Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

'Candidatus Phytoplasma manihotii' sp. nov. associated with cassava oversprouting disease and the phytoplasma mystery of cassava frogskin disease

Saulo Alves Santos Oliveira

saulo.oliveira@embrapa.br

Brazilian Agricultural Research Corporation

Samar Sheat

Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures

Vanessa Pierote Silva

Federal University of Recôncavo da Bahia

Maria Selma Alves Silva Diamantino

Brazilian Agricultural Research Corporation

Cintia Paula Feitosa Souza

Brazilian Agricultural Research Corporation

Paolo Margaria

Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures

Claudia Fortes Ferreira

Brazilian Agricultural Research Corporation

Stephan Winter

Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures

Article

Keywords: Host-Pathogen Interaction, Metagenomics, Molecular Diagnostics, Phytoplasma Diversity

Posted Date: March 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9033736/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

"*Candidatus* Phytoplasma manihotii" sp. nov. associated with cassava oversprouting Disease and the phytoplasma mystery of cassava frogskin disease

Saulo Alves Santos de Oliveira[1,2*†], Samar Sheat[2†], Vanessa Pierote Silva[3], Maria Selma Alves Silva Diamantino[1], Cintia Paula Feitosa Souza[1], Paolo Margaria[2], Claudia Fortes Ferreira[1], Stephan Winter[2]

[1]Embrapa Mandioca e Fruticultura, Rua Embrapa s/n, CP 007, Chapadinha, 44380000 Cruz das Almas, Bahia, Brazil.

[2]Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Inhoffenstraße 7B, 38124 Braunschweig, Niedersachsen, Germany.

[3]Universidade Federal do Recôncavo da Bahia, Rua Rui Barbosa 710, 44380-000, Cruz das Almas, Bahia, Brazil.

*Corresponding author: Saulo Alves Santos de Oliveira < saulo.oliveira@embrapa.br>

†These authors contributed equally to this work.

Saulo Alves Santos de Oliveira - ORCID: 0000-0002-2992-1570

Vanessa Pierote Silva - ORCID: 0000-0003-1351-3019

Samar Sheat - ORCID: 0000-0002-5649-7795

Maria Selma Alves Silva Diamantino - ORCID: 0000-0002-0938-179X

Cintia Paula Feitosa Souza - ORCID: 0000-0002-3506-7059

Paolo Margaria - ORCID: 0000-0001-8670-1331

Claudia Fortes Ferreira: 000-0001-5476-2236

Stephan Winter – ORCID: 0000-0003-0309-9934

Abstract

Cassava is a key food and energy crop in the tropical and subtropical regions. In South America, oversprouting disease (OVSPD) and cassava frogskin disease (CFSD) occur frequently, however, their etiologies remain unknown. OVSPD is characterized by excessive shoot proliferation, chlorosis, and dwarfism, whereas CFSD largely affects roots, causing cork-like tissues with honeycomb patterns and reduced starch content. In Brazil, phytoplasma organisms were found to be associated with both diseases, with sequences of the 16SrIII-B subgroup phytoplasmas linked to OVSPD and the 16SrIII-L and 16SrIII-A subgroups linked to CFSD. To clarify the identity and taxonomic position of the putative phytoplasma in the diseases, high-throughput sequencing (HTS) of total plant DNA from cassava plants with OVSPD and others with CFSD symptoms was performed, and the complete genome of a novel phytoplasma was assembled from OVSPD plants. In contrast, HTS from CFSD affected plants revealed short sequence fragments of diverse phytoplasmas, however, the assembly of larger scaffolds failed. Genome analyses of two OVSPD phytoplasma assemblies confirmed a novel phytoplasma species within the 16SrIII group. It has an average nucleotide identity (ANI), below the 95% species demarcation threshold relative to the genomes of other phytoplasma species and we propose the name '*Candidatus* Phytoplasma manihotii' sp. nov. A broader screening showed that this phytoplasma occurs only in OVSPD plants, suggesting that the genome fragments detected in CFSD plants may have resulted from other sources, such as genomic integration or horizontal gene transfer.

Keywords: Host-Pathogen Interaction, Metagenomics, Molecular Diagnostics, Phytoplasma Diversity.

Introduction

Cassava (*Manihot esculenta* Crantz) is an essential staple crop for millions of people across tropical and subtropical regions, including Brazil. Vegetatively propagated root crops are exposed to a range of diseases that significantly reduce yield and harvest quality. Cassava diseases associated with phytoplasmas have been reported in Brazil, Costa Rica, Paraguay, and Argentina[1–5], as well as in Southeast Asia, Indonesia, the Philippines, Lao PDR, and Vietnam[6,7]. While these organisms have been mostly identified using partial 16S-rRNA sequences, further verification from independent tests is still pending. Furthermore, the findings of 16S-rRNA sequences specific for phytoplasmas in cassava frogskin, witches' broom, and oversprouting diseases were not

substantiated with biological evidence, and since further pathogens were also identified in diseased plants, the contribution of phytoplasmas to the disease remains unclear.

In cassava frogskin disease (CFSD), a number of diverse viruses from the Secoviridae, Alphaflexiviridae, and Luteoviridae families[8,9] and phytoplasma sequences have been identified. Oliveira et al. (2014)[4] reported the presence of a 16SrIII-L phytoplasma associated with CFSD symptoms in cassava in Brazil, while Souza et al. (2014)[10] identified sequences from the 16SrIII-A phytoplasma subgroup, with the nearest relative being '*Ca*. phytoplasma pruni'. The first report of oversprouting disease (OVSPD) in cassava was by Silberschmidt & Campos in 1944, and it was named "*superbrotamento*." In 2013, Flôres et al. (2013)[3] identified a 16SrIII-B phytoplasma as the causative agent of the disease, and similar to CFSD, the OVSPD-associated phytoplasma was most related to *'Ca.* phytoplasma pruni'. The "cassava witches' broom" disease was first reported in Southeast Asia[7], and phytoplasma sequences from 16SrII and 16SrVIII subgroups are closely related to '*Ca*. phytoplasma luffae'[6] were associated with the disease. CWBD was considered a phytoplasma disease; however, recent studies from Southeast Asia[13,14], French Guiana[11], and Brazil[12] refuted the phytoplasma etiology and proved the fastidious fungus *Rhizoctonia theobromae* (*Ceratobasidium theobromae*) as the causal agent of CWBD.

Interestingly, in English, both OVSPD and CWBD are referred to as "witches' broom," while in Portuguese distinct terms are used; "superbrotamento" for oversprouting symptoms and "vassoura-de-bruxa" for witches' broom symptoms[12]. This differentiation in terminology reflects distinct symptoms and indicates different etiologies. OVSPD is characterized by excessive shoot proliferation, stunted plant growth, and leaf malformation, leading to significant losses in cassava production, particularly in regions such as Northeast Brazil, where root yield reductions can reach 90%. Symptoms of witches' broom, on the other hand, include dwarfism, proliferation of weak spindly shoots, shortened internodes, and most prominently, vascular necrosis as the key differentiating feature (Fig. 1). The early symptoms of these diseases are similar, but they differ significantly in terms of disease progression and severity. Oversprouting is characterized by excessive shoot proliferation and stunted growth, whereas witches' broom-like leaf arrangements on stems and vascular necrosis, leading to plant decline, are characteristic features of CBSD[12]. In contrast, CFSD on aboveground plant parts is inconspicuous and only becomes obvious at harvest, when the woody, cork-like tuberous roots with a honeycomb pattern of fissures[1,8,9] provide

evidence of the disease. Recently, Jimenez et al. (2024)[15] provided evidence that CFSD is caused by a cassava torrado-like virus 2, without considering a role of phytoplasma in the disease complex.



Fig. 1. Typical disease symptoms of cassava plants from which samples were collected. (A) Frog skin disease symptoms on roots from Bahia state (B) Shoot proliferation (oversprouting) in the field – in Ceará State (C) Witches' broom on leaf and stems in the field in Amapá state.

To resolve this conundrum of phytoplasmas associated with OVSPD and CFSD, in our research, high-throughput sequencing (HTS) of total DNA of diseased plants to search for pathogen sequences without *a priori* knowledge of the target organism[16], was used. Following HTS, genome assemblies were created and genome drafts were generated for two phytoplasma isolates associated with cassava oversprouting disease in Southern Ceará, Brazil. The genomic differences between these phytoplasma isolates warrant taxonomic assignment to a novel species, for which we propose '*Candidatus* Phytoplasma manihotii' sp. nov. causing cassava oversprouting disease in Brazil.

