Al-Sadi, A. M., Azadvar, M., and Abeysinghe, S. (Eds.), Diversity, Distribution, and Current Status (Volume I). Academic Press, London, pp. 1–30. https://doi.org/10.1016/B978-0-323-91896-1.00016-7

Tiwari, A.K., Tripathi, S., Singh, J., Kirdat, K., Reddy, M.G., Suryanarayana, V., Yadav, A., Rao, G.P., 2023b. The diversity, distribution, and status of phytoplasma diseases in India, in: Tiwari, A. K., Caglayan, K., Al-Sadi, A. M., Azadvar, M., and Abeysinghe, S. (Eds.), Diversity, Distribution, and Current Status (Volume I). Academic Press, London, pp. 281–320. https://doi.org/10.1016/B978-0-323-91896-1.00001-5

Toro, N., Molina-Sánchez, M.D., Fernández-López, M., 2002. Identification and characterization of bacterial class E group II introns. Gene 299, 245–250. https://doi.org/10.1016/s0378-1119(02)01079-x.

Toruño, T.Y., Seruga Musić, M., Simi, S., Nicolaisen, M., Hogenhout, S.A., 2010. Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. Mol. Microbiol. 77, 1406–1415. https://doi.org/10.1111/j.1365-2958.2010.07296.x

Tourasse, N.J., Kolstø, A.-B., 2008. Survey of group I and group II introns in 29 sequenced genomes of the Bacillus cereus group: insights into their spread and evolution. Nucleic Acids Res. 36, 4529–4548. https://doi.org/10.1093/nar/gkn372

Tran-Nguyen, L.T.T., Kube, M., Schneider, B., Reinhardt, R., Gibb, K.S., 2008. Comparative genome analysis of '*Candidatus* Phytoplasma australiense' (subgroup tuf-Australia I; rp-A) and '*Ca*. Phytoplasma asteris' strains OY-M and AY-WB. J. Bacteriol. 190, 3979–3991. https://doi.org/10.1128/JB.01301-07

Tripathy, B.C., Oelmüller, R., 2012. Reactive oxygen species generation and signaling in plants. Plant Signal Behav. 7, 1621–1633. https://doi.org/10.4161/psb.22455

Tullius, M. V, Harth, G., Horwitz, M.A., 2001. High extracellular levels of *Mycobacterium tuberculosis* glutamine synthetase and superoxide dismutase in actively growing cultures are due to high expression and extracellular stability rather than to a protein-specific export mechanism. Infect. Immun. 69, 6348–6363. https://doi.org/10.1128/IAI.69.10.6348-6363.2001

Tully, J.G., 1993. International Committee on Systematic Bacteriology Subcommittee on the Taxonomy of Mollicutes. Minutes of the interim meetings, 1 and 2 Aug. 1992, Ames, Iowa. Int. J. Syst. Bacteriol 43, 394–397. https://doi.org/10.1099/00207713-43-2-394

Wandersman, C. 1992. Secretion across the bacterial outer membrane. Trends Genet., 8(9), 317–322. https://doi.org/10.1016/0168-9525(92)90264-5

Wei, W., Zhao, Y., 2022. Phytoplasma Taxonomy: Nomenclature, Classification, and Identification. Biology (Basel) 11, 1119. https://doi.org/10.3390/biology11081119

White, D.T., Blackall, L.L., Scott, P.T., Walsh, K.B., 1998. 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. Evol. Microbiol. 48, 941–951. https://doi.org/10.1099/00207713-48-3-941

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Xu L, Dong Z, Fang L, et al. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 47: W52–W58. https://doi.org/10.1093/nar/gkz333

Yadav, A., Bhale, U., Thorat, V., Shouche, Y., 2014. First report of a new subgroup 16Sr II-M '*Candidatus* Phytoplasma aurantifolia' associated with witches'-broom disease of *Tephrosia purpurea* in India. Plant Dis. 98, 990. https://doi.org/10.1094/PDIS-11-13-1183-PDN.

Yadav, A., Thorat, V., Bhale, U., Shouche, Y., 2015a. Association of 16SrII-C and 16SrII-D subgroup phytoplasma strains with witches' broom disease of *Parthenium hysterophorus* and insect vector *Orosius albicinctus* in India. Australas. Plant Dis. Notes 10, 31. https://doi.org/10.1007/s13314-015-0181-2

Yadav, A., Thorat, V., Shouche, Y., 2015b. '*Candidatus* Phytoplasma aurantifolia' (16SrII group) associated with Witches' Broom disease of Bamboo (*Dendrocalamus strictus*) in India. Plant Dis. 100, 209.

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., Chun, J., 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int. J. Syst. Evol. Microbiol. 67, 1613. https://doi.org/10.1099/ijsem.0.001755

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., Yu, J., 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics 28, 416–418. https://doi.org/10.1093/bioinformatics/btr655

Zimmerly, S., Hausner, G., Wu, X., 2001. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res 29, 1238–1250. https://doi.org/10.1093/nar/29.5.1238.