Results

Cassava samples with frogskin disease and oversprouting symptoms were analyzed using HTS and metagenome analysis to reveal the microbial composition associated with these diseases. The Venn diagram of the microbial communities associated with the samples collected from different regions indicated that the diversity of microbial species associated with CFSD (n= 319) was greater than that associated with OVSPL-leaves (n= 105) and OVSPS-stems (n= 55) (Table 4). Nine organism groups were shared across all conditions, including *Fusarium*, *Pseudomonas*,

Streptomyces, *Alternaria*, *Mycobacterium*, *Trichoderma*, *Colletotrichum*, *Aspergillus* and *Diaporthe*. Plant samples from oversprouting diseased cassava leaves (OVSPL) and stems (OVSPS) shared twelve families, among which, Achoplasmatacea, comprised the genus "*Candidatus* Phytoplasma" (Fig. 2A).

There was a predominance of distinct organism families. In roots of CFSD cassava, Streptomytaceae, Sphingomonadaceae and Nocardiaceae and a high number of unidentified taxonomic units, were found. Acholeplasmataceae was the most abundant family in symptomatic leaves from OVSPD plants, while in stems, the most abundant families were Caulobacteraceae and Saccotheciaceae (Fig. 2B).

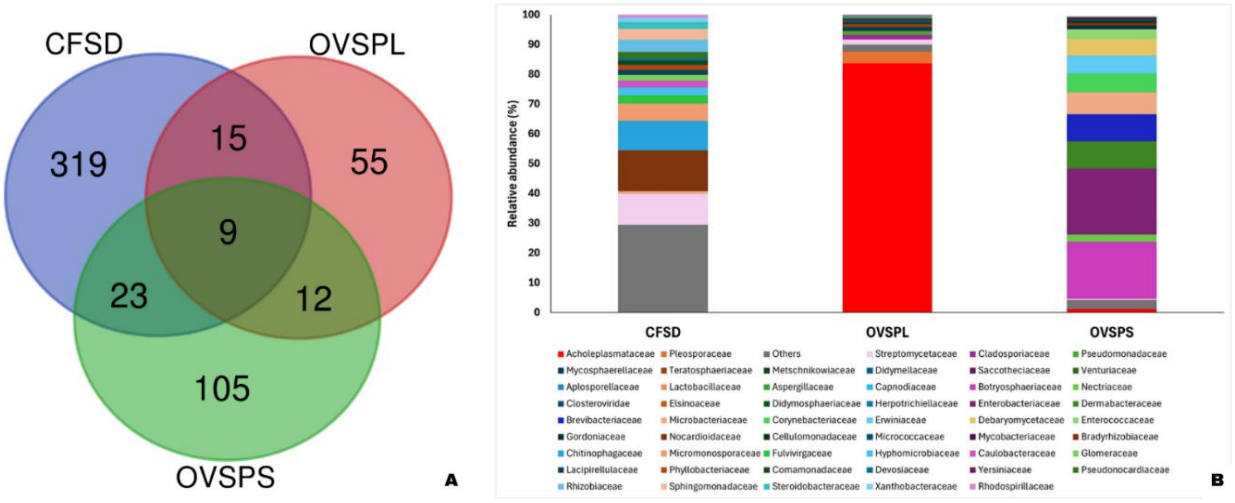


Fig. 2. (A) Venn diagram showing unique and shared microbial groups between CFSD, OVSPL, and OVSPS. (B) Relative abundance of major microbial families in each condition.

Whole genome assembly

The HTS Illumina sequencing results for CFSD, OVSP-L01 (Genbank n°: SAMN52870867), and OVSP-S01 (Genbank n°: SAMN52869104) are presented in Table 1. The total number of reads ranged from 48.8 to 62.6 million across all samples, amounting to approximately 29.6 to 33.3 million reads after normalization. Mapping the reads to the cassava reference sequence showed varying percentages, with the CFSD sample exhibiting the highest percentage of mapped reads (84.91%), while the OVSP-L01 and OVSP-S01 samples presented values of 70.76% and 78.78%, respectively.

Assembly analysis revealed that the CFSD sample produced the largest number of contigs (6,020), followed by OVSP-L01 (5,250) and OVSP-S01 (4,975), with a maximum read length

reached for OVSP-S01 of 53,407 bp. A high number of contigs in OVSP-L01 were assigned to phytoplasma sequences; in contrast, only twenty contigs were assigned to putative phytoplasma sequences in CFSD.

16S phylogeny and 'in silico' RFLP classification

Phytoplasma 16S rDNA sequences extracted from HTS sequences from two pools (OVSP-L01 and OVSP-S01) obtained from 18 pooled cassava samples were compared with phytoplasma 16SrDNA sequences available at NCBI/GenBank (Supplementary Table S1), including a reference sequence for a *Manihot esculenta* oversprouting phytoplasma (GU193976, accessed on February 02, 2025). Supplementary Table S1 contains detailed information on the phytoplasma strains, including their geographic location, diseases associated with each strain, and year of collection.

The 16S rDNA sequences of OWBL and OWBS clustered in the same Sr III clade (Fig. 3) as phytoplasmas associated with "Cassava Witches' Broom Disease" (16SrIII-L) and *Manihot esculenta* Witches' Broom (16SrIII-B) Supplementary Fig. S1 presents a dendrogram based on the 16S rDNA gene, constructed using the UPGMA method.

The 16S rDNA sequences of OVSP S01 and OVSP L01 samples grouped within the 16SrIII clade, close to the 16SrIII-L subgroup, also including sequences from the cassava witches' broom disease phytoplasma and *Manihot esculenta* oversprouting phytoplasma.

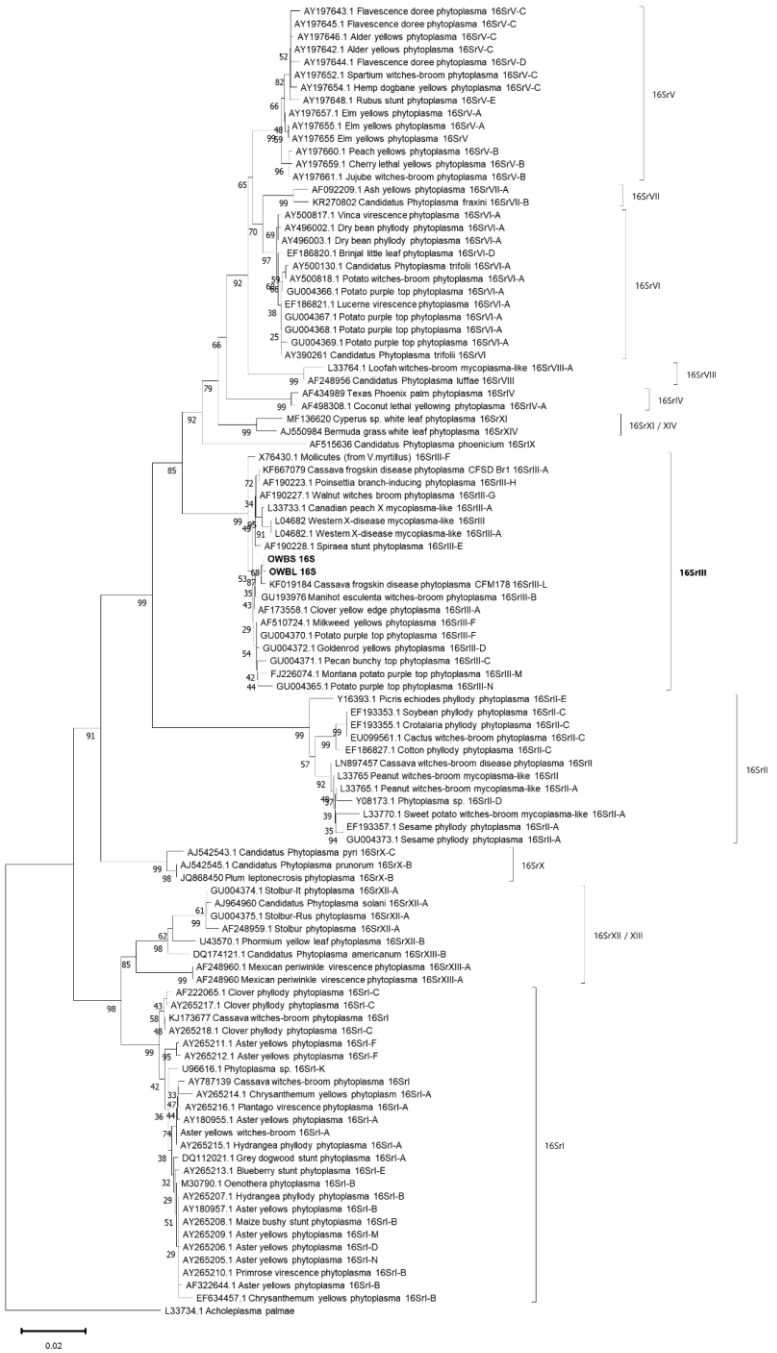


Fig. 3. Phylogenetic tree inferred from 16S rRNA gene sequences (1,515 nucleotides) using the Maximum Likelihood method under the GTR+G model. The dataset comprised sequences retrieved from the NCBI, with *Acholeplasma palmae* included as an outgroup. Bootstrap values (1,000 replicates) are shown next to the branches, and branch lengths (substitutions per site) are indicated by the scale bar at the bottom of the figure. Cassava oversprouting phytoplasma isolates (OVSP-S01 and OVSP-L01) are highlighted in bold.

secY based phylogeny

Phylogenetic analysis of cassava phytoplasma sequences based on *secY* gene sequences available at NCBI/Genbank (suppl. Table S2) placed OWBL *secY* and OWBS *secY* together with the *Manihot esculenta phytoplasma* sequence LC228771 reported from cassava in Japan (Supplementary Fig. S2).

Multi-locus sequenMulti-locuss (MLSA)

Multi-locus analysis (Supplementary Fig. S3) with concatemeric 16S rDNA and *secY* sequences confirmed the grouping enclosing '*Ca.* Phytoplasma pruni*'* to form a distinct clade of closely related sequences (Supplementary Fig. S3).

Whole genome comparison

A comparison between the newly assembled OVSP-S01 and OVSP-L01 genomes in relation to *'Ca.* Phytoplasma pruni' is shown in Fig. 4. The reference genome of the phytoplasma is represented in purple, whereas the genes from the stem (A) and leaf (B) samples are shown in orange, indicating the direction of the sequences. Black lines indicate GC content. The sequencing data for OVSP-S01 and OVSP-L01 are presented in Table 2.

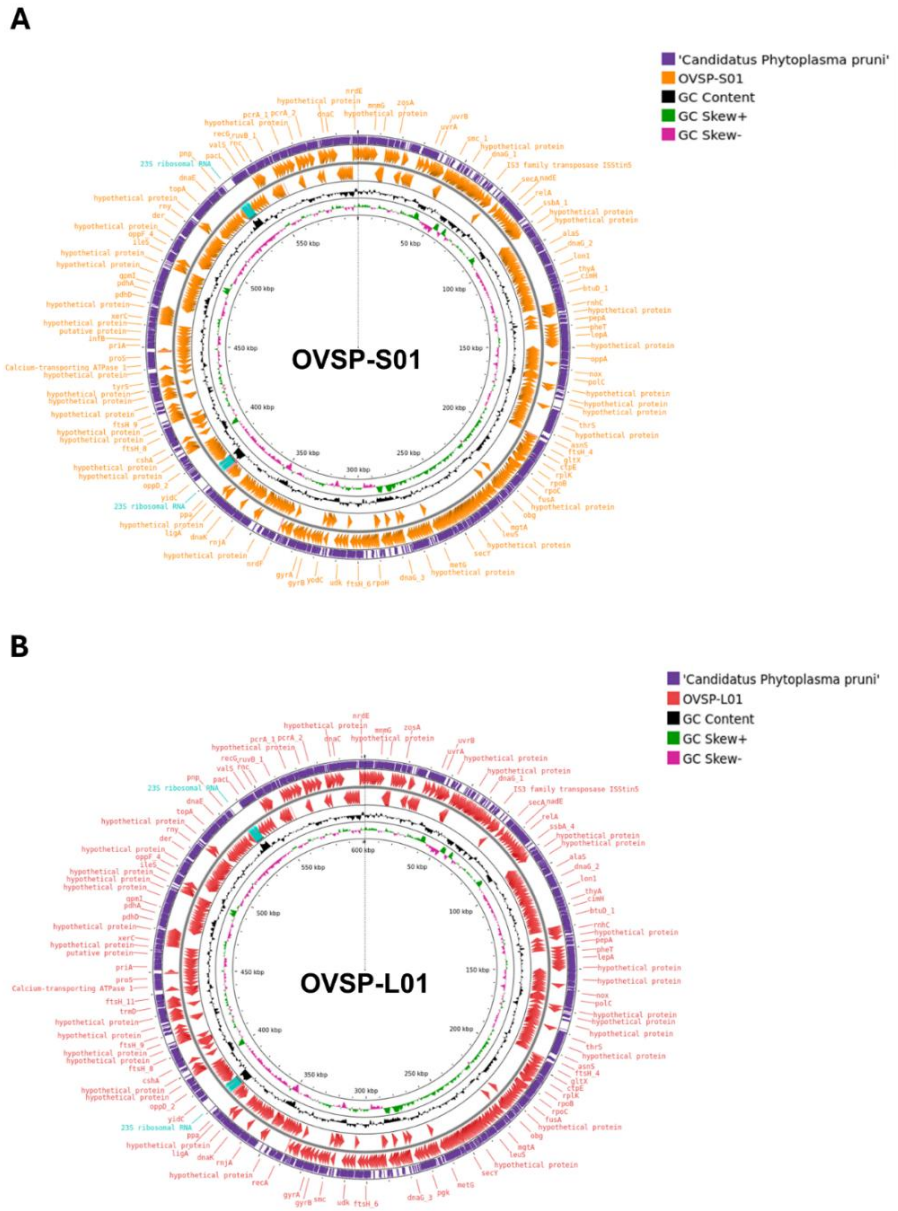


Fig. 4. Circular genomic representation of cassava oversprouting phytoplasma isolates OVSP-S01 (A) and OVSP-L01 (B). Coding sequences (CDSs), rRNA genes, and tRNA genes were identified using PROKSEE on the CGView Server with Prokka annotation, displayed on the outermost circle. The outer ring corresponds to the genomic representation of *'Candidatus Phytoplasma pruni'* (purple), followed by the genome representations of OVSP-S01 (orange) and OVSP-L01 (red), which were assigned to *'Candidatus* Phytoplasma manihotii' sp. nov. The inner rings depict GC content (black) and GC skew, with positive values in green (GC skew+) and negative values in pink (GC skew-).

Although no complete phytoplasma genome was identified in the "frog skin" samples, sequence fragments with similarities to phytoplasmas were detected, and these partial sequences are depicted in figure 4A and B as light blue blue fragments. These sequence fragments may result from processes such as DNA transfer, the presence of related bacteria, and genomic integration into cassava DNA, but highlight the lack of further evidence for phytoplasmas in the "frog skin" samples (Supplementary Fig. S4).

Fig. 5 shows a dendrogram based on the average nucleotide identity (ANI) percentage. The red dotted line marks the 95% threshold commonly used for species delimitation. The OVSP-S01 and OVSP-L01 isolates clustered together, indicating high similarity, but fell below the 95% threshold established by Bertaccini et al. 2022[19] when compared to '*Ca*. Phytoplasma pruni and other species in the dendrogram. This strongly supports the classification of these isolates as belonging to a distinct species. Based on these findings a new species is proposed to be the most likely pathogen involved in the oversprouting symptoms, here named as *'Candidatus* Phytoplasma manihotii' sp. nov.