Zimmerly, S., Semper, C., 2015. Evolution of group II introns. Mob. DNA 6, 1–19. https://doi.org/10.1186/s13100-015-0037-5

Zreik, L., Carle, P., BOV, J.M., Garnier, M., 1995. Characterization of the Mycoplasma-like organism associated with witches'-broom disease of lime and proposition of a '*Candidatus*' taxon for the organism, '*Candidates* Phytoplasma aurantifolia'. Int. J. Syst. Evol. Microbiol. 45, 449–453. https://doi.org/10.1099/00207713-45-3-449

Table 1. Comparative analysis of the genomic characteristics of different strains belonging to the Peanut Witches' Broom (PWB) group of phytoplasmas. The genomes of six strains, namely *'Ca*. P. partheni' (PR34), *'Ca*. P. australasia' (NCHU2014, NTU2011 SS02 and PR08), and *'Ca*. P. aurantifolia' (WBDL) were analyzed. The genome sizes, number of contigs, coding density, and G+C content were determined, and the presence of proteins, rRNA, and tRNA was evaluated.

Isolate ID	Accession Number	No. of. Contigs	Genome Size (bp)	Proteins	rRNA	tRNA	Coding density	% G+C
PR34	CP097206	1	614,574	474	6	28	70.61	24.65
PR08	CP097207	1	588,746	468	6	27	72.74	24.36
NCHU2014	CP040925	1	639,808	471	6	24	71.88	24.54
NTU2011	AMWZ00000000	14	566,694	448	6	27	73.31	24.37
SS02	JAHBAJ000000000	60	553,228	449	3	17	72.25	23.68
WBDL	MWKN00000000	98	474,669	385	0	19	73.76	23.9

Table 2. OrthoANI and dDDH values shared between *'Ca*. P. partheni' (PR34) and other PWB isolates, including *'Ca*. P. aurantifolia' (WBDL), *'Ca*. P. australasia' (NCHU2014, NTU2011 SS02 and PR08) along with the percentage of shared genomic segments. The observed differences in ANI, dDDH values, and shared genomic segments highlight the genetic diversity within the PWB (16SrII) group.

Isolate ID	ANI values with closest relative	dDDH values with closest relative	Genomic segments shared (%)
PR34	100	100	98.88
WBDL	95.45	62.1	65.17
NCHU2014	87.52	31.6	86.52
NTU2011	87.25	30.8	80.34
PR08	86.99	30.5	85.39
SS02	86.98	30.6	75.84