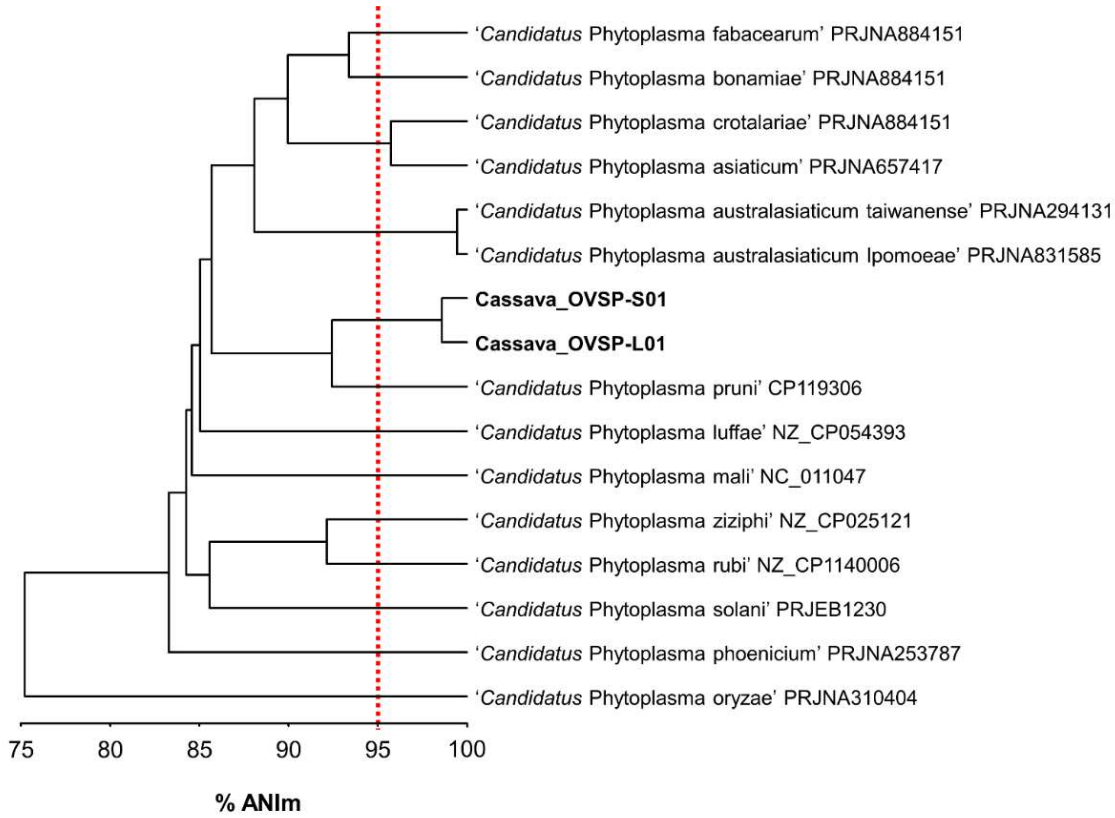


Fig. 5. UPGMA dendrogram of phytoplasma strains associated with oversprouting disease and other '*Candidatus* Phytoplasma' species references using whole-genome comparisons based on the 'ANIm' values. The 'ANIm' was calculated using the percentage estimated for the Average Nucleotide Identity based on MUMmer. Red line shows the boundary for species differentiation (ANIm < 95%) based on the species description´s guidelines established by Bertaccini et al (2022)[19].

Therefore, comparative genomic analysis showed that '*Ca.* P. manihotii' shares strong synteny and overall gene content with *'Ca. Phytoplasma pruni,',* the closest known relative. Both species harbor conserved informational modules, such as ribosomal proteins, chaperone systems, and Sec secretion machinery, but differ in the arrangement of several hypothetical protein clusters and the apparent absence of a complete potential mobile unit (PMU), as seen in *'Ca. P. mali'* (Kube et al. 2008). Genes encoding *dnaB*, *dnaG*, *tmk*, *ssbA*, and *ftsH/hflB*, were detected in both *"Ca.* P. manihotii' genomes, suggesting the presence of vestigial or fragmented mobile elements rather than an intact PMU.

Prediction of putative secreted effectors

Potential pathogenic effectors were identified using an integrated approach. BLASTp searches against the NCBI nr database revealed that all predicted effectors shared high similarity to hypothetical proteins of closely related phytoplasmas, particularly *'Ca. Phytoplasma pruni'* and *'Conium maculatum' witches'- broom phytoplasma*. The strongest matches included proteins with 96–99% identity to *'Ca. P. pruni'* hypothetical proteins (e.g., WP_244948362.1, WP_178734048.1, W90%734413.1) and > 90% identity to *'C. maculatum'* witches'-broom phytoplasma proteins (e.g., MEC4558719.1, MEC4558753.1, MEC4558770.1). None of the sequences displayed significant similarity to the canonical phytoplasma effectors SAP05, SAP11, SAP54, or TENGU, underscoring the distinct effector repertoire of *'Ca. P. manihotii'*.

These findings indicate that *'Ca. Phytoplasma manihotii'* retains a functional secretion system but employs a unique set of small secreted proteins likely responsible for its characteristic witches'- broom symptoms in cassava. Further experimental validation is required to elucidate their host-target interactions and contributions to pathogenicity.

Discussion

The unclear etiology of CFSD and OVSPD and their association with phytoplasmas prompted further investigation. In Brazil, Oliveira et al. (2014)[4] associated CFSD with subgroup 16SrIII-L phytoplasmas, whereas Souza et al. (2014)[10] linked phytoplasma sequences from subgroup 16SrIII-A to the disease. Only recently, Jimenez et al. (2024)[15] reported that CFSD in Colombia was caused by a single torradovirus infection with other viruses or phytoplasma organisms not associated with the disease. Nevertheless, the causes of CFSD remain unclear. OVSPD in Brazil was associated with 16SrIII phytoplasmas (Flores et al. (2013)[3] while 16SrI and 16SrII and 16SrVIII phytoplasma sequences were found in cassava from Southeast Asia[6] further adding to confusion on the causal association of phytoplasmas to the diseases. The notion of phytoplasmas associated with both witches' broom / oversprouting diseases may have been driven by the association of symptoms with short 16S rRNA sequence fragments amplified from those plants.

In our HTS study, we assembled complete phytoplasma genomes from OVSPD diseased plants that were exclusively found symptomatic samples. Phylogenetic analysis based on the 16S gene to compare with other phytoplasmas confirmed that OVSP-S01 and OVSP-L01 isolates belonged to the 16SrIII group, specifically clustering within the 16SrIII-B/L subgroup. The *in silico* restriction fragment length polymorphism (RFLP) analysis supported this classification and furthermore differentiated OVSP-S01 and OVSP-L01 from other known phytoplasmas including those previously associated with both diseases, OVSPD and CFSD.

The *secY* gene provides a higher resolution for phytoplasma classification and the phylogenetic reconstruction using the *secY* gene corroborated the results placing OVSP-S01 and OVSP-L01 within the 16SrIII-L subgroup. Sec translocation system genes (*secA, secY, and secE*), chaperone operons (*dnaK–dnaJ–grpE* and *groEL–groES*), and extensive sets of ribosomal proteins (Rpl/Rps), along with full rRNA and tRNA complements were identified in the genomes of both phytoplasma isolates. Partial CDS were found near the contig termini, indicating minor fragmentation but no major gene loss indicating that isolates represent near-complete genomes of '*Ca.* P. manihotii.' Multi-locus sequence analysis (MLSA), which integrates multiple genetic markers, further enhances the resolution of phytoplasma classification and confirmed that OVSP-S01 and OVSP-L01 are genetically distinct from other 16SrIII phytoplasmas and represent novel species.

We failed to reconstruct phytoplasma genomes from nucleic acid extracts of CFSD cassava showing only, "the commonly found" sequences attributed to phytoplasma-like sequences. The diversity of those sequences also raises questions concerning genome integration, and environmental contamination rather than active infection. Thus, it is safe to state that phytoplasmas are not implicated in CFSD however, considering the diversity of complete virus genomes found in CFSD, their involvement and contribution to the disease is still debatable despite the ground breaking work of Jimenez et al., 2024.

The average nucleotide identity (ANI) analysis highlighted significant genetic divergence between OVSP-S01, OVSP-L01, and other described phytoplasma species. Both isolates fell below the 95% ANI threshold used for species delimitation when compared to *'Ca.* Phytoplasma pruni' and other 16SrIII phytoplasmas, as postulated by Bertaccini et al. (2022)[19]. This supports the taxonomic assignment of OVSP-S01 and OVSP-L01 to distinct species within the 16SrIII group.

We can conclude from our study that OVSPD and CFSD are indeed distinct entities based on their unique symptoms and their pathogens. The presence of phytoplasmas that were exclusively present in oversprouting samples with no other sequence present and the phylogenetic and genomic differentiation, classifies '*Ca.* Phytoplasma manihotii' sp. nov. as a novel species within the 16SrIII group. The failure to retrieve more than short sequences to reconstruct complete genomes from CFSD samples falsified an association of phytoplasmas with the disease. It further emphasizes the power of HTS to exclude a potential pathogen with low uncertainty. Future studies should focus on biological aspects of OVSPD, transmission and spread paramenters and the pathogenic mechanisms of '*Candidatus* Phytoplasma manihotii' sp. nov.

Material and Methods

Surveys were conducted in 2023 to scout for symptoms of shoot proliferation and witches` broom in cassava-growing regions in the States of Bahia and Ceará, Brazil (Table 4 and Fig. 6). Samples of plants with CFSD, OVSPD and witches' broom symptoms are shown in Fig. 1 (A-C). Samples with OVSPD were collected from cassava planting areas in the municipality of Laje (1 sample), a city located in the rural area of the State of Bahia, and in three rural locations in the State of Ceará: Serra do Salitre (5 samples), Serra dos Estevam (2 samples), and Chapada Pitombeira (1 sample). Additionally, in the city of Cruz das Almas, Bahia, samples of plants with typical symptoms of CFSD (Fig. 3C), were also collected. All samples were taken to the Molecular

Biology Laboratory at Embrapa Mandioca e Fruticultura – Cruz das Almas/BA, flash frozen in liquid nitrogen, transferred to a lyophilizer (72 h at 54 µHg and -53º C), and then stored at -80 °C prior to shipment to DSMZ Germany for further High-Throughput Sequencing (HTS) analysis.

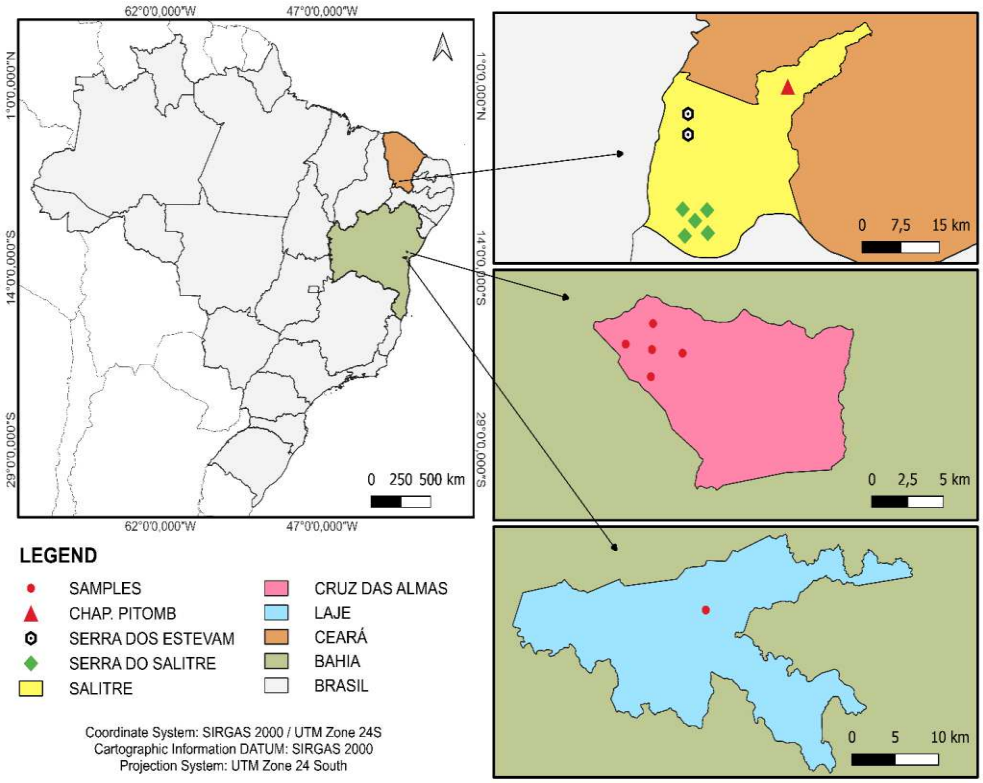


Fig. 6. Map showing the distribution of samples collected in the municipalities of Cruz das Almas and Laje – BA, and Salitre – CE. Dots indicate the collection site.

The cassava symptoms assigned to the diseases included root symptoms of frog skin disease, shoot proliferation, bushy stunt, and internodal shortening for oversprouting and witches' broom disease.

Nucleic acid isolation, library preparation and HTS

All nucleic acid isolation, library preparation, and HTS experiments were conducted at the Plant Virus Department, German Collection of Microorganisms and Cell Cultures (*Deutsche Sammlung von Mikroorganismen und Zellkulturen* -DSMZ, Germany). In the present study, two distinct methodologies were employed for the detection of DNA- and RNA-based pathogens in cassava, utilizing high-throughput sequencing (HTS) to facilitate comprehensive surveillance.

Both approaches were optimized for the specific nature of the nucleic acid under investigation, ensuring robust pathogen detection across a range of potential viral and microbial contaminants.

Total DNA was extracted from each sample using NucleoSpin® Plant II (MACHEREY-NAGEL GmbH & Co. KG) kit, following the manufacturer's protocol. Approximately 150 mg of plant tissue was placed in 2 mL microtubes containing two stainless steel beads (one small bead placed over a larger bead). The tubes were then immersed in liquid nitrogen to ensure complete freezing of the samples. Once frozen, the samples were ground into a fine powder using a TissueLyser LT at 50 rpm for 4 min. All subsequent steps followed the "Genomic DNA from Plant" protocol provided by the NucleoSpin® Plant II kit. DNA was eluted in 100 µL of elution buffer and assessed by electrophoresis on a 1% agarose gel. The samples were then stored at −20 °C for future use.

For RNA extraction, approximately 150 mg of plant tissue was placed in 2 mL microtubes containing two stainless steel beads (one small and one large). The tubes were immersed in liquid nitrogen to freeze the samples completely. Once frozen, the samples were ground into a fine powder using a TissueLyser LT at 50 rpm for 4 min. Total RNA was isolated using the GenCatch™ Total RNA Extraction Kit (Epoch Life Science, Inc.) following the manufacturer's protocol. Cell lysis was performed using PRX buffer, and RNA was eluted in 50 µL of RNase-free deionized water. RNA quantity and quality were assessed using a spectrophotometer and gel electrophoresis on a 1% agarose gel. Extracted RNA samples were stored at −20 °C until further use. Prior to library preparation, equimolar amounts of RNA and DNA from individual sample preparations (Table 1) were combined into pools. DNA and RNA concentrations were determined using NanoDrop and Qubit spectrophotometry.

RNA-based pathogen detection involves the synthesis of complementary DNA (cDNA) from total RNA. First, ribosomal RNA was depleted using Qiaseq FastSelect -rRNA Plant Kit (Qiagen), followed by reverse transcription with Maxima H Minus Reverse Transcriptase (Thermo Scientific), RiboLock RNase Inhibitor (Thermo Scientific), dNTPs (Invitrogen,), an Octamer Primer (Eurofins), and RNase H (Invitrogen). Second-strand cDNA synthesis was subsequently performed using the NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (NEB), followed by purification using the NucleoSpin Gel and PCR Cleanup kit (Macherey-Nagel). The generated libraries were quantified using a DNA kit on an Agilent 2100 Bioanalyzer according to the manufacturer's instructions.

Library preparation for both DNA and RNA samples was performed using the Illumina DNA Prep Kit (FlexM), coupled with IDT® for Illumina® DNA/RNA Unique Dual (UD) Indexes. Sequencing was conducted on a NextSeq 2000 instrument using paired-end reads (2 × 150 nt with P2 Reagents v3, 300 cycles). Quality control steps included adapter trimming, error correction, and normalization. Processed reads were then mapped against the *Manihot esculenta* reference genome {(National Center for Biotechnology Information, 2021 – *Manihot esculenta* v8.1, NCBI RefSeq assembly GCF_001659605.2, Annotation Release 101) [cited 2025 Mar 10]. Available from: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001659605.2} with unmapped reads retained for *de novo* assembly using Geneious Prime (v.2024.0; Biomatters).