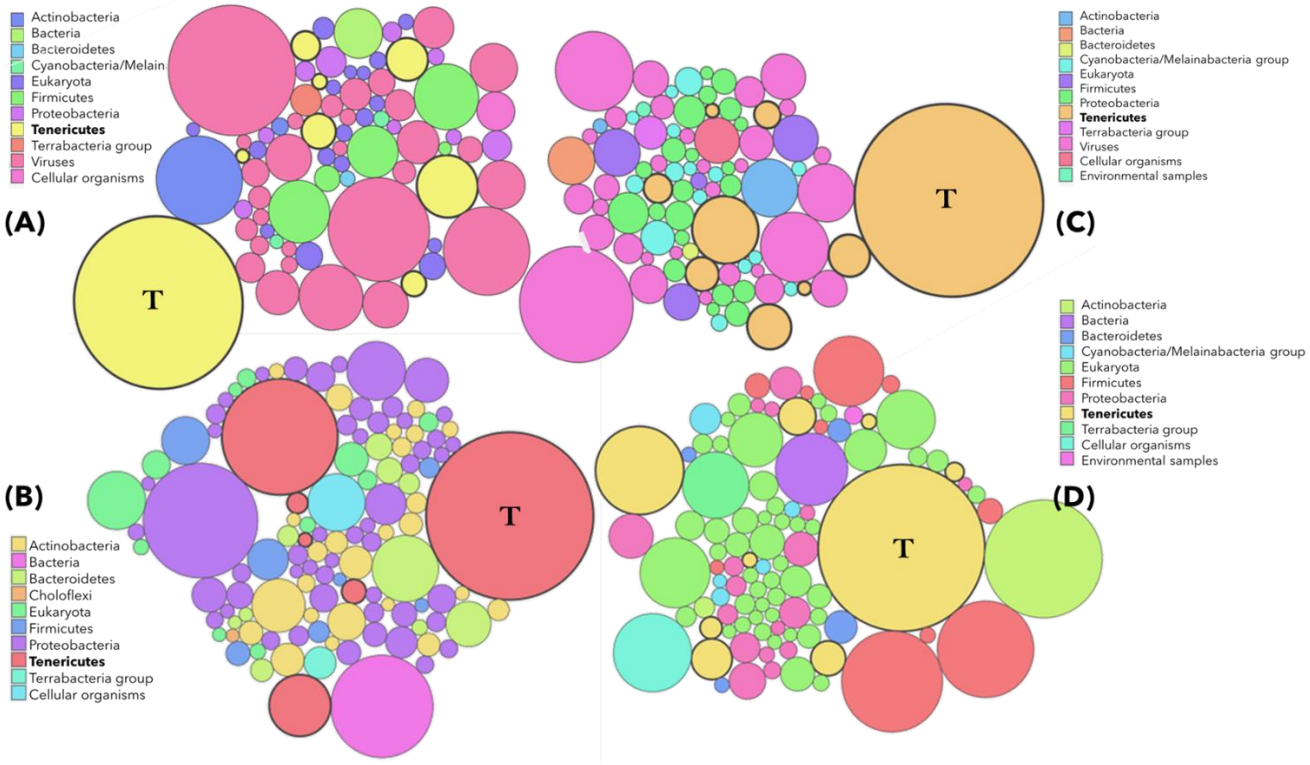


Figure 1. Bubble plots depicting the number and relative abundance of taxa in the 'enriched' genomic DNA samples were generated using the Kaiju server (https://kaiju.binf.ku.dk). Kaiju plots for the isolates PR08 and PR34 assigning Illumina raw reads, A and C; Oxford Nanopore Technologies raw read, B and D; respectively. The size of each bubble is scaled logarithmically to reflect the number of raw reads directly assigned to the corresponding taxon. Each bubble represents a species, with its diameter representing the relative abundance of taxa in the dataset. 'T' represents *Tenericutes*, indicating the fair abundance of raw reads assigned to phytoplasmas while the scarcity of raw reads assigned to 'cellular organism' or 'Eukaryota' signifies the successful enrichment of 'prokaryotic' DNA.

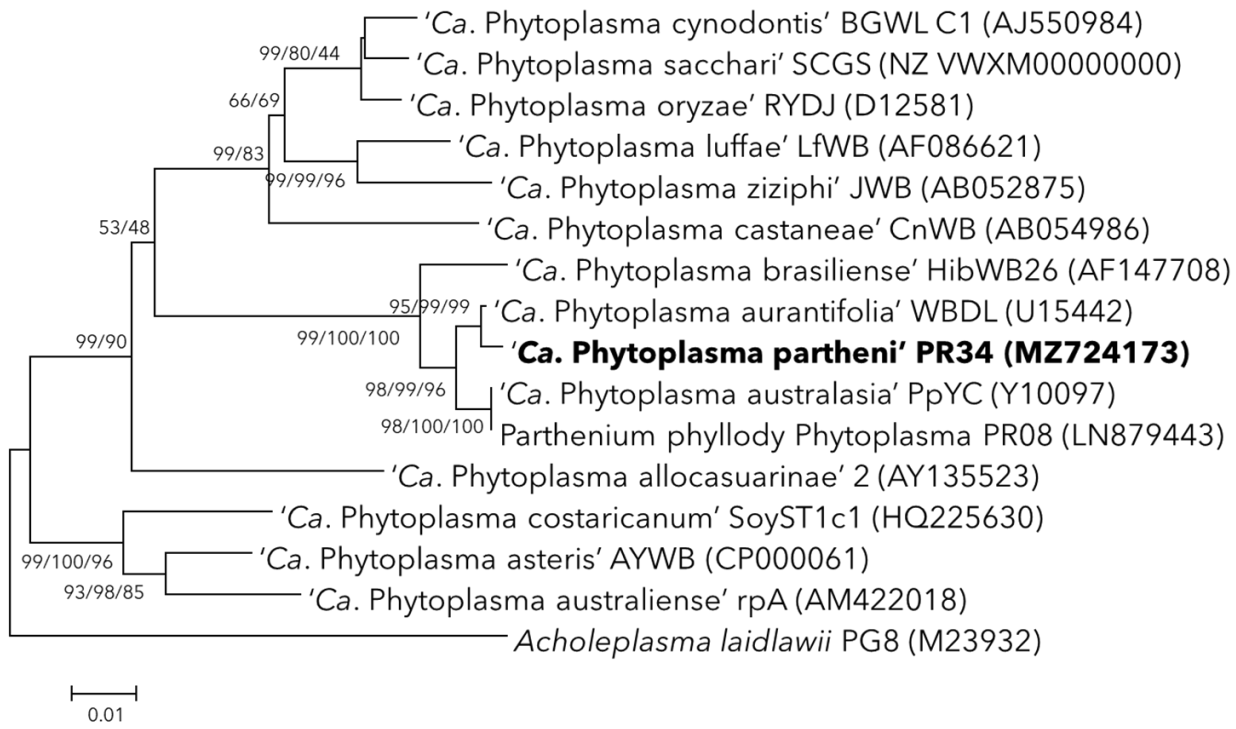


Figure 2. Phylogenetic position of '*Ca*. Phytoplamsa partheni' sp. nov. isolate PR34 inferred from analysis of reference 16S rRNA gene sequences of published provisional species of '*Ca*. Phytoplamsa'. The Neighbour-Joining (NJ), Maximum-Likelihood (ML), and Maximum-Parsimony (MP) methods were employed, utilizing the Tajima-Nei, JTT, and Subtree-Pruning-Regrafting (SPR) models, respectively, in MEGA 7. The topologies of the trees were evaluated by bootstrap analysis based on 1000 replicates. Figures at nodes of the branches indicate the percentage of replicate trees obtained from NJ, ML and MP methods respectively in which the associated taxa clustered together in the bootstrap test. There were a total of 1285 positions in the final dataset. The 16S rRNA sequence of *Acholeplasma laidlawii* PG-8A (M23932) was used as an outgroup. The bar indicates the number of nucleotide substitutions per site.

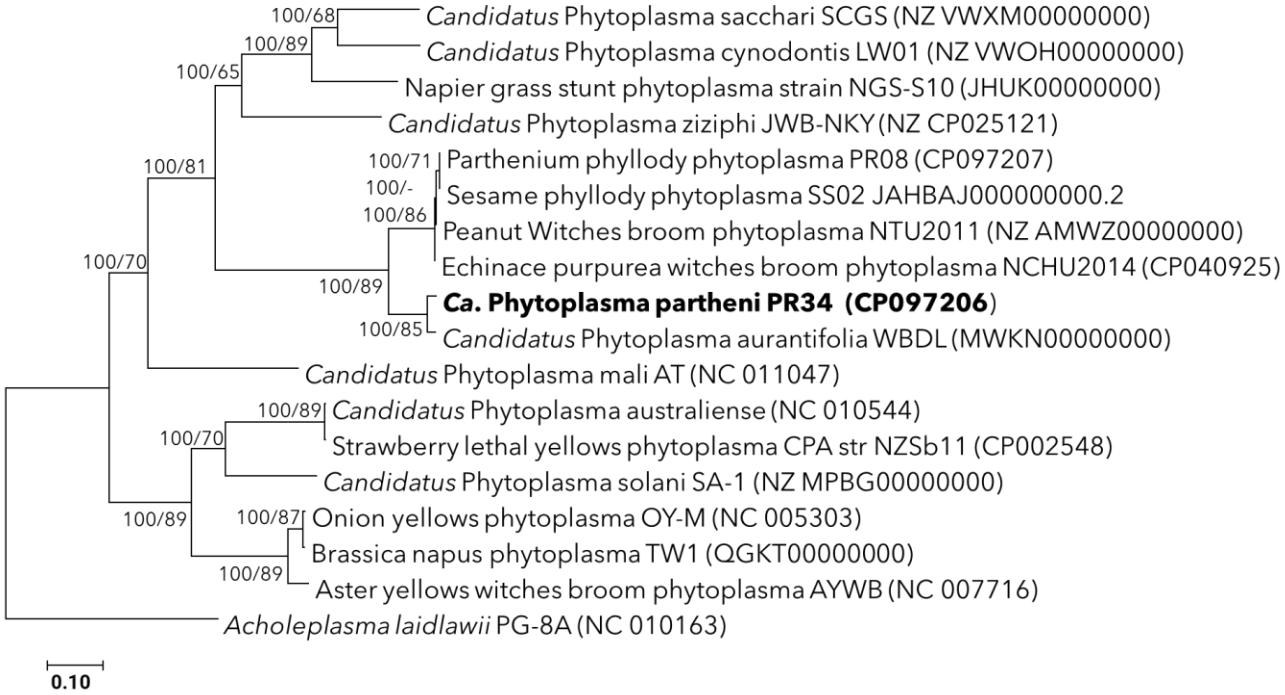


Figure 3. The combined pan-genome phylogenetic tree of phytoplasma is based on orthologous protein and gene sequences derived using BPGA and UBCG tools. The BPGA based phylogenetic tree was constructed in MEGA 7 using the Neighbour-Joining method (1000 bootstrap replicates) involving a concatenated alignment file obtained from the USEARCH Clustering algorithm. The BPGA concatenated alignment file contains a total of 10,387 amino acid positions in the final dataset. The length of UBGC concatenated alignment was 86,628 containing 89 marker genes identified using HMMER and predicted using Prodigal search. Figures at branch points are bootstrap values seen in trees obtained using UBGC and BPGA, respectively. The genome sequence of *Acholeplasma laidlawii* PG-8A (NC 010163) was used as an outgroup. Bar indicates the number of substitutions per site.

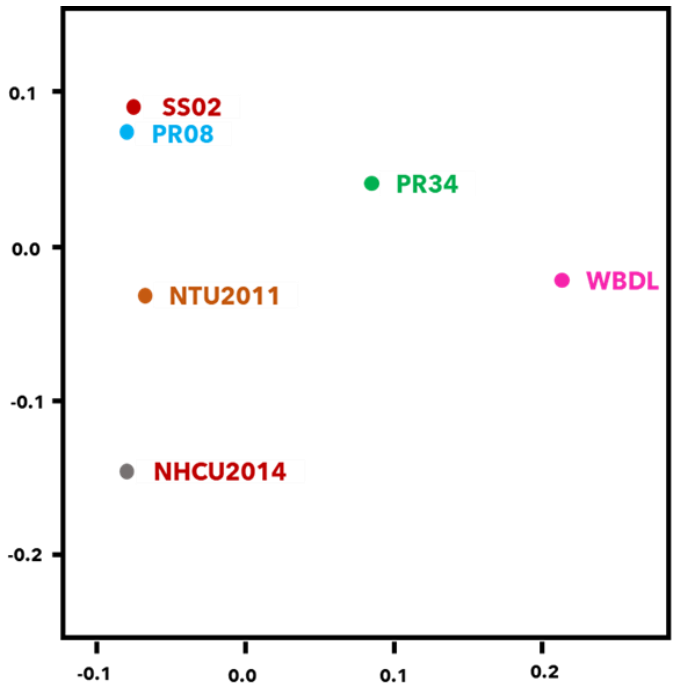


Figure 4. Principal Coordinate Analysis (PCoA) of PWB (16SrII) group of phytoplasma genomes highlighting the distinct position of *'Ca*. Phytoplamsa partheni' PR34. PCoA was performed to compare the gene content among PWB genomes. A matrix consisting of six genomes and 476 gene clusters, obtained from OrthoMCL analysis (Fischer et al., 2011), was transformed into a Jaccard distance matrix using the VEGAN package in R (Dixon, 2003) to assess genome dissimilarity. The Jaccard distance matrix was processed using the PCoA function in the APE package (Paradis and Schliep, 2019). The analysis reveals a significant genetic divergence, with '*Ca*. P. partheni' PR34 exhibiting a distinctive gene composition compared to other PWB genomes.

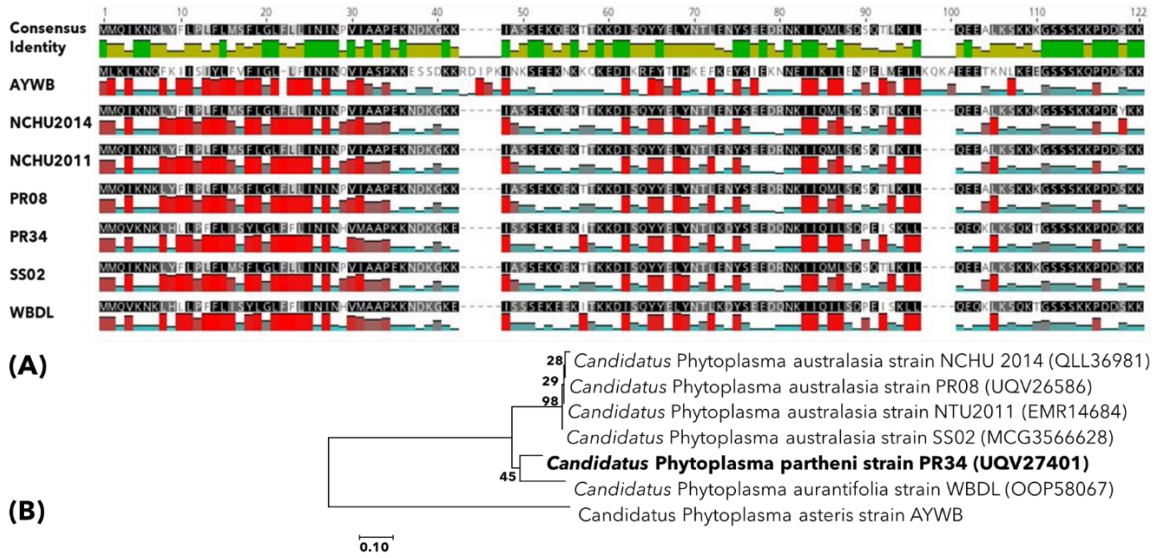


Figure 5. Hydrophobicity pattern and phylogenetic analysis of SAP11 homologs in representative isolates of PWB (16SrII) group of phytoplasmas. The amino acid sequences of SAP11 protein homologs from all PWB genomes were aligned, and their hydrophobicity patterns were visualized using Geneious Prime (2022.1). The hydrophobicity patterns of the SAP11 homologs exhibit a higher degree of conservation unlike their corresponding amino acid sequences (A), The phylogenetic analysis SAP11 homologs of PWB phytoplasmas using the Neighbour-joining method in MEGA7 reveals that the SAP11 homolog of *'Ca*. P. partheni' isolate PR34 has evolved differently from the SAP11 homologs of *'Ca*. P. aurantifolia' and *'Ca*. P. australasia' (B).

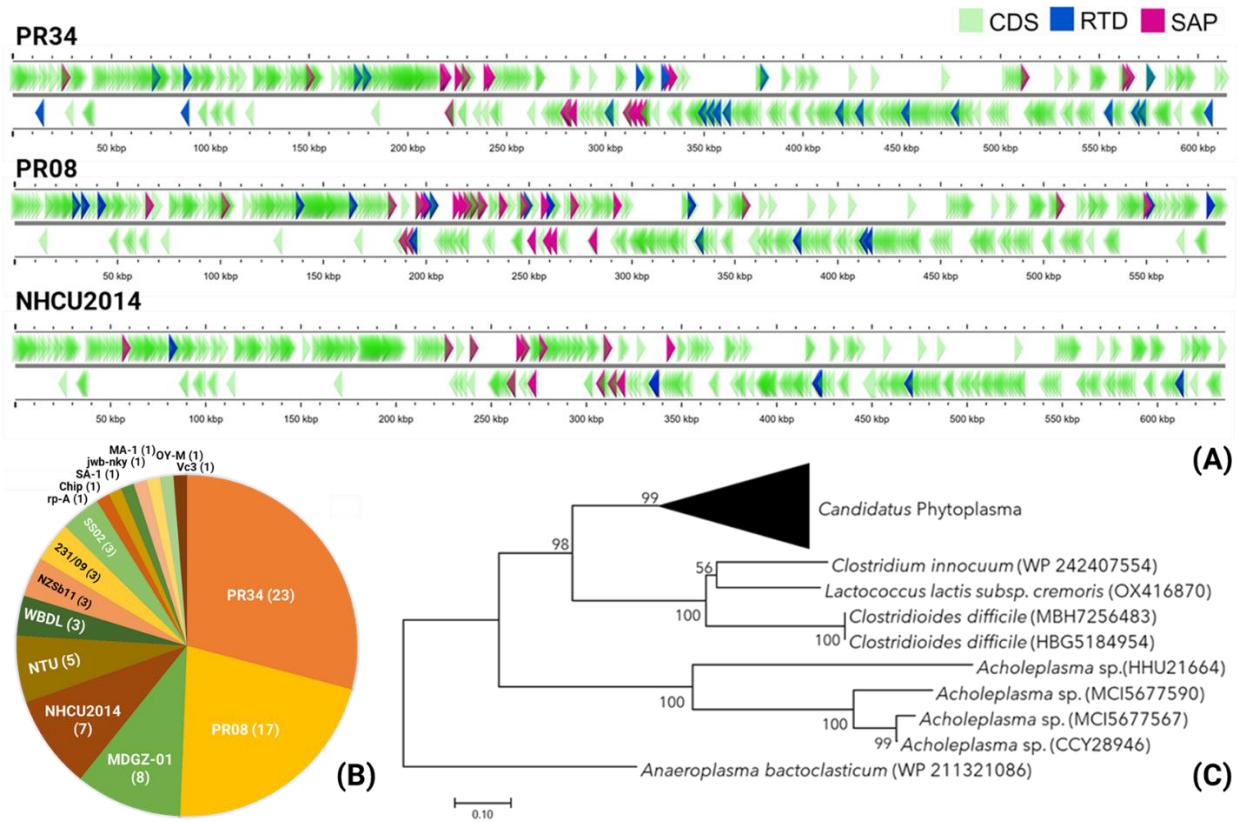
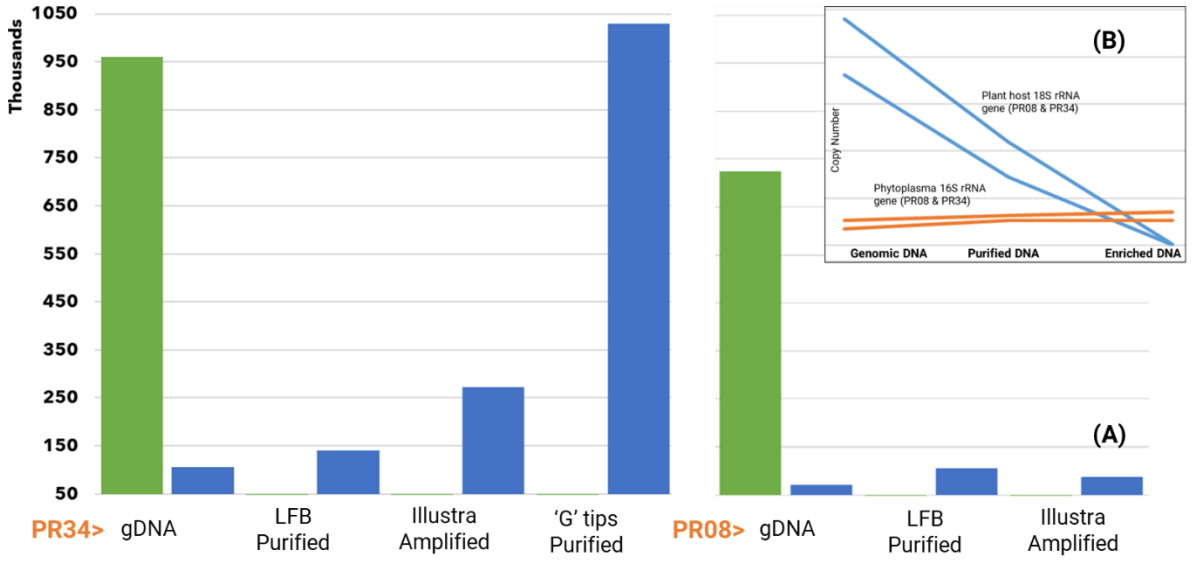


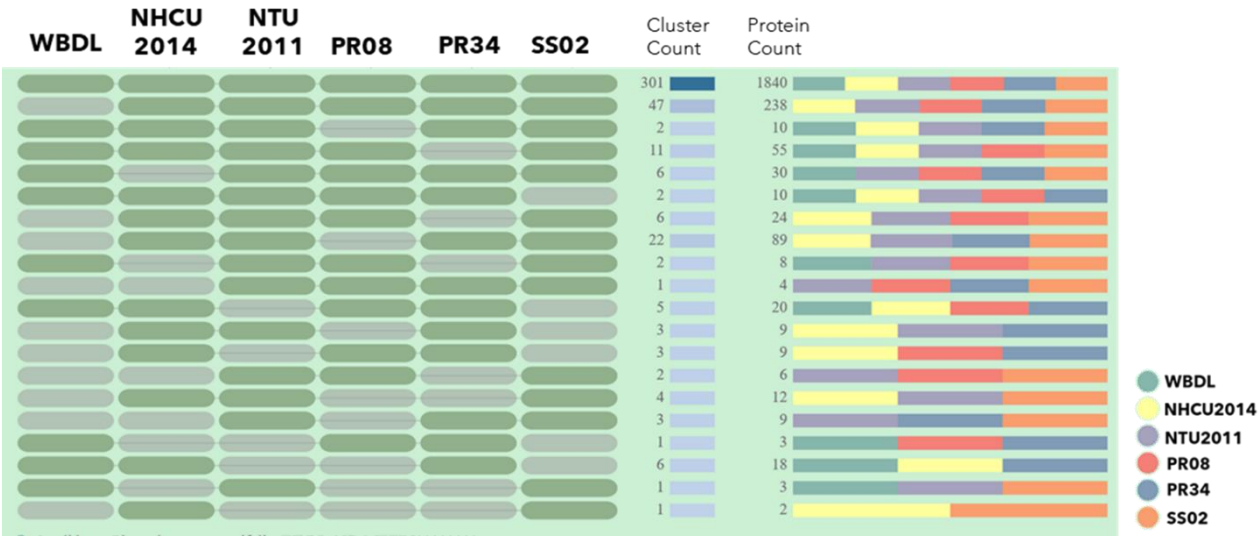
Figure 6. Distribution, abundance, and phylogenetic position of Intron-encoded proteins (IEPs) in complete genomes of PWB (16SrII) phytoplasmas. The figure illustrates the genome-wide distribution of Reverse Transcriptase Domain (RTD) containing proteins (blue), locally distributed secretary proteins (pink), and protein-coding sequences (CDSs, green) in three PWB isolates: PR34, PR08, and NCHU2014 (A). The abundance of RTD sequences as distinct genomic characteristics of PWB phytoplasma genomes, less prevalent in other phytoplasmas (B). Phylogenetic analysis reveals the closer evolutionary relationship of phytoplasmal IEPs to *Clostridium* and *Lactobacillus* species compared to *Acholeplasma*, suggesting horizontal gene transfer as the likely mechanism for IEP acquisition. The evolutionary history was inferred using the Neighbor-Joining method in MEGA7, with bootstrap support shown at nodes. The outgroup sequence of *Anaeroplasma bactoclasticum* (WP 211321086) was used, and the bar indicates the number of substitutions per site. The final dataset comprised 254 positions (C).

Supplementary Table S1. List of phytoplasma isolates -related to proposed '*Ca*. P. partheni', reference isolate PR34 (MZ724173).

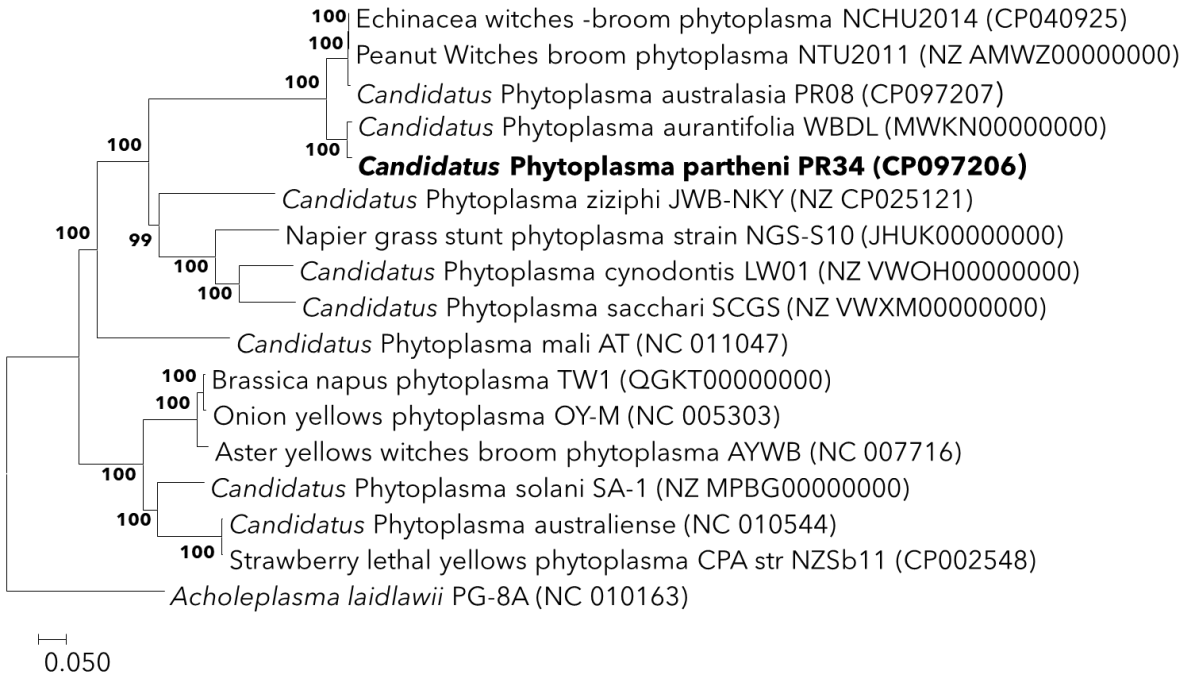
Phytoplasma name and isolate(s)	Accession	Country	Reference
Acacia Witches' broom Tripura1 & Tripura2	MH644006 MH644007	India	Rao et al., 2020
Bamboo Witches' broom BB10	LN811707	India	Yadav et al., 2015b
Bamboo Witches' broom BWB UP1	MZ295214	India	Ravi et al., 2022
Blue snakeweed Witches' broom Gvt	MK603205	India	Pramesh et al., 2020
Cactus Witches'-broom YN08, YN16	AJ293216 EU099561	China	Cai et al., 2008
Chickpea phyllody CPAP-PP2, CPAP-PP6, CPP-P8 & CPP-UP7	MN551488 MN551489 MT420257 MT420258	India	Reddy, 2020
Chickpea phyllody CKKu-29	KX151128	India	Unpublished
Coconut lethal Yellow VCP	JQ868437	N/A	Makarova et al., 2012
Faba bean phyllody FBP	HQ589188	Sudan	Unpublished
Marigold witches'-broom MMWB-ND1 & MMWB-ND2	MW377691 MW377692	India	Unpublished
Parthenium Phyllody PhUP-1	MZ424215	India	Ravi et al., 2022
Parthenium Phyllody PR01	LN811709	India	Yadav et al., 2015b
Parthenium phyllody PR04, PR15, PR35	LN879439MT940957 MZ724174	India	Unpublished
Phyllanthus little leaf PNLL-ND1 & PNLL-ND2	MW377693 MW377694	India	Unpublished
Sesame phyllody CG, Dsh1	KF773145 MW272567	India	Unpublished
Sesame phyllody Kushinagar-3	KF728952	India	Nabi et al., 2015
Sesame phyllody TKG-421 & TKG-431	KF322275 KF322278	India	Unpublished
Soybean phyllody SOYP	EF193353	Thailand	Martini et al., 2007
Soybean witches'-broom PPsoymoz13	HQ840717	Mozambique	Lava Kumar et al., 2011
Tephrosia purpurea phyllody TP02	LN878981	India	Yadav et al., 2014



Supplementary Figure S1. The graph showing the qPCR-determined enrichment efficiency of different methods used in this study. Genomic DNA extracted from symptomatic parthenium plants (PR08 and PR34) underwent various techniques, and the copy numbers of phytoplasma 16S rRNA and host 18S rRNA genes were quantified using quantitative PCR. The copy numbers are represented in green and blue for phytoplasma 16S rRNA gene and host 18S gene, respectively. For PR34 genomic DNA, the first and second bars represent copy numbers obtained using the CTAB method. The third and fourth bars display copy numbers after CTAB extraction followed by processing with LFB (Long Fragment buffer) (SQK-LSK109, Oxford Nanopore Technologies, UK) and NEBNext microbiome-mediated enrichment (E2612L; New England BioLabs, USA). The fifth and sixth bars show copy numbers after additional processing with the illustra ready-to-go V3 amplification kit (GE25-6601-96; Merck, Germany). The fifth and sixth bars show copy numbers after CTAB extraction and purification using QIAGEN Genomic-tips (10223, Qiagen, Germany), followed by NEBNext microbiome-mediated enrichment. Similar DNA preprocessing was carried out for isolates PR08, with the seventh and eighth bars representing copy numbers after processing with LFB, and the ninth and tenth bars representing copy numbers after illustra amplification (A). The line graph demonstrates a decline in copy number for the host 18S rRNA gene compared to a relatively stable copy number of phytoplasma 16S rRNA genes throughout the NEBNext microbiome-mediated enrichment processes for both PR34 and PR08 isolates (inset, B).



Supplementary Figure S2. Comparison of orthologous gene clusters among the genomes of PWB phytoplasmas made using OrthoVenn2 (Xu et al. 2019). The genomes used for this comparison were '*Ca*. P. aurantifolia' WBDL (MWKN00000000), Echinacea purpurea witches' - broom phytoplasma NCHU2014 (CP040925), Peanut Witches' broom NTU2011 (NZ AMWZ00000000), Parthenium phyllody phytoplasma, PR08 (CP097207); '*Ca*. P. partheni' PR34 (CP097206) and 'Sesame phyllody phytoplasma SS02 (JAHBAJ000000000). The occurrence table presents the pattern of top 20 shared orthologous groups among these PWB genomes. The left side of the table indicates the species present in each cluster. The cluster count represents the number of shared clusters between species, while the protein count indicates the number of proteins within these shared clusters. This visual representation provides insights into the variations in protein content across these genomes.



Supplementary Figure S3. Phylogenetic tree of concatenated full-length amino acid sequences of five individual housekeeping genes obtained from selected phytoplasma genomes. The tree involves five amino acid sequences of Replication initiation protein DnaD (*dnaD*), DegV family protein (*degV*), TIGR00282 family metallophosphoesterase, Preprotein translocase SecY (*secY*), and RluA family pseudo uridine synthase (*rluA*) genes totalling to 1132 positions in the final dataset. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model in MEGA7. Figures at nodes of the branches indicate the percentage of replicate trees obtained from ML method. Bootstrap analysis was carried out using 1000 replicates. The respective sequences of *Acholeplasma laidlawii* PG-8A (NC 010163) were used as an outgroup. Bar indicates the number of substitutions per site.