These complementary approaches—one based on direct genomic DNA extraction and the other on reverse-transcribed RNA–provide a comprehensive framework for detecting both DNA and RNA pathogens in cassava, leveraging HTS technologies and advanced bioinformatics tools to enable the detection of a broad spectrum of potential pathogens in plant tissues.

Nanopore libraries were prepared following the protocols provided by Oxford Nanopore Technologies (ONT, UK) and sequenced using the MinION platform. Raw reads were base-called and demultiplexed using Guppy (v6.0.1), followed by quality control with 'Porechop' for adapter trimming, error correction, and normalization. The reads were mapped against the *M. esculenta* reference genome (NCBI RefSeq assembly GCF_001659605.2) using 'Minimap2', with unmapped reads retained for further downstream analyses. For a general overview, the *de novo* contig assemblies from Illumina sequencing and nanopore reads were analyzed using the DIAMOND BLASTx tool to identify potential homologs of the predicted proteins. The resulting BLASTx output was then imported into MEGAN (MEtaGenome ANalyzer) to obtain a metagenomic overview of species and assign taxonomic identities to the contigs. This provided insight into the diversity and composition of the species present and enabled a comprehensive analysis of putative protein-coding sequences. To verify species assignment, trimmed reads and *de novo* assembled contigs were subjected to a BLAST search against a custom reference database (constructed in Geneious Prime v. 2022.0.1) created for each putative pathogenic species identified.

16S phylogeny and 'in silico' RFLP classification

The two 16S rDNA regions (rRNA operons rrnA and rrnB) were aligned with sequences from various *Candidatus Phytoplasma* species using MUSCLE, implemented in MEGA11[23]. The

alignment was compared among the different phytoplasma sequences and with representatives of the '16S subgroups.' A Maximum Likelihood phylogenetic tree was constructed to illustrate the relationships between sequences from cassava and other phytoplasma groups.

Additionally, *in silico* restriction fragment length polymorphism (RFLP) analysis and similarity coefficient (F) comparisons were performed for both rRNA operons (rrnA and rrnB) using the iPhyClassifier tool[24] to support classification and group and subgroup placement. Similarly, RFLP analyses were simulated using 16S rRNA region sequences from type specimens belonging to the phytoplasma groups[25]. The sequences were aligned using MUSCLE and subsequently restricted in silico using the pDRAW 32 tool (REF) to obtain cleavage patterns for the tested restriction enzymes.

Seventeen restriction enzymes were evaluated (*Alu*I, *Bam*HI, *Bfa*I, *Bst*KTI, *Bst*UI, *Dra*I, *Eco*RI, *Hae*II, *Hha*I, *Hinf*I, *Hpa*I, *Hpa*II, *Kpn*I, *Mse*I, *Rsa*I, *Ssp*I and *Taq*I). After *in silico* digestion, virtual 2% agarose gel images were generated using pDRAW 32 program[26]. The resulting bands were analyzed using UPGMA clustering based on Euclidean genetic distance, and cluster analysis was performed in R software[27] using the '*tidyverse', 'ape',* and *'ggdendro'* packages, allowing the construction of a dendrogram to visualize the relationships between the restriction profiles of different phytoplasma isolates.

secY based phylogeny

For phylogenetic analysis based on the *secY* gene[25], sequences were extracted from the HTS assemblies and aligned using MUSCLE in MEGA 11. This alignment served as the basis for establishing evolutionary relationships. Different evolutionary models were tested, and the most appropriate model was selected based on the likelihood criterion for phylogenetic tree construction. A phylogenetic tree was generated using the Maximum Likelihood (ML) method, and the reliability of the phylogenetic groupings was assessed through bootstrap analysis with 10,000 replicates.

Multi-locus sequence analysis (MLSA)

To enhance phylogenetic classification, multi-locus sequence analysis (MLSA) was conducted by incorporating additional housekeeping genes alongside the 16S rRNA gene. Although the 16S rRNA gene is a universal marker, its limited length can lead to genetic saturation and insufficient resolution for fine-scale differentiation. To address this issue, sequences from *secY*

and 16S were included because of their high evolutionary rates and ability to provide detailed strain differentiation. Sequences were identified and extracted from the assemblies aligned using MUSCLE to test different evolutionary models and select the most appropriate model based on the likelihood criterion for phylogenetic tree construction. Phylogenetic trees were generated using both the Neighbor-Joining and Maximum Likelihood methods, with the ML analysis employing 10,000 bootstrap replicates.

Genome assembly and annotation

Complete phytoplasma genome assemblies were generated using a hybrid approach that integrated long- and short-read sequencing data to optimize both structural resolution and sequence accuracy. Oxford Nanopore long reads facilitated the assembly of contiguous genomic scaffolds and resolution of repetitive regions, while Illumina short reads provided high-accuracy base calling and error correction. Raw sequencing reads were first mapped to the *Manihot esculenta* reference genome (NCBI RefSeq assembly GCF_001659605.2) to remove the host-derived sequences. The unmapped reads were retained and subjected to de novo assembly. Illumina reads were assembled using SPAdes (v3.15.5), and Nanopore reads were assembled using Flye (v2.9). Contigs generated from both assemblies were aligned against the reference genome of *'Candidatus* Phytoplasma pruni*'* (NCBI accession number CP119306) using BLAST. Only contigs exhibiting significant homology to the reference phytoplasma genome were retained for downstream hybrid assemblies.

The hybrid assembly was constructed by merging the filtered contigs into a consensus sequence. The draft assembly was polished through multiple rounds of Illumina read mapping, enabling the correction of sequencing errors and removal of low-confidence or redundant contigs. Structural validation was conducted by aligning the final assembly with *'Ca.* Phytoplasma pruni*'* reference genome to assess synteny and completeness.

The draft genomes of both isolates OVSP-L01 and OVSP-S01 were annotated using Prodigal, Bakta, and Prokka (Hyatt et al. 2010; Schwengers et al. 2021; Seemann 2014), followed by manual inspection and BLASTp searches against the NCBI nr database. The assemblies totaled 601,536 bp (L01) and 599,901 bp (S01), comprising 624 and 621 predicted genes, respectively, of which 588 and 585 were protein-coding sequences. Both contain two copies of 16S rRNA and 23S rRNA genes and 32 tRNA gen9%%ccount4%for approximately 0.9% and 0.4% of their respective genomes (Table 2). Coding regions represent ~81–82% of the genome, with mean CDS lengths

near 835 bp, consistent with other phytoplasmas. For ORF identification, GC content calculation, and GC skew analysis, Proksee (https://proksee.ca), a platform for bacterial genome assembly, annotation, and visualization, was used. The Prokka tool[28] was used within Proksee for the rapid annotation of protein-coding genes, tRNAs, rRNAs, and other genomic features of the phytoplasma. Completeness was estimated heuristically based on the presence of conserved genetic markers.

Genome comparison

Comparative genomic analysis was performed to assess the relatedness between the newly assembled cassava-associated phytoplasma genomes and the known phytoplasma reference genomes. Average Nucleotide Identity (ANI) values based on genome pairwise comparisons were calculated using the JSpeciesWS web server (http://jspecies.ribohost.com/jspeciesws), following the methods described by Richter and Rosselló-Móra (2015)[29]. This analysis enabled a quantitative assessment of genome-wide similarity, supporting the taxonomic classification and phylogenetic placement of the detected phytoplasmas within the broader '*Candidatus Phytoplasma*' clade.

Prediction of putative secreted effectors

Potential pathogenic effectors were identified using an integrated approach combining SignalP 6.0 (Almagro-Armenteros et al. 2019) with manual curation and BLASTp validation. All annotated protein sequences from OVSP-L01 and OVSP-S01 were submitted to SignalP 6.0, and proteins predicted to contain an N-terminal signal peptide without additional transmembrane regions, were retained as potential effectors. Candidate lists from both isolates were compared to identify the consensus effectors shared by both genomes. For each consensus candidate, the longest representative sequence was retained, and redundant sequences removed through clustering (≥ 99% identity) (Table 3).

BLASTp searches against the NCBI nr database revealed that all predicted effectors shared high similarity to hypothetical proteins of closely related phytoplasmas, particularly *'Ca. Phytoplasma pruni'* and *'Conium maculatum' witches'-broom phytoplasma*. The strongest matches (99%) included proteins with 96–99% identity to *'Ca. P. pruni'* hypothetical proteins (e.g., WP_244948362.1, WP_90%4048.1, WP_178734413.1) and > 90% identity to *'C. maculatum'* witches'-broom phytoplasma proteins (e.g., MEC4558719.1, MEC4558753.1, MEC4558770.1). None of the sequences displayed significant similarity to the canonical phytoplasma effectors

SAP05, SAP11, SAP54, or TENGU, underscoring the distinct effector repertoire of *'Ca. P. manihotii'*.

References

1. Alvarez, E., Meija, J.F., Calle, F., Llano, G.A., Loke, J.B., Calari, A. *et al.* Characterization of a phytoplasma associated with frogskin disease in cassava. *Plant Dis.* **93**, 1139–1145 (2009). https://doi.org/10.1094/PDIS-93-11-1139
2. Alvarez, E., Pardo, J.M., Dufour, D.L. & Moreno, J.L. The metabolism of carbohydrates in roots of cassava (*Manihot esculenta* Crantz) infected with frogskin disease. *CGIAR Research Report* (2014). Available: https://hdl.handle.net/10568/51760
3. Flôres, D., Haas, I.C., Canale, M.C., Fantin, G.M., Bedendo, I.P. *et al.* Identificação molecular de um fitoplasma 16SrIII-B associado à doença da vassoura-de-bruxa da mandioca. *Eur. J. Plant Pathol.* **137**, 237–242 (2013). https://doi.org/10.1007/s10658-013-0250-3
4. Oliveira, S.A.S., Abreu, E.F.M., Araújo, T.S., Oliveira, E.J., Andrade, E.C., Garcia, J.M.P. *et al.* First report of a phytoplasma associated with cassava frogskin disease in Brazil. *Plant Dis.* **97**, 1268 (2014). https://doi.org/10.1094/PDIS-05-13-0499-PDN
5. Téllez, L.C.D., Pardo, J.M., Torres, A. & Alvarez, E. First report of a 16SrIII phytoplasma associated with frogskin disease in cassava (*Manihot esculenta*) in Paraguay. *Plant Dis.* **100**, 1951 (2016). https://doi.org/10.1094/PDIS-09-15-1102-PDN
6. Dolores, L.M., Langres, J.A., Pinili, M.S., Caasi-Lit, M.T., Cortaga, C.O. *et al.* Incidence, distribution, and genetic diversity of 'Candidatus Phytoplasma luffae'-related strain (16SrVIII) associated with cassava witches' broom disease in the Philippines. *Crop Prot.* **169**, 106244 (2023). https://doi.org/10.1016/j.cropro.2023.106244
7. Landicho, D.M., Montañez, R.J.M., Camagna, M., Neang, S., Bulasag, A.S., Magdaraog, P.M. *et al.* Status of cassava witches' broom disease in the Philippines and identification of potential pathogens by metagenomic analysis. *Biology* **13**, 522 (2024). https://doi.org/10.3390/biology13070522
8. Carvajal-Yepes, M., Olaya, C., Lozano, I., Cuervo, M., Zúñiga, C., Castaño, M. *et al.* Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. *Virus Res.* **186**, 76–86 (2014). https://doi.org/10.1016/j.virusres.2013.12.011
9. Oliveira, S.A.S., Ferreira, C.F., Diamantino, M.S.A.S., Oliveira, E.J., Andrade, E.C. *et al.* First report of cassava torrado-like virus, cassava polero-like virus and cassava new alphaflexivirus

associated with cassava frogskin disease in Brazil. *J. Plant Pathol.* **102**, 247 (2020). https://doi.org/10.1007/s42161-019-00384-6

10. Souza, A.N., Silva, F.N., Bedendo, I.P. & Carvalho, C.M. A phytoplasma belonging to a 16SrIII-A subgroup and dsRNA virus associated with cassava frogskin disease in Brazil. *Plant Dis.* **98**, 771–779 (2014). https://doi.org/10.1094/PDIS-04-13-0440-RE
11. Pardo, J.M., Gil-Ordóñez, A., Leiva, A.M., Cuervo, M., Lozano, I. *et al.* First report of cassava witches' broom disease and *Ceratobasidium theobromae* in the Americas. *New Dis. Rep.* **50**, e70002 (2024). https://doi.org/10.1002/ndr2.70002
12. Oliveira, S.A.S., Sheat, S., Margaria, P., Winter, S., Cuervo, M., Pardo, J.M. *et al. Rhizoctonia theobromae* associated with a severe witches' broom outbreak in cassava of the Brazilian rainforest and evidence of its Southeast Asian origin. *bioRxiv* (2025). https://doi.org/10.1101/2025.02.14.638269
13. Leiva, A.M., Pardo, J.M., Arinaitwe, W., Cuervo, M., Quintero, A., Lozano, I. *et al. Ceratobasidium* sp. is associated with cassava witches' broom disease, a re-emerging threat to cassava cultivation in Southeast Asia. *Sci. Rep.* **13**, 22500 (2023). https://doi.org/10.1038/s41598-023-49735-5
14. Landicho, D.M., Montañez, R.J.M., Manayon, V.S., Sendaydiego, M.K.Y.C., Gaylo, J.O., Dimayacyac, A.C., Sato, I., Takemoto, D., Pinili, M.S. & Chiba, S. Detection, isolation, and identification of *Rhizoctonia theobromae* associated with cassava witches' broom disease in the Philippines. *Physiol. Mol. Plant Pathol.* **140**, 102872 (2025). https://doi.org/10.1016/j.pmpp.2025.102872
15. Jimenez, J., Caicedo, S., Pardo, J.M., Gil-Ordóñez, A., Cuervo, M. *et al.* Single torradovirus infections explain the mysterious cassava frogskin disease in the Americas. *Sci. Rep.* **14**, 29648 (2024). https://doi.org/10.1038/s41598-024-81142-2
16. Nafea, A.M., Wang, Y., Wang, D., Salama, A.M., Aziz, M.A., Xu, S. & Tong, Y. Application of next-generation sequencing to identify different pathogens. *Front. Microbiol.* **14**, 1329330 (2024). https://doi.org/10.3389/fmicb.2023.1329330
17. Espindola, A.S. Simulated high throughput sequencing datasets: A crucial tool for validating bioinformatic pathogen detection pipelines. *Biology (Basel)* **13**, 700 (2024). https://doi.org/10.3390/biology13090700

18. Maina, S., Zheng, L. & Rodoni, B.C. Targeted genome sequencing (TG-Seq) approaches to detect plant viruses. *Viruses* **13**, 583 (2021). https://doi.org/10.3390/v13040583
19. Bertaccini, A., Arocha-Rosete, Y., Contaldo, N., Duduk, B., Fiore, N., Montano, H.G. *et al.* Revision of the 'Candidatus Phytoplasma' species description guidelines. *Int. J. Syst. Evol. Microbiol.* **72**, 005353 (2022). https://doi.org/10.1099/ijsem.0.005353
20. Pardo, J.M., Alvarez, E., Lopez-Lavalle, L.A.B., Cuervo, M., Gil, J. *et al.* Cassava frogskin disease: Current knowledge on a re-emerging disease in the Americas. *Plants* **11**, 1841 (2022). https://doi.org/10.3390/plants11141841
21. Graziosi, I., Minato, N., Alvarez, E., Ngo, D.T., Hoat, T.X., Le, H.T. *et al.* Emerging pests and diseases of Southeast Asian cassava: A comprehensive evaluation of geographic priorities, management options and research needs. *Pest Manag. Sci.* **72**, 1071–1089 (2016). https://doi.org/10.1002/ps.4250
22. Calvert, L., Cuervo, M. & Lozano, I. Identification of three strains of a virus associated with cassava plants affected by frogskin disease. *J. Phytopathol.* **156**, 647–653 (2008). https://doi.org/10.1111/j.1439-0434.2008.01412.x
23. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021). https://doi.org/10.1093/molbev/msab120
24. Zhao, Y., Wei, W., Lee, I.-M., Shao, J.Y., Suo, X. & Davis, R.E. Construction of an interactive online phytoplasma classification tool, iPhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). *Int. J. Syst. Evol. Microbiol.* **59**, 2582–2593 (2009). https://doi.org/10.1099/ijs.0.010249-0
25. Lee, I.-M., Bottner-Parker, K.D., Zhao, Y. & Davis, R.E. Phylogenetic analysis and delineation of phytoplasmas based on secY gene sequences. *Int. J. Syst. Evol. Microbiol.* **65**, 1382–1392 (2010). https://doi.org/10.1099/ijs.0.019695-0
26. Aagaard, K. *pDRAW32 DNA analysis software*, v1.1.147 (AcaClone Software, 2025). Available at: http://www.acaclone.com
27. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available: https://www.R-project.org/ (2024).
28. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014). https://doi.org/10.1093/bioinformatics/btu153

29. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl Acad. Sci. USA* **106**, 19126–19131 (2009). https://doi.org/10.1073/pnas.0906412106

Acknowledgements

We thank Dr. Jaeveson da Silva and Mr. José Leonardo Nascimento for collecting and kindly providing part of the samples used in this study.

Funding

This work was supported by Embrapa and DSMZ. Saulo A.S.O. received a Research Productivity Fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grant no. 315172/2021-5) and an Experienced Research Fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)–Alexander von Humboldt (AvH).

Author contributions statement

S.A.S.O., V.P.S., M.S.A.S.D., and C.P.F.S. were responsible for sample collection and on-site disease identification. S.A.S.O., S.S., and S.W. conceptualized the high-throughput sequencing approach for pathogen identification. Data analysis was performed by S.A.S.O., V.P.S., P.M., and S.S., while data curation was performed by S.A.S.O., S.S., P.M., C.F.F., and S.W. The methodology was developed by S.W., P.M., and S.S. Funding acquisition was managed by S.A.S.O., C.R.J., and S.W. All authors contributed to the original draft and participated in reviewing and editing the manuscript. S.A.S.O. and S.S. contributed equally to this work.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable requests.

Competing interests

The author(s) declare no competing interests.

Tables

Table 1. Summary of results from Illumina sequencing for CFSD, OVSP-L01, and OVSP-S01 samples.

Illumina data	Samples		
	CFSD	**OVSP-L01**	**OVPS-S01**
Total reads	62,648,108	48,823,878	60,609,962
Total reads after normalization	29,673,340	29,698,224	33,353,050
% Reads mapped to cassava	84,91%	70,76%	78,78%
# Reads unmapped	4,476,529	8,684,988	7,076,949
Number of produced contigs	6,020	5,250	4,975
Minimum Length (bp)	489	431	357
Maximum Length (bp)	42,266	53,407	32,731
'Phytoplasma-like' contigs	20	1,300	1,280

CFSD= Cassava Frogskin Disease
OVSP-L01= Over Sprouting Phytoplasma from Leaves
OVSP-S01= Over Sprouting Phytoplasma from Stems
Total reads = number of total reads obtained after quality check, filtering, and trimming
 # Reads unmapped = number of reads not belonging to the host

Table 2 General features of the genome of the "*Candidatus* Phytoplasma manihotii" sp. nov. isolates OVSP-S01 (obtained from stems) and OVSP-L01 (obtained from leaves) associated with cassava oversprouting disease.

OVSP-S01 (total size = 599,901)

Gene	**Number**	**Total length**	**Average length**	**Percentage of genome**
Gene	621	502325	808	83.74
CDS	585	491115	839	81.87
tRNA	32	2523	79	0.42
23S rRNA	2	5641	2820,5	0.94
16S rRNA	2	3046	1523	0.51

OVSP-L01 (total size = 601,536)

Gene	**Number**	**Total length**	**Average length**	**Percentage of genome**
Gene	624	501849	804	83.43
CDS	588	490605	834	81.56
tRNA	32	2523	79	0.42
23S rRNA	2	5676	2838	0.94
16S rRNA	2	3045	1522	0.51

CDS= Protein-code regions

606 **Table 3.** Potential pathogenic effectors of *'Candidatus* Phytoplasma manihotii' sp. nov.

Protein ID	Length (aa)	Description	Closest Phytoplasma	Accession
OINLKOPJ_00131 and EFLOPLJH_00131	255	hypothetical protein	*Ca.* Phytoplasma pruni	WP_244948362.1
OINLKOPJ_00019 and EFLOPLJH_00363	217	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558719.1
OINLKOPJ_00361 and EFLOPLJH_00363	117	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558919.1
OINLKOPJ_00423 na EFLOPLJH_00425	218	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558753.1
OINLKOPJ_00446 and EFLOPLJH_00425	246	hypothetical protein	*Ca.* Phytoplasma pruni	WP_178734048.1
OINLKOPJ_00474 and EFLOPLJH_00425	210	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558753.1
OINLKOPJ_00403 and EFLOPLJH_00405	413	hypothetical protein	*Ca.* Phytoplasma pruni	WP_178734413.1
OINLKOPJ_00360 and EFLOPLJH_00362	171	hypothetical protein	*Ca.* Phytoplasma pruni	WP_178733878.1
OINLKOPJ_00360 and EFLOPLJH_00553	343	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558770.1
OINLKOPJ_00360 and EFLOPLJH_00550	163	hypothetical protein	*Ca.* Phytoplasma pruni	WP_178733878.1
OINLKOPJ_00589 and EFLOPLJH_00362	324	hypothetical protein	*Ca.* Phytoplasma pruni	WP_178733997.1
OINLKOPJ_00360 and EFLOPLJH_00462	231	hypothetical protein	Italian clover phyllody phytoplasma	WP_017192863.1
OINLKOPJ_00555 and EFLOPLJH_00362	344	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558770.1
OINLKOPJ_00552 and EFLOPLJH_00550	158	hypothetical protein	*Ca.* Phytoplasma pruni	WP_178734426.1
OINLKOPJ_00459 and EFLOPLJH_00362	230	hypothetical protein	Italian clover phyllody phytoplasma	WP_017192863.1
OINLKOPJ_00154 and EFLOPLJH_00445	95	hypothetical protein	Conium maculatum witches'-broom phytoplasma	MEC4558805.1

607 aa= amino acids

608

609

610

611

612

613

614

615

616

617

618

619

Table 4. Overview of samples from Bahia and Ceará from different municipalities in 2023, categorized by symptoms, tissue type, and collection site.

Pool	Sample n°	Symptoms	Municipality	State	Tissue	Collection year
CFSD	216	Frogskin	Cruz das Almas	Bahia	Stem	2023
CFSD	324	Frogskin	Cruz das Almas	Bahia	Secondary root	2023
CFSD	400	Frogskin	Cruz das Almas	Bahia	Tuberous root	2023
CFSD	418	Frogskin	Cruz das Almas	Bahia	Tuberous root	2023
CFSD	420	Frogskin	Cruz das Almas	Bahia	Leaf	2023
OVSPL	15B	Oversprouting	Laje	Bahia	Petiole	2023
OVSPL	29	Oversprouting	Serra do Salitre	Ceará	Petiole	2023
OVSPL	1054	Oversprouting	Serra do Salitre	Ceará	Petiole	2023
OVSPL	1060	Oversprouting	Serra do Salitre	Ceará	Petiole	2023
OVSPL	1059	Oversprouting	Serra dos Estevam	Ceará	Petiole	2023
OVSPS	1052	Oversprouting	Chap. Pitomb.	Ceará	Stem	2023
OVSPS	1053	Oversprouting	Serra do Salitre	Ceará	Stem	2023
OVSPS	1055	Oversprouting	Serra do Salitre	Ceará	Stem	2023
OVSPS	1058	Oversprouting	Serra dos Estevam	Ceará	Stem	2023

CFSD = A set of samples containing total DNA extracted from different cassava tissues expressing frog skin disease symptoms in tuberous roots and no visible symptoms in the leaves or stems.
OVSPL = A set of samples containing total DNA extracted from the leaves of plants showing shoot proliferation symptoms.
OVSPS = A set of samples containing total DNA extracted from the stems of plants showing shoot proliferation symptoms.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx