UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

FCA
FACULDADE DE CIÊNCIAS AGRONÔMICAS
UNESP | BOTUCATU

MARCOS ROBERTO RIBEIRO JUNIOR

DETECTION OF GLOBAL SOYBEAN VIRUSES IN METAGENOMIC SEQUENCE DATA USING MICROBE FINDER (MiFi®) & THRIPS VARIABILITY ON SOYBEAN

Botucatu

2022

MARCOS ROBERTO RIBEIRO JUNIOR

DETECTION OF GLOBAL SOYBEAN VIRUSES IN METAGENOMIC SEQUENCE DATA USING MICROBE FINDER (MiFi®) & THRIPS VARIABILITY ON SOYBEAN

Thesis presented to São Paulo State University (Unesp) School of Agriculture, Botucatu, to obtain the degree: Doctor in Agronomy - Plant Protection.

Advisor: Renate Krause Sakate
Co-advisor: Francisco M. Ochoa Corona

Botucatu
2022

R484d

Ribeiro Junior, Marcos Roberto

Detection of global soybean viruses in metagenomic sequence data using Microbe Finder (MiFi®) & Thrips variability on soybean / Marcos Roberto Ribeiro Junior. -- Botucatu, 2022

113 p. : il., tabs., fotos

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agronômicas, Botucatu

Orientadora: Renate Krause Sakate

Coorientador: Francisco Manuel Ochoa Corona

1. next-generation sequencing. 2. EDNA. 3. Thysanoptera. 4. Glycine max. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agronômicas, Botucatu. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

FCA
FACULDADE DE CIÊNCIAS AGRONÔMICAS
UNESP | BOTUCATU

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DETECTION OF GLOBAL SOYBEAN VIRUSES IN METAGENOMIC SEQUENCE DATA USING MICROBE FINDER (MiFi®) & THRIPS VARIABILITY ON SOYBEAN

AUTOR: MARCOS ROBERTO RIBEIRO JUNIOR
ORIENTADORA: RENATE KRAUSE SAKATE
COORIENTADOR: FRANCISCO MANUEL OCHOA CORONA

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (PROTEÇÃO DE PLANTAS), pela Comissão Examinadora:

Prof.ª Dr.ª RENATE KRAUSE SAKATE (Participaçao Virtual)
Protecao Vegetal / Faculdade de Ciencias Agronomicas de Botucatu UNESP

Prof. Dr. ANDRES ESPINDOLA CAMACHO (Participaçao Virtual)
Entomology & Plant Pathology / Oklahoma State University

Pesquisadora Dr.ª CRISTIANE MÜLLER (Participaçao Virtual)
Entomologia / Corteva Agriscience

Pesquisador Dr. BRUNO ROSSITTO DE MARCHI (Participaçao Virtual)
. / Eurofins Agrosciences Services

Prof.ª Dr.ª MONIKA FECURY MOURA (Participaçao Virtual)
Coordenação de Planejamento e Pesquisa / Fundação Centro de Referência em Educação Ambiental Escola Bosque

Botucatu, 27 de junho de 2022

Dedicated to my Grandfather.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my adviser Dr. Renate Krause-Sakate for her guidance from the beginning of my journey and for giving me this amazing opportunity to undertake this challenging endeavor. I am thankful for her support and encouragement during hard times in research. I also would like to express my sincere thanks to my co-adviser Dr. Francisco Ochoa-Corona for his guidance during my time at Oklahoma State University. He is an example of a professional model of what I hope to embody in the future spaces I join. I thank Dr. Andres Espindola, developer of EDNA MiFi, for his support and advice with sequencing and bioinformatics. He was exceptionally helpful in every step of my EDNA research. I would like to extend my gratitude to Dr. Cristiane Muller from Corteva Agriscience for her assistance, financial support, and for sending thrips specimens to our collection. To Dr. Fabricio Lima Bezerra for his willingness in helping us to identify the species of thrips and for sharing with us his specimens to build our thrips database.

Additionally, I want to thank the Department of Plant Protection (DPP) and the Graduate Program in Agronomy - Plant Protection (PPGA-PP) at São Paulo State University (Unesp), the Department of Entomology and Plant Pathology (EPP), and the Institute for Biosecurity & Microbial Forensics (IBMF) at Oklahoma State University (OSU). I also want to thank all members of the Laboratory of Plant Virology and Virus-Vector-Plant Interaction (LAVIV) at Unesp, the Virus Chasers Research Group at OSU, students, staff, and professors who I had contact with. You were incredibly important during all these years.

Last but not the least, I am immensely grateful to my beloved parents, brothers, grandma, wife, and my four-legged friends. I am blessed to have you in my life and I could not accomplish this without each of you.

This thesis was possible thanks to the scholarship granted from the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), in the scope of the Program CAPES-PrInt, process number 88887.310463/2018-00, mobility number 88887.570292/2020-00.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Marcos Roberto Ribeiro Junior was a recipient of CAPES and CNPq (National Council for Scientific and Technological Development) Ph.D. fellowships (grant numbers: 88882.180603/2018-01 and 140130/2019-5).

“Evolution forged the entirety of sentient life on this planet using only one tool: the mistake”.

WESTWORLD: Season 1, Episode 1 - The Original. Created and executive produced by Jonathan Nolan & Lisa Joy. New York, NY: HBO, 2016.

ABSTRACT

Soybean (*Glycine max*) is an important crop worldwide used as a source of oilseed and protein. Nevertheless, each year, soybean growers lose significant yields to diseases caused by viruses and damage caused by thrips, therefore, precise diagnosis is crucial for management. Passiflora virus Y (PaVY) was detected naturally infecting soybean for the first time in Brazil and the complete genome sequencing shows a 9679nt long virus that shares 84.4% nt identity with a previously reported PaVY isolate from *Passiflora* spp. The virus was successfully transmitted by *Myzus persicae,* and sap transmitted to indicator hosts. Electronic probes (e-probes) ranging from 19-60 nucleotides in length were designed for the 46 viruses that infect soybean using MiProbe within MiFi® platform. *In silico* results validate the usage of e-probes to detect up to 46 viruses in metagenomic soybean data. *In vitro analyses* revelated nine viruses present in Brazilian HTS Illumina soybean samples. The curated soybean virus e-probes are free of access for diagnostics through MiDetect within the MiFi® platform. Nine thrips species, collected from 20 different plant species in Brazil, including soybean, were identified by sequencing partial COI gene. Specific primers based on COI and 28S sequences were obtained for *Caliothrips phaseoli* and *Frankliniella occidentalis* species, respectively. Bayesian phylogenetic analyses demonstrated that specimens of the same species were clustered under their respective clade, but not necessarily under the same genus clade. Our results demonstrate that PCR and sequencing can be used for identification of well described species, but traditional taxonomy is still necessary for the identification of new species and must be combined with molecular techniques.

Keywords: next-generation sequencing; EDNA; Thysanoptera; *Glycine max*.

RESUMO

A soja (*Glycine max*) é uma importante cultura utilizada mundialmente como fonte de óleo e proteína. No entanto, a cada ano, os produtores de soja sofrem perdas significativas na produtividade devido a doenças causadas por vírus e danos causados por tripes. Esse trabalho apresenta o primeiro relato de passiflora virus Y (PaVY) infectando naturalmente soja Brasil. O genoma do PaVY foi sequenciado completamente e apresentou 9679nt de comprimento e 84,4% de identidade com um outro isolado PaVY coletado em *Passiflora* spp. e relatado anteriormente. O vírus foi transmitido por afídeos da espécie *Myzus persicae* e por inoculação em hospedeiros indicadores usando extrato vegetal infectado. Hospedada dentro da plataforma MiFi®, a ferramenta MiProbe foi utilizada para a projeção de sondas eletrônicas (e-probes), com comprimento variado entre 19 a 60 nucleotídeos, para detectar simultaneamente todos os vírus descritos infectando soja no mundo, através de dados de Next-Generation Sequencing. Os resultados *in silico* demonstraram que as e-probes são sensíveis o bastante para detectar até 46 vírus em dados metagenômicos de soja. Análises *in vitro* revelaram nove vírus presentes em amostras brasileiras de soja sequenciadas utilizando a plataforma Illumina. As e-probes de soja foram validadas e estão disponíveis para outros pesquisadores através da função MiDetect na plataforma MiFi®. Nove espécies de tripes, coletadas em 20 espécies vegetais diferentes, incluindo a soja, foram identificadas por sequenciamento parcial do gene COI. Primers específicos baseados nas sequências COI e 28S foram obtidos para as espécies *Caliothrips phaseoli* e *Frankliniella occidentalis*, respectivamente. Análises filogenéticas e bayesianas demonstraram que insetos pertencentes à mesma espécie foram agrupados sob seu respectivo clado, mas não necessariamente sob o mesmo clado de gênero. Análises de sequencias pareadas do mtCOI indicaram que *F. schultzei* é a espécie com sequência de nucleotídeos mais variável entre seus indivíduos. Nossos resultados demonstram que PCR e sequenciamento podem ser utilizados para identificação de espécies de tripes previamente descritas na literatura e que possuam sequencias disponíveis em bancos de dados, mas a taxonomia tradicional ainda é necessária para a identificação de novas espécies, devendo ser combinada com técnicas moleculares.

Palavras-chave: sequenciamento de nova geração; EDNA; Thysanoptera; *Glycine max*.

SUMMARY

GENERAL INTRODUCTION 17

CHAPTER 1 - MOLECULAR AND BIOLOGICAL CHARACTERIZATION OF AN ISOLATE OF THE POTYVIRUS PASSIFLORA VIRUS Y NATURALLY INFECTING SOYBEAN (*Glycine max*) IN BRAZIL 20

REFERENCES 33

CHAPTER 2 - AN ATTEMPT TOWARD THE GLOBAL MONITORING OF SOYBEAN VIRUSES USING MICROBE FINDER (MiFi®) 37

2.1 INTRODUCTION 39

2.2 MATERIAL AND METHODS 40

2.3 RESULTS 48

2.4 DISCUSSION 57

REFERENCES 58

CHAPTER 3 - MOLECULAR IDENTIFICATION AND PHYLOGENETIC ANALYSES OF THRIPS SPECIES COLLECTED IN BRAZIL 63

3.1 INTRODUCTION 64

3.2 MATERIAL AND METHODS 66

3.3 RESULTS 70

3.4 DISCUSSION 84

3.5 CONCLUSION 85

REFERENCES 86

FINAL CONSIDERATIONS 91

REFERENCES 93

APPENDIX A - MOLECULAR AND BIOLOGICAL CHARACTERIZATION OF APIUM VIRUS Y INFECTING PARSLEY IN BRAZIL 96

GENERAL INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is a major economic crop worldwide. This staple crop is cultivated in about 6% of the world's arable land and is among the five most important food crops in the world (HARTMAN; WEST; HERMAN, 2011; SAVARY et al., 2019). The leading worldwide producer of soybean is Brazil, which reached 125.00 million metric tons in the 2021/22 season, corresponding to 35.64% of global soybean output. Brazil is the world's largest global exporter of soybeans with 61.69% of its production exported (77 million tons) (USDA, 2022), followed by the United States (USDA, 2022). Soybean fields are scattered all over the country and it is the main agricultural product and export item in value (PROINDE, 2020).

Soybean is a versatile oilseed supplies the domestic market with edible food products, cooking oil, biofuels, and meal for animal feed (PROINDE, 2020). However, soybean growers must deal with several adversities, such as the occurrence of diseases that severely affect production (BANDARA et al., 2020).

Among these plant diseases, viruses represent a challenge for management. Farmers are limited to preventing the entry of the virus and possible insect vectors, like whiteflies and thrips. In this scenario, biosecurity adopts measures for controlling the spread of plant diseases. Plant material for phytosanitary inspection and testing in the borders is required before it enter in the country. This way biosecurity measure makes possible to avoid the introduction of diseases that are not present in the country (OLIVER, 2022).

At least 46 virus species are described to have infected soybean worldwide (HILL; WHITHAM, 2014). Kitajima, (2020) described 15 viruses infecting soybean in Brazil until 2018: alfafa mosaic virus (AMV), bean golden mosaic virus (BGMV), bean rugose mosaic virus (BRMV), bean yellow mosaic virus (BYMV), cowpea mild mottle virus (CPMMV), cowpea severe mosaic virus (CPSMV), euphorbia yellow mosaic virus (EuYMV), okra mottle virus (OMoV), sida micrantha mosaic virus (SimMV), sida mottle virus (SiMoV), southern bean mosaic virus (SBMV), soybean chlorotic spot virus (SCSV), soybean mosaic virus (SMV), tobacco streak virus (TSV), and tomato severe rugose virus (ToSRV).

Groundnut ringspot virus (GRSV) was founded on soybean fields in Brasília, Distrito Federal, in 2017. It is possible the presence of viruliferous thrips, such as *Frankliniella schultzei*, may be associated with disease spread in the crop (FONTES et al., 2019).

The last virus reported in soybean in Brazil was passiflora virus Y (PaVY) (RIBEIRO-JUNIOR et al., 2022). To present, 18 soybean viruses are reported in Brazil, and the other 28 are described infecting soybean in other countries. Biosecurity methods need to be adopted to prevent the introduction of new viruses in countries where they are not described yet.

Enzyme-Linked-Immunosorbent Assay (ELISA) and polymerase chain reaction (PCR) are the gold standards methods used to detect plant viruses (POSTNIKOVA et al., 2008). However, these methods present limitations, if intended for detection of simultaneous pathogens (up to 46) (POSTNIKOVA et al., 2008).

Next-Generation Sequencing (NGS) is a novel alternative, which allows almost all pathogen in sample to be sequenced, however, the bioinformatic processing of data is still challenging (JONES, 2010; KREUZE et al., 2009; TYSON et al., 2004). To overcome this gap between the high-volume data generated by NGS and pathogen detection, the E-probe Diagnostic for Nucleic acid Analysis (EDNA), hosted in the platform called Microbe Finder (MiFi®), was developed (ESPINDOLA; CARDWELL, 2021; STOBBE et al., 2013).

MiFi® can be used by diagnosticians without advanced coding skills. It requires e-probes to be developed, curated, and validated for the target pathogens (ESPINDOLA; CARDWELL, 2021). Given the importance of soybean and the high number of viruses associated with the crop, MiFi® is an alternative for quick detection of multiple viruses present in a sample.

At present, e-probes are available to detect rose and cucurbit viruses, water-borne plant viruses, viruses associated with citrus leprosis syndrome, grapevine, and blueberries pathogens, fungus and oomycetes, and also the aflatoxin production in the soil (BOCSANCZY et al., 2018; ESPINDOLA et al., 2019, 2015, 2018; ESPINDOLA; CARDWELL, 2019, 2021; PROANO; ESPINDOLA; GARZON, 2018; ZUNIGA et al., 2017, 2018).

Additional challenges related to soybean viruses are the emergent insect vectors, thrips. There is a lack of information about the main thrips species that affect

soybean in Brazil. Two genera, *Frankliniella* and *Thrips*, are responsible for orthotospovirus transmission, but more studies are needed for the correct molecular identification and phylogenetic classification of these insects (PAPPU; JONES; JAIN, 2009).

This thesis was divided into three chapters: Chapter 1 “Molecular and biological characterization of an isolate of passiflora virus Y (*Potyvirus*) naturally infecting soybean (*Glycine max*) in Brazil”, Chapter 2 “An attempt toward the global monitoring of soybean viruses using Microbe Finder (MiFi®)”, and Chapter 3 “Molecular identification and phylogenetic analyses of thrips species collected in Brazil”.

CHAPTER 1

MOLECULAR AND BIOLOGICAL CHARACTERIZATION OF AN ISOLATE OF THE POTYVIRUS PASSIFLORA VIRUS Y NATURALLY INFECTING SOYBEAN (*Glycine max*) IN BRAZIL*

Marcos Roberto Ribeiro-Junior[1,2,3], Felipe Barreto da Silva[1], Julio Massaharu Marubayashi[1], Juliana Uzan[1], Angelica Maria Nogueira[1], Cristiane Muller[4], Daniele Maria Nascimento[1,3], Valdir Atsushi Yuki[5], Nobuyoshi Narita[6], Marcelo Agenor Pavan[1], Francisco Manuel Ochoa-Corona[2,3], Renate Krause-Sakate[1].

[1]Department of Plant Protection, São Paulo State University, School of Agriculture, Botucatu, SP, Brazil.
[2]Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, USA.
[3]Institute for Biosecurity and Microbial Forensics, Oklahoma State University, Stillwater, OK, USA.
[4]Corteva Agriscience, Mogi Mirim, SP, Brazil.
[5]Agronomic Institute of Campinas, Campinas, SP, Brazil.
[6]APTA Regional Research and Development Unit of Presidente Prudente, Presidente Prudente, SP, Brazil.
Corresponding author (e-mail: marcos.ribeiro@unesp.br) https://orcid.org/0000-0002-5024-1801

Archives of Virology (2022) 167:2743–2747
https://doi.org/10.1007/s00705-022-05605-5

* Chapter written according to the journal **Archives of Virology**

Abstract

Passiflora virus Y was detected naturally infecting soybean (*Glycine max*) for the first time in Brazil. Here, we report the nearly complete genome sequence and molecular and biological properties of the PaVY-Br isolate. The nearly complete genome sequence is 9679 nt long and shares 84.4% nt sequence identity with a previously reported PaVY isolate from *Passiflora* sp. PaVY-Br induced chlorotic spots and systemic mosaic on soybean and chlorotic local lesions on yellow passion fruit (*Passiflora edulis*) and sesame (*Sesamum indicum*). The virus was successfully transmitted by *Myzus persicae*, indicating that this aphid vector can contribute to the spread of PaYV from passion fruit to soybean plants. Additional epidemiological research is in progress to investigate the distribution of PaVY in soybean production areas in Brazil.

Keywords: *Potyviridae,* PaVY, surveys.

Soybean (*Glycine max*) is one of the most important crops worldwide. Since 2020, Brazil has been the largest producer of soybeans in the world, producing 114.2 million tons in the 2019/2020 growing season [1]. Several diseases affect soybean production, and viruses cause major problems, with at least 46 viruses identified infecting soybean worldwide [2]. About 20 viruses have been described in soybean in Brazil, belonging to species of the genera *Alfamovirus* (1 species), *Begomovirus* (5), *Bromovirus* (1), *Carlavirus* (1), *Comovirus* (1), *Cucumovirus* (1), *Ilarvirus* (1), *Orthotospovirus* (2), *Potyvirus* (5), *Sobemovirus* (1), and *Tobamovirus* (1) [3,4,5,6,7,8,9,10].

In January 2020, soybean plants were found showing symptoms of chlorosis on young leaves in the municipality of Taciba, São Paulo, Brazil (22° 24′ 38″ S, 51° 19′ 30″ W). Total RNA was extracted from two symptomatic leaves, using a Purelink Viral RNA/DNA Mini Kit (Thermo Fisher Scientific). The total RNA was used for the construction of a cDNA library, using a Complete ScriptSeq Kit (Epicenter, Illumina). Transcriptome sequencing was performed on an Illumina HiSeq2500 platform, using HiSeq Flow Cell v4 and HiSeq SBS v4 (Illumina) in paired-end mode (2 × 100 bp). The reads generated (39,551,276 total) were trimmed and assembled de novo using the CLC Genomics Workbench software v9.0.3. The 16,409 contigs that were obtained (average length, 887 bp) were used as queries in a BLAST search of the NCBI virus genome database using Geneious software v 11.1.6. This analysis showed the presence of potyvirus-like contig sequences. Based on the sequence analysis, the previously extracted RNA was used in a single-step RT-PCR reaction with the universal potyvirus primers W-CIEN (5′-ATG GTT TGG TGY ATY GAR AAT-3′) and PV-1 (5′-GAT TTA GGT GAC ACT ATA GT(17)-3′) [11, 12], which allow the amplification of a fragment of approximately 850 bp that corresponds to part of the capsid protein (CP)-coding region and the 3′ untranslated region (UTR). The amplified fragment was sequenced (GenBank accession no. MZ190342) and found to share 94% identity with an isolate of passiflora virus Y (PaVY) from *Passiflora foetida* in Australia (AY461661).

The first complete genome sequence of PaVY (MW165064) [13] served as the reference for mapping HTS data. A total of 2,835,543 reads were mapped to the PaVY genome, covering the entire sequence with a mean coverage depth of 17277×. No other potyvirus sequences were identified in the HTS output. The nearly complete

PaYV-Br sequence (MZ190341) is 9679 nt long, with a 5′ UTR of 169 nt and a 3′ UTR of 257 nt, excluding the poly(A) tail.

PaVY was first reported infecting stinking passionflower (*Passiflora foetida*) in the Indonesian province of Papua and commercial passion fruit (*P. edulis*) in Australia [14]. Symptoms include mosaic, chlorotic spots, ringspot, leaf deformation, and stunting in passionflower plants [14, 15]. The virus is transmitted non-persistently by *Aphis gossypii* and has been sap transmitted to *Passiflora* spp., *Glycine max*, *Nicotiana benthamiana*, and *Macroptilium lathyroides* [14].

This study reports an isolate of PaVY naturally infecting soybean plants growing near a commercial passion fruit crop. The molecular and biological properties of this soybean isolate were studied.

The putative ORF of PaVY starts at an AUG codon at nucleotide positions 170–173 and ends with an UGA codon at nt 9419–9422, encoding a polyprotein of 3083 amino acids (aa) with a molecular weight of 352.16 kDa. Nine putative protease cleavage sites were identified in the PaVY-Br polyprotein, giving rise to 10 putative mature proteins: P1 (318 aa), HC-Pro (457 aa), P3 (347 aa), 6K1 (52 aa), CI (634 aa), 6K2 (52 aa), VPg (190 aa), Nia-Pro (242 aa), Nib (516 aa), and CP (273 aa). The cleavage sites are, in order, VRHY/S, YRVG/G, VVAQ/G, VRVQ/S, VQLQ/S, VATE/G, VEME/S, VEVQ/S, and VTLQ/S (Fig. 1). The cleavage sites are similar (but not identical) between the Brazilian and Chinese (MW165064) isolates. In the cleavage site between P1 and HC-Pro, Ile is replaced by Val; between VPg and Nla-Pro, Val is replaced by Met; and between Nib and CP, Val is replaced by Thr. An ORF for the putative protein PIPO was identified within the P3 region in the +2 reading frame, starting at position 2953 and ending at position 3178.

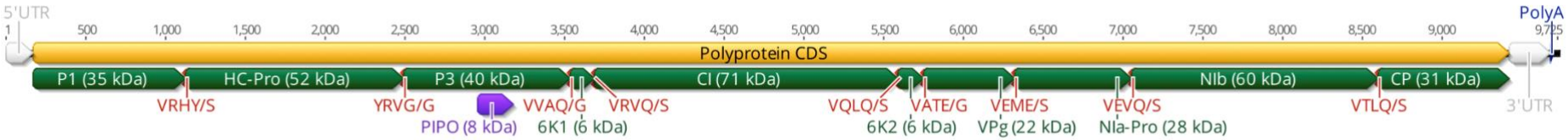


Figure 1 – Diagram showing the genomic organization of the PaVY-Br isolate with its predicted polyprotein cleavage sites. White boxes indicate the 5′- and 3′- untranslated regions, the yellow box shows the large open reading frame (ORF), the ten putative mature proteins and sizes are indicated by green boxes and the putative protein PIPO is indicated by a purple box. The location of each putative cleavage site on the polyprotein is red marked.

Pairwise comparisons were performed with SDT v1.2 [16], using the complete polyprotein nucleotide sequence from 37 potyviruses belonging to the bean common mosaic virus (BCMV) subgroup as described previously [13] (Fig. 2). The PaVY-Br isolate shares 84.6% nucleotide and 96.5% amino acid sequence identity with the Chinese PaVY isolate (MW165064) [13] and 84.8% nucleotide and 92% amino acid sequence identity with passiflora foetida virus Y (PfVY) (LC466655) [17]. The Chinese PaYV isolate and PfVY share 98.8% nt sequence identity, which shows that these two viruses are members of the same species. According to the International Committee on Taxonomy of Viruses (ICTV) [18], members of the same potyvirus species share more than 76% nucleotide and 82% amino acid sequence identity in the complete ORF. PaVY and PfVY are unassigned members of the genus Potyvirus. The name PaYV was proposed in 2004 [14], and since it was used for the completely sequenced isolate from China (MW165064) [13], we chose to use this name for the Brazilian potyvirus studied here as well (PaVY-Br).

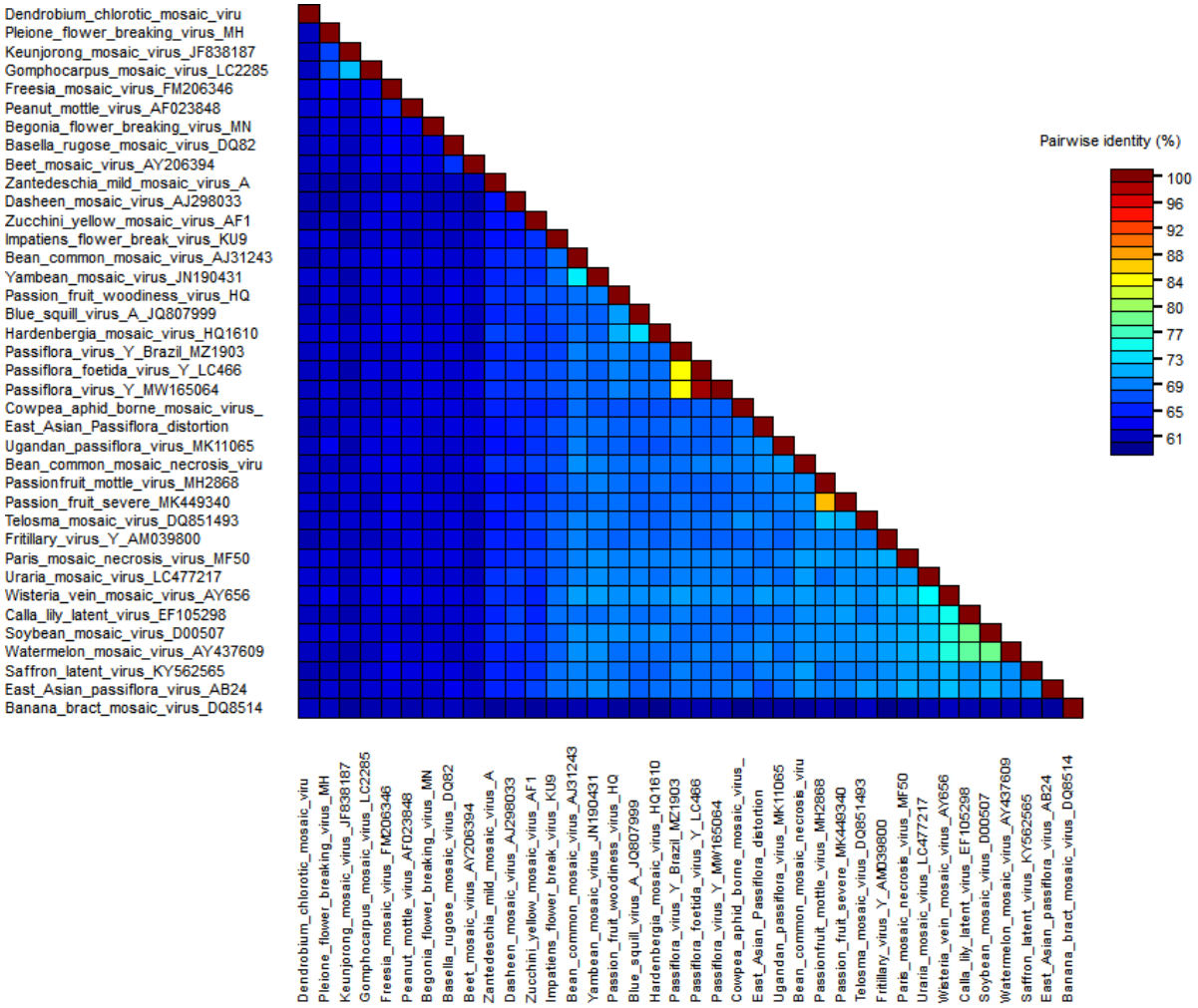


Figure 2 – Percentage of nucleotide sequence identity of the Brazilian isolate of passiflora virus Y (PaVY-Br) (MZ190341) to all other potyviruses belonging to the BCMV subgroup.

A Bayesian phylogenetic tree was constructed using the complete nucleotide sequence dataset described above. The sequences were aligned using MUSCLE v.3.8.425 within the Geneious v.11.1.5 software, and phylogenetic analysis was performed using Mr. Bayes 3.2.2 [19]. The program MrModeltest v.2.2 [20] was used to select the nucleotide substitution model GTR+I+G using the Akaike information criterion (AIC). Two independent runs were conducted simultaneously using 10 million generations and excluding 25% of the resulting trees as burn-in. The phylogenetic tree was constructed using FigTree v1.4.4, and the results were in agreement with those of a previous study [13]. PaVY (and PfVY) formed an independent evolutionary lineage distinct from those of the other potyviruses (Fig. 3).

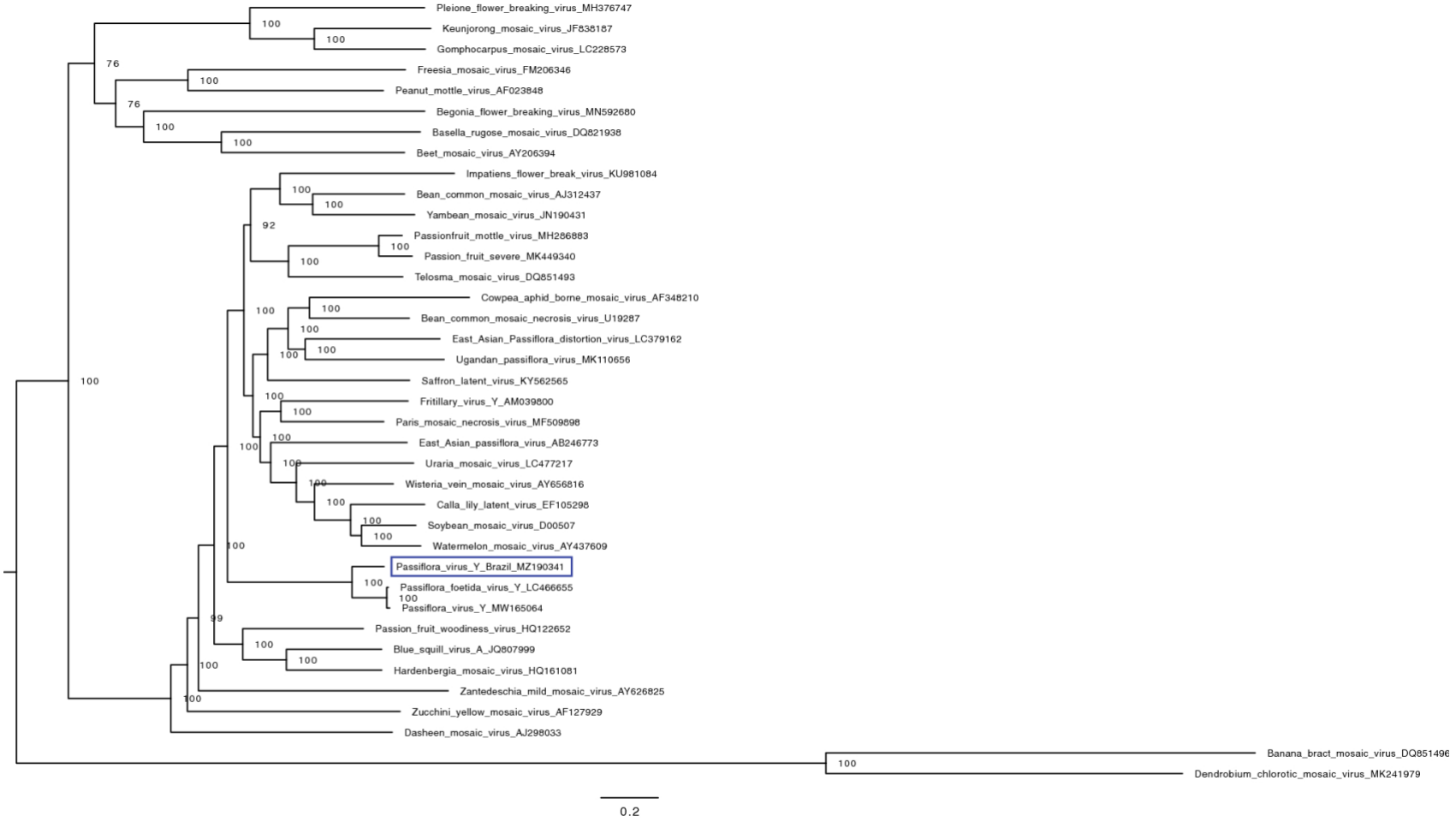


Figure 3 – Bayesian phylogenetic tree based on complete nucleotide sequences of the Brazilian isolate of passiflora virus Y (PaVY-Br) (MZ190341) and all other potyviruses belonging to the BCMV subgroup. PaVY-Br is highlighted in blue.

A new set of specific primers was designed for the PaVY-Br isolate using Primer3 2.3.7. within Geneious software. Alignments of the genome sequence with those of the main potyviruses that infect soybean revealed that the PaVY-Br HC-Pro coding region was the most appropriate for designing new primers. These primers, named PaVY 1988F (TAT CCG TGC TGC TGT GT AAC) and PaVY 2244R (TTT GGG ATT AGA GCG TC ACG), were used for RT-PCR, with the following parameters: reverse transcription at 42°C for 30 min, followed by initial denaturation at 95°C for 5 min, and then 40 cycles of 95°C for 30 s, 59°C for 30 s, and 72°C for 30 s, followed by a final extension step at 72°C for 10 min. The resulting amplicons were purified and sequenced directly by the Sanger method (ON228170).

To confirm the infectivity of the isolate, the original soybean tissue samples that were stored at −80°C were immediately ground in phosphate buffer and used for sap inoculation of soybean plants of cv. TMG 7063. After two weeks, the inoculated plants developed symptoms of mosaic and chlorotic spots (Fig. 4), and the results were confirmed by RT-PCR. Leaf tissue from these plants was used as an inoculum source for further host range and aphid transmission assays.



Figure 4 – Symptoms of mosaic and chlorotic spots on young leaves of soybean infected with passiflora virus Y (PaVY-Br).

Symptomatic soybean leaves were used to produce infectious plant extracts needed for further inoculations. Sets of 10 plants of 14 indicator hosts were tested: *Nicotiana clevelandii*, *N. tabacum* cv. Xanthi, *N. tabacum* cv. Havana, *N. tabacum* cv. TNN, *N. tabacum* cv. Samsun, *N. rustica* var. *rustica* 'Turkish Tabacco', *N. glutinosa*, common bean (*Phaseolus vulgaris*), zucchini (*Cucurbita pepo*), tomato (*Solanum lycopersicum* cv. Santa Clara), passion fruit (*Passiflora edulis*), sweet pepper (*Capsicum annuum* cv. Magali), and sesame (*Sesamum indicum*). Soybean plants were used as positive controls. Two plants of each indicator host were mock-inoculated and used as negative controls. All plants were maintained in a greenhouse at 25 ± 2°C, and symptoms were visually assessed weekly.

The PaVY-Br isolate induced chlorotic spots and systemic mosaic on soybean two weeks after inoculation. No symptoms were observed in Nicotiana species. Sesame and passion fruit showed chlorotic local lesions.

Previous reports indicated that PaVY induced symptoms of mosaic and leaf deformation in the perennial wild legume siratro (*Macroptilium atropurpureum*) and plant stunting in rhyncho (*Rhynchosia minima*) and wild mung bean (*Vigna trilobata*) [15]. Symptoms of chlorotic spots were associated with PaVY-infected butterfly pea (*Clitoria ternatea*) and stinking passionflower (*P. foetida*) [15]. Chlorotic spots, mosaic symptoms, and ringspot in stinking passionflower were also observed in this study.

The PaVY-Br isolate was successfully transmitted by *Myzus persicae*. Adults of *M. persicae* and *Aphis gossypii* were made to fast for one hour before feeding on PaVY-infected soybean leaves for one hour for virus acquisition, followed by overnight inoculation access (10 aphids per plant) to 30 soybean plants

(cv. TMG 7063). The aphids were sprayed with insecticide (flupyradifurone) after inoculation access. Plants inoculated by *M. persicae* developed symptoms two weeks post-inoculation, and PaVY infection was confirmed by RT-PCR in 3 out of 15 plants. Three weeks after aphid inoculation, soybean plants inoculated by *M. persicae* developed the same chlorotic symptoms observed in the soybean leaves that served as the inoculum source. Previous research demonstrated the transmission of PaVY by *A. gossypii* to *Passiflora edulis* [14]. However, PaVY-Br was not transmitted by *A. gossypii* (0/15 plants) in our study.

PaVY has been found naturally infecting Passiflora species [13,14,15, 17]. Our study found PaVY naturally infecting soybean plants for the first time. How PaVY disperses naturally through soybean plants requires further study. The results presented here point towards the involvement of an aphid vector and suggest that PaYV may spread from passion fruit to soybean plants and vice-versa. It is worth mentioning that PaYV had never been reported in Brazil, even infecting passion fruit. The presence of PaYV was reported to the Brazilian Ministry of Agriculture, Livestock and Food Supply (MAPA) as required by federal law. Its distribution in soybean production areas in Brazil needs to be studied further.

Data availability

The nucleotide sequences reported in this manuscript have been deposited in the GenBank database under accession numbers MZ190342, MZ190341, and ON228170 and are available in the Electronic Supplementary Material.

REFERENCES

1. FAO (2021) Food and Agriculture Organization of the United Nations. http://www.fao.org/

2. Hill JH, Whitham SA (2014) Control of virus diseases in soybeans, 1st edn. Elsevier Inc., Amsterdam

3. dos Figueira AR, Geraldino-Duarte PS, Pinzon Nunez AM et al (2019) Characterization of soybean yellow shoot virus, a new member of the family Potyviridae infecting soybean plants in Brazil. Plant Dis 103:1172–1180

4. Zanardo LG, Silva FN, Bicalho AAC et al (2014) Molecular and biological characterization of Cowpea mild mottle virus isolates infecting soybean in Brazil and evidence of recombination. Plant Pathol 63:456–465

5. Almeida ÁMR, Sakai J, Hanada K et al (2005) Biological and molecular characterization of an isolate of tobacco streak virus obtained from soybeans in Brazil. Fitopatol Bras 30:366–373

6. Fontes MG, da Silva GFA, Lima MF et al (2019) First report of groundnut ringspot orthotospovirus infecting soybeans in Brazil. Embrapa Hortaliças-Nota Técnica/Nota Científica (ALICE)

7. Souto ER, Almeida ÁMR, Bianchini A et al (2002) Análise molecular de segmento do RNA2 de comovirus isolados de soja no estado do Paraná. Fitopatol Bras 27:525–527. https://doi.org/10.1590/S0100-41582002000500014

8. Fernandes FR, Cruz ARR, Faria JC et al (2009) Three distinct begomoviruses associated with soybean in central Brazil. Arch Virol 154:1567–1570

9. Almeida AM, Bianchini A, Costa AS, Vega J (1982) Calico mosaic, a new virus disease of soybean in Brazil. Fitopatol Bras 7:133–138

10. Coco D, Calil IP, Brustolini OJB et al (2013) Soybean chlorotic spot virus, a novel begomovirus infecting soybean in Brazil. Arch Virol 158:457–462. https://doi.org/10.1007/S00705-012-1499-0

11. Gibbs A, Mackenzie A (1997) A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. J Virol Methods 63:9–16

12. Mota LDC, dellaVecchia MGS, Gioria R et al (2004) Pfaffia mosaic virus: a new potyvirus found infecting Pfaffia glomerata in Brazil. Plant Pathol 53:368–373

13. Chen B, Wu D, Zheng H et al (2021) Complete genome sequence of passiflora virus Y infecting passion fruit in China. Adv Virol 166:1489–1493. https://doi.org/10.1007/s00705-021-05013-1

14. Parry JN, Davis RI, Thomas JE (2004) Passiflora virus Y, a novel virus infecting Passiflora spp in Australia and the Indonesian Province of Papua. Australasian Plant Pathol 33:423. https://doi.org/10.1071/AP04042

15. Coutts BA, Kehoe MA, Webster CG et al (2011) Indigenous and introduced potyviruses of legumes and Passiflora spp. From Australia: biological properties and comparison of coat protein nucleotide sequences. Adv Virol 156:1757–1774. https://doi.org/10.1007/s00705-011-1046-4

16. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PloS ONE 9:e108277. https://doi.org/10.1371/34assif.pone.0108277

17. Nakasato K, Fujioka S, Sugawara Y et al (2020) First detection of two potyviruses, uraria mosaic virus and passiflora mosaic virus Y, from passionfruit in Japan. J Gen Plant Pathol 86:401–404. https://doi.org/10.1007/s10327-020-00932-4

18. Inoue-Nagata AK et al (2022) ICTV virus taxonomy profile: Potyviridae 2022. J Gen Virol 103(5):001738

19. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

20. Nylander J (2004) MrModeltest v. 2.2 b. Department of Systematic Zoology, Uppsala University, Sweden

Funding

MRRJ was a recipient of CNPq and CAPES-PrInt Ph.D. fellowships (grant numbers: 140130/2019-5 and 88887.570292/2020-00); FBS is a recipient of a FAPESP Ph.D. fellowship (grant number: 2019/24642-8); AMG is a recipient of a FAPESP post-doctoral fellowship (grant number: 2019/18639-4); RKS hold CNPq research fellowship (grant number: 303411/2018-0). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Contributions

MRRJ and FBS initiated the project; MRRJ, AMN, CM, FMOC, MAP, and RKS analyzed the data; MRRJ, DMN, and RKS contributed to writing the paper. All authors have read and agreed to the published version of the manuscript.

Corresponding author

Correspondence to Marcos Roberto Ribeiro-Junior

Ethics declarations

Conflict of interest: The authors have no relevant financial or non-financial interests to disclose.

Ethics approval: This article does not contain any studies with human participants or animals performed by any of the authors.

CHAPTER 2

AN ATTEMPT TOWARD THE GLOBAL MONITORING OF SOYBEAN VIRUSES USING MICROBE FINDER (MiFi®)*

Marcos Roberto Ribeiro-Junior[1,2,3], Francisco Manuel Ochoa-Corona[1,2], Andres Espindola[1,2], Daniele Maria Nascimento[1,2] and Renate Krause-Sakate[3].

[1]Institute for Biosecurity and Microbial Forensics, Oklahoma State University, Stillwater, OK, USA.
[2]Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, USA.
[3]Department of Plant Protection, São Paulo State University, UNESP, School of Agriculture, Botucatu, SP, Brazil.
Corresponding author (e-mail: maribei@okstate.edu) https://orcid.org/0000-0002-5024-1801

Abstract

Soybean (*Glycine max* (L.) Merr.) is an important crop worldwide being a cornerstone source of oilseed and protein. Brazil and the U.S. are the two largest producers in the world. Nevertheless, each year, soybean growers lose a significant volume of yields to diseases caused by viruses. This is the reason why precise diagnosis is crucial for management. To date, 46 virus species have been described as infecting soybean worldwide. Plant virus identification based on serological and traditional molecular techniques is time-consuming and does not allow simultaneous identification of all viruses present in a single sample. Microbe Finder (MiFi®) is a web application for the quick detection and identification of

* Chapter written according to the journal **Phytopathology®**

predetermined pathogen species. MiFi® uses highly curated, target-specific electronic probes (e-probes) which allow querying raw, unassembled High-throughput sequencing (HTS) metagenomic data. E-probes ranging from 19-60 nucleotides in length were designed for the 46 viruses reported to infect soybean using MiProbe, another application within the MiFi® platform. The curation and validation of the designed e-probes were performed using mock metagenomic databases assembled with each of the virus reference positive controls and sourced host genome sequences. The results obtained *in silico* allowed us to validate the use of e-probes for the detection of 46 viruses in metagenomic soybean data. *In vitro* analyzes were performed using Illumina HTS data obtained from soybean plants collected in Brazil. Nine viruses were detected in the Brazilian soybean samples. The results were confirmed by PCR, RT-PCR, and mapping reads to virus reference genomes using Geneious software. This research proposes a sequence-based method for rapid (less than an hour) detection and identification of all viruses so far described infecting soybean. The curated soybean virus e-probes are free of access for diagnostics through MiDetect within the MiFi® platform.

Keywords: plant virus, detection, high-throughput sequencing, e-probe

2.1 INTRODUCTION

The relevance of soybean as a crop is described by adjectives, golden bean, super legume, the crop of the planet, or just the miracle bean. Soybean is a staple crop worldwide and is used for food, feed, and fuel (Rahman et al. 2022).

The global soybean output 2021/22 reached 350.72 million metric tons, from which 35.64% (125.00 million tons) was produced in Brazil, the leading producer, and soybean exporter worldwide. Brazil produces 61.69% of the whole exported production (77 million tons), closely followed by the United States, with 34.41% of world production (120.71 million tons) (USDA 2022).

Plant viruses represent a threat to soybean production. At least 46 viruses are known to infect soybean worldwide (Hill and Whitham 2014). Among these viruses, soybean mosaic virus (SMV), naturally transmitted by aphid species and infected seeds, can cause yield losses of up to 86% and it is present worldwide (Hajimorad et al. 2018).

Rapid and precise diagnosis is the first step for the detection and abatement of the spreading of new viruses in countries where viruses are not reported yet (Espindola and Cardwell 2021). In 2022, a new potyvirus (PaVY) was identified in soybean crops in Brazil, emphasizing the need to improve current detection methods (Ribeiro-Junior et al. 2022).

Biosecurity can prevent the spreading of diseases by minimizing risks of entry and infection of susceptible crops, monitoring, restricting, and controlling the movement of pathogens across national and regional borders (Oliver 2022). Sequence-based detection technology is a new alternative for use in routine detection by plant quarantine agencies, which could overcome the screening of numerous viruses simultaneously if equipped with this technology (Maree et al.

2018). In this scenario, high-throughput sequencing (HTS), also known as next-generation sequencing, has become an indispensable tool for virus diagnosis based on the possibility of broad virus detection in samples (Kutnjak et al. 2021). However, the volume of data generated by HTS becomes a limiting factor, extending the time needed to make a final diagnostic decision (Stobbe et al. 2013).

This gap between HTS and pathogen detection is addressed by E-probe Diagnostic Nucleic acid Analysis (EDNA), which is hosted in the platform called Microbe Finder (MiFi®) (Espindola and Cardwell 2021). MiFi® is composed of two parts (MiProbe and MiDetect) and uses EDNA to query the metagenome raw data and identify known pathogen sequences (Espindola and Cardwell 2021).

This study describes the development of a bioinformatics tool combining HTS and bioinformatics to detect multiple viruses previously reported infecting soybean worldwide using MiFi®, which provides quick detection of known species in unassembled HTS metagenomic data. E-probes need to be designed, curated, and validated *in silico* and *in vitro* for this purpose. This study describes the development of *in silico*-curated e-probes for 46 virus species for the global detection of soybean viruses.

2.2 MATERIAL AND METHODS

A MiProbe pipeline described by Espindola and Cardwell (2021) was used to develop and validate e-probes to detect all viruses previously described as infecting soybean in the world. Whole genomes of all the viruses (and their isolates) were downloaded from NCBI public database (Table 1).

Table 1 - Genus, virus species, acronym/e-probe ID, genome size (Kb), and reference genome of all viruses described infecting soybean to present worldwide used in this study.

Genus	Viral species	Acronym/ E-probe ID	Genome size (Kb)	Reference genome
Alfamovirus	*Alfalfa mosaic virus*	AMV	3.64 2.59 2.04	NC_001495.1 NC_002024.2 NC_002025.1
Begomovirus	*Abutilon mosaic virus*	AbMV	2.63 2.59	NC_001928.2 NC_001929.2
	Bean golden mosaic virus	BGMV	2.62 2.58	NC_004042.1 NC_004043.1
	Euphorbia mosaic virus	EuMV	2.61 2.6	NC_008304.1 NC_008305.1
	Horsegram yellow mosaic virus	HgYMV	2.73 2.68	NC_005635.1 NC_005636.1
	Mungbean yellow mosaic virus	MYMV	2.72 2.66	NC_001983.1 NC_001984.1
	Okra mottle virus	OMV	2.66 2.55	NC_011181.1 NC_011182.1
	Rhynchosia golden mosaic virus	RhGMV	2.55 2.6	NC_010293.1 NC_010294.1
	Sida micrantha mosaic virus	SimMV	2.68 2.66	NC_005330.1 NC_005331.1

	Soybean chlorotic spot virus	SoCSV	2.62 2.59	NC_018457.1 NC_018456.1
	Soybean crinkle leaf virus	SCLV	2.74	NC_003357.1
Bromovirus	*Cowpea chlorotic mottle virus*	CCMV	3.17 2.77 2.17	NC_003543.1 NC_003541.1 NC_003542.1
Carlavirus	*Cowpea mild mottle virus*	CPMMV	8.13	NC_014730.1
Comovirus	*Bean pod mottle virus*	BPMV	6 3.66	NC_003496.1 NC_003495.1
	Broad bean true mosaic virus	BBTMV	6.07 3.53	NC_022004.1 NC_022006.1
	Cowpea mosaic virus	CPMV	5.89 3.48	NC_003549.1 NC_003550.1
	Cowpea severe mosaic virus	CPSMV	5.96 3.73	NC_003545.1 NC_003544.1
Cucumovirus	*Cucumber mosaic virus*	CMV	3.36 3.05 2.22	NC_002034.1 NC_002035.1 NC_001440.1
	Peanut stunt virus	PSV	3.36 2.95 2.19	NC_002038.1 NC_002039.1 NC_002040.1
Enamovirus	*Pea enation mosaic virus*	PEMV	5.71	NC_003629.1
Fabavirus	*Broad bean wilt virus*	BBWV	5.82 3.45	NC_005289.1 NC_005290.1
Gammacarmovirus	*Soybean yellow mottle mosaic virus*	SYMMV	4.01	NC_011643.1

Ilarvirus	*Tobacco streak virus*	TSV	3.49	NC_003844.1
			2.93	NC_003842.1
			2.21	NC_003845.1
Luteovirus	*Bean leaf roll virus*	BLRV	5.96	NC_003369.1
	Soybean dwarf virus	SbDV	5.85	NC_003056.1
Nepovirus	*Tobacco ringspot virus*	TRSV	7.51	NC_005097.1
			3.93	NC_005096.1
	Tomato ringspot virus	ToRSV	8.21	NC_003840.1
			7.27	NC_003839.2
Orthotospovirus	*Groundnut ringspot virus*	GRSV	8.63	NC_043503.1
			3.4	NC_043502.1
			0.78	NC_043504.1
	Soybean vein necrosis virus	SVNV	9.01	NC_055179.1
			4.96	NC_055180.1
			2.6	NC_055178.1
	Tomato spotted wilt virus	TSWV	8.9	NC_002052.1
			4.82	NC_002050.1
			2.92	NC_002051.1
Potyvirus	*Azuki bean mosaic virus*	AzMV	No full genome available	
	Bean common mosaic virus	BCMV	9.99	NC_003397.1
	Bean yellow mosaic virus	BYMV	9.53	NC_003492.1
	Blackeye cowpea mosaic virus	BlCMV	9.92	AY575773.1
	Passiflora virus Y	PaVY	9.67	MW165064
	Passion fruit woodiness virus	PWV	9.62	NC_014790.2
	Peanut mottle virus	PMotV	9.71	NC_002600.1
	Peanut stripe virus	PStV	10.09	U34972.1

	Soybean mosaic virus	SMV	9.59	NC_002634.1
	Soybean yellow shoot virus	SoyYSV	9.05	MH890547.1
Sobemovirus	*Southern bean mosaic virus*	SBMV	4.13	NC_004060
	Soybean yellow common mosaic virus	SYCMV	4.15	NC_016033.1
Soymovirus	*Soybean chlorotic mottle virus*	SbCMV	8.17	NC_001739.2
Tobamovirus	*Sunn-hemp mosaic virus*	SHMV	6.43	D84000.1
	Tobacco mosaic virus	TMV	6.39	NC_001367.1
Tobravirus	*Tobacco rattle virus*	TRV	6.79 3.85	NC_003805.1 NC_003811.1

2.2.1 Raw e-probe design

Forty-six viruses (Table 1) were selected as target genomes for e-probe generation. For each of them, a set of near neighbors (closely related viruses) were included, not predicted to be detected, once only the target genome would be expected to be positive or negative.

Complete or partial genomes of all or most virus isolates were concatenated into a single multi-fasta file, named "target". The near neighbors were also concatenated (named "near neighbors"). These two files selected were uploaded to the MiFi® platform and the MiProbe function was used to generate short e-probes sequences, ranging from 20nt to 60nt (lengths recommended for virus detection) (Espindola and Cardwell 2021).

2.2.2 E-probe curation: specificity and cross-reactivity

Curation is defined as the process of eliminating any e-probe sequence that matches with another pathogen, causing a false positive hit (Espindola and Cardwell 2021). Initially, a higher number of e-probe sequences (raw e-probes) is generated for the target virus. Raw e-probes were tested to assess their specificity and cross-reactivity. Since each e-probe set must match specifically only to the target virus, each e-probe set was tested against near neighbors, all other soybean viruses, and the soybean genome (host) sequence. MetaSim software was used for simulations (Richter et al. 2008). These simulations were performed for each of the 46 soybean viruses in the database, by creating a mock metagenome database with soybean genome, near neighbor, and other soybean virus sequences.

Each e-probe set was individually tested to assess cross-reactivity. E-probes that were less specific (usually, short e-probes, with lengths ranging from

20 to 30nt) to the targeted virus were removed from the set. When available, longer e-probes are more specific, but those presenting cross-reactivities were removed from the set.

Subsequently, the developed e-probes were searched through the National Center for Biotechnology Information (NCBI) nt database using the BLASTn algorithm. E-probes that matched with non-target organisms in the NCBI nt database, causing a false positive hit, were removed.

2.2.3 Limit of detection and validation *in silico*

The limit of detection (LoD), defined as the lowest pathogen concentration, was established for all the e-probe sets. For this purpose, *in silico* validation was performed using MetaSim (Richter et al. 2008) simulated HTS data composed of soybean genome and different concentrations of each of the 46 viruses (ranging from 1 to 5,000 pathogen reads). Each concentration was made of 10 replicates and at least nine of them must be positive to consider the lower detectable ratio as LoD. The simulations were capped at 1 million total reads and the results were transformed into Minimum Relative Abundance of Pathogen Reads to Detect (MRAD) of pathogen reads. For example, if the LoD of the virus is 10 pathogen reads when transformed to relative abundance, it is 0.001% for 1 million reads metagenomes.

2.2.4 Validation with field samples

Symptomatic leaf samples were collected in soybean fields located in São Paulo State (Brazil) during 2018 – 2020. Total RNA was extracted from symptomatic leaves using Purelink Viral RNA/DNA Mini Kit (Thermo Fisher Scientific) generating seven samples used for the construction of a cDNA library utilizing the Complete ScriptSeq Kit (Epicenter, Illumina). The transcriptome sequencing was performed with the Illumina HiSeq2500 platform using HiSeq Flow Cell v4 and HiSeq SBS v4 (Illumina) in paired-end mode (2×100 bp).

The seven generated Illumina outputs were analyzed with the newly cured e-probes and the gold standard method to assure the validity of the performance of the e-probes.

The metagenomes were uploaded to the MiFi® platform, and the MiDetect function was used to set the suggested parameters for virus detection (Espindola and Cardwell 2021). E-value was set to 1e-1 and the minimum number of reads was set to sensitive (10 hits).

To make possible the comparison of the results, the generated metagenomes were paired, trimmed and *de novo* assembled using the CLC Genomics Workbench software v9.0.3. To demonstrate the efficiency of MiDetect, all viruses detected with MiFi had their reference genome downloaded from NCBI. The HTS reads obtained were used to map the reference virus genome using Geneious software v 11.1.6., proving the presence of virus reads in soybean Illumina HTS samples.

To demonstrate that no other viruses were present in the sample and escaping MiFi detection, the obtained contigs were subjected to BLASTn against the NCBI virus genome database using Geneious software v 11.1.6.

2.3 RESULTS

2.3.1 E-probe design and curation

For some viruses, such as CMV, the number of generated e-probes sequences was reduced by 50% after curation (Table 3). The number of e-probes for AbMV, BPMV, BBTMV, CPSMV, SYMMV, TSV, ToRSV, AzMV, PaVY, SoyYSV, SbCMV, and SHMV remained the same before and after curation process.

E-probes were selected to be 20-60nt long. For BCMV, BICMV and PStV, e-probes of variable length had to be designed manually examining genomic sequences, as it was not possible to obtain a set of specific e-probes with fixed length targeting a single virus. In this case, e-probes of variable sizes were obtained through the creation of a consensus sequence of the targeted virus species, and subsequent alignments made with sequences obtained from near neighbors showed the most appropriate regions (specific regions) for e-probe design.

Table 3 – Number of raw e-probes and e-probes obtained after curation. The number of e-probes in each virus set is unique for the virus species and ranges from 19 to 60 nt.

Viral species	**Raw e-probes**	**Curated e-probes**	**E-probe length (nt)**
Alfalfa mosaic virus (AMV)	1197	982	40
Abutilon mosaic virus (AbMV)	19	19	40
Bean golden mosaic virus (BGMV)	414	299	60
Euphorbia mosaic virus (EuMV)	73	26	60
Horsegram yellow mosaic virus (HgYMV)	29	28	60
Mungbean yellow mosaic virus (MYMV)	220	114	40
Okra mottle virus (OMV)	14	10	30
Rhynchosia golden mosaic virus (RhGMV)	60	34	60
Sida micrantha mosaic virus (SimMV)	75	73	40
Soybean chlorotic spot virus (SoCSV)	4	2	40
Soybean crinkle leaf virus (SCLV)	12	4	30
Cowpea chlorotic mottle virus (CCMV)	87	86	40
Cowpea mild mottle virus (CPMMV)	506	483	40
Bean pod mottle virus (BPMV)	284	284	40
Broad bean true mosaic virus (BBTMV)	64	64	30
Cowpea mosaic virus (CPMV)	42	37	40
Cowpea severe mosaic virus (CPSMV)	99	99	40
Cucumber mosaic virus (CMV)	7472	3884	40
Peanut stunt virus (PSV)	252	250	40
Pea enation mosaic virus (PEMV)	305	304	40
Broad bean wilt virus (BBWV)	1282	1120	40
Soybean yellow mottle mosaic virus (SYMMV)	98	98	60
Tobacco streak virus (TSV)	444	444	40
Bean leaf roll virus (BLRV)	125	99	40
Soybean dwarf virus (SbDV)	663	662	40
Tobacco ringspot virus (TRSV)	145	135	60
Tomato ringspot virus (ToRSV)	22	22	40
Groundnut ringspot virus (GRSV)	83	27	60
Soybean vein necrosis virus (SVNV)	105	104	60
Tomato spotted wilt virus (TSWV)	1092	1050	40
Azuki bean mosaic virus (AzMV)	4	4	20
Bean common mosaic virus (BCMV)	15	15[1]	19 (2)[2] 20 (1)[2] 21 (1)[2] 22 (1)[2] 24 (1)[2] 25 (3)[2] 27 (2)[2] 29 (1)[2] 31 (2)[2] 40 (1)[2]
Bean yellow mosaic virus (BYMV)	601	374	60

Blackeye cowpea mosaic virus (BlCMV)	15	15[1]	19 (2)[2] 20 (1)[2] 21 (1)[2] 24 (2)[2] 25 (3)[2] 27 (2)[2] 28 (1)[2] 30 (1)[2] 31 (1)[2] 40 (1)[2]
Passiflora virus Y (PaVY)	9	9	60
Passion fruit woodiness virus (PWV)	58	42	60
Peanut mottle virus (PMotV)	107	94	60
Peanut stripe virus (PStV)	25	25[1]	19 (4)[2] 21 (2)[2] 22 (3)[2] 23 (1)[2] 24 (2)[2] 25 (1)[2] 26 (1)[2] 28 (1)[2] 29 (1)[2] 30 (1)[2] 31(1)[2] 35 (1)[2] 36 (1)[2] 40 (5)[2]
Soybean mosaic virus (SMV)	199	196	40
Soybean yellow shoot virus (SoyYSV)	14	14	30
Southern bean mosaic virus (SBMV)	73	64	40
Soybean yellow common mosaic virus (SYCMV)	170	122	40
Soybean chlorotic mottle virus (SbCMV)	48	48	40
Sunn-hemp mosaic virus (SHMV)	30	30	40
Tobacco mosaic virus (TMV)	442	111	40
Tobacco rattle virus (TRV)	767	759	40

[1]E-probes manually designed. [2]Number of e-probes manually designed for the respective length.

2.3.2 Limit of detection and validation *in silico*

The minimum pathogens reads serially diluted in a million host reads ranged from one to 5000, or 0.0001% to 0.5% (Table 4).

Table 4 – Minimum Relative Abundance of Pathogen Reads to Detect (MRAD), calculated by *in silico* simulations over a million total reads.

Viral species	**E-Probe Sequences**	**MRAD (LoD)**
Alfalfa mosaic virus (AMV)	982	0.0003%
Abutilon mosaic virus (AbMV)	19	0.002%
Bean golden mosaic virus (BGMV)	299	0.001%
Euphorbia mosaic virus (EuMV)	26	0.001%
Horsegram yellow mosaic virus (HgYMV)	28	0.0011%
Mungbean yellow mosaic virus (MYMV)	114	0.001%
Okra mottle virus (OMV)	10	0.0025%
Rhynchosia golden mosaic virus (RhGMV)	34	0.001%
Sida micrantha mosaic virus (SimMV)	73	0.001%
Soybean chlorotic spot virus (SoCSV)	2	0.5%
Soybean crinkle leaf virus (SCLV)	4	0.45%
Cowpea chlorotic mottle virus (CCMV)	86	0.001%
Cowpea mild mottle virus (CPMMV)	483	0.0007%
Bean pod mottle virus (BPMV)	284	0.001%
Broad bean true mosaic virus (BPMV)	64	0.0006%
Cowpea mosaic virus (CPMV)	37	0.001%
Cowpea severe mosaic virus (CPSMV)	99	0.001%
Cucumber mosaic virus (CMV)	3884	0.0001%
Peanut stunt virus (PSV)	250	0.001%
Pea enation mosaic virus (PEMV)	304	0.001%
Broad bean wilt virus (BBWV)	1120	0.0009%
Soybean yellow mottle mosaic virus (SYMMV)	98	0.0017%
Tobacco streak virus (TSV)	444	0.0008%
Bean leaf roll virus (BLRV)	99	0.001%
Soybean dwarf virus (SbDV)	662	0.0006%
Tobacco ringspot virus (TRSV)	135	0.0009%
Tomato ringspot virus (ToRSV)	22	0.013%
Groundnut ringspot virus (GRSV)	27	0.0025%
Soybean vein necrosis virus (SVNV)	104	0.0015%
Tomato spotted wilt virus (TSWV)	1050	0.0009%
Azuki bean mosaic virus (AzMV)	4	0.08%
Bean common mosaic virus (BCMV)	15	0.003%
Bean yellow mosaic virus (BYMV)	374	0.001%
Blackeye cowpea mosaic virus (BICMV)	15	0.005%
Passiflora virus Y (PaVY)	9	0.004%
Passion fruit woodiness virus (PWV)	42	0.0014%
Peanut mottle virus (PMotV)	94	0.0011%
Peanut stripe virus (PStV)	25	0.0040%
Soybean mosaic virus (SMV)	196	0.0015%
Soybean yellow shoot virus (SoyYSV)	14	0.0035%
Southern bean mosaic virus (SBMV)	64	0.0007%
Soybean yellow common mosaic virus (SYCMV)	122	0.0006%
Soybean chlorotic mottle virus (SbCMV)	48	0.0015%
Sunn-hemp mosaic virus (SHMV)	30	0.0025%
Tobacco mosaic virus (TMV)	111	0.001%
Tobacco rattle virus (TRV)	759	0.0008%

The minimum relative abundance of pathogen reads to detect was lower when the number of e-probes was higher. For cucumber mosaic virus, the e-probe set containing 3,884 e-probe sequences and just one pathogen read was needed which, transformed to a relative abundance, is 0.0001% (Table 4).

2.3.3 Validation with field samples

Nine viruses were detected with the e-probes developed with MiDetect and results are presented in Table 5, assuring that the respective e-probes sets will perform well, being able to be adopted as a regular diagnostic method. All the other viruses were not present in these Brazilian samples tested.

Table 5 - Viruses detected in samples collected in Brazil with the curated e-probes in MiDetect

Genus	Species names	E-probe ID	MiFi® Diagnostic	Samples
Begomovirus	*Bean golden mosaic virus*	BGMV	Positive	3
	Sida micrantha mosaic virus	SimMV	Positive	5
Bromovirus	*Cowpea chlorotic mottle virus*	CCMV	Positive	1
Carlavirus	*Cowpea mild mottle virus*	CPMMV	Positive	1, 2, 3, 4, 5, 6, 7
Cucumovirus	*Cucumber mosaic virus*	CMV	Positive	1, 3, 4, 5, 6, 7
Irlavirus	*Tobacco streak virus*	TSV	Positive	3, 4
Orthotospovirus	*Groundnut ringspot virus*	GRSV	Positive	1, 3, 5, 6, 7
Potyvirus	*Bean common mosaic virus*	BCMV	Positive	2
	Passiflora virus Y	PaVY	Positive	7

Brazilian samples confirmed as positive had the corresponding reads successfully mapped to reference genomes using Geneious software (Table 6), proving the presence of virus reads in the samples. The concentration of reads in the samples ranged from 4,068 (GRSV, sample 5) to 9,391,113 (CPMMV, sample 1).

Table 6 - Number of reads-matches with reference virus genomes and percentage of genome coverage.

	Samples (total reads per HTS sample)						
Positive e-probes and their reference virus accession number	1 (27,674,874)	2 (33,944,812)	3 (30,636,554)	4 (33,774,040)	5 (35,606,110)	6 (41,590,402)	7 (37,621,186)
BGMV: NC_004042.1, NC_004043.1	-	-	3,213[a] 92.4%[b] 194[a] 99.3%[b]	-	-	-	-
SimMV: NC_005330.1, NC_005331.1	-	-	-	-	938[a] 100%[b] 2,446[a] 100%[b]	-	-
CCMV: NC_003543.1, NC_003541.1, NC_003542.1	1,365[a] 98.3%[b] 2,785[a] 98.1%[b] 15,068[a] 97,8%[b]	-	-	-	-	-	-
CPMMV: NC_014730.1	9,391,113[a] 100%[b]	6,394,631[a] 100%[b]	3,161,251[a] 100%[b]	1,1686,900[a] 100%[b]	1,152,927[a] 96.4%[b]	2,922,714[a] 100%[b]	4,815,528[a] 100%[b]
CMV: NC_002034.1, NC_002035.1, NC_001440.1	320[a] 97.3%[b] 86[a] 86.5%[b] 16,175[a] 98.3%[b]	-	36,065[a] 57%[b] 100,380[a] 100%[b] 14,197[a] 100%[b]	36,065[a] 75.2%[b] 108,628[a] 100%[b] 102,457[a] 100%[b]	51,855[a] 93.7%[b] 87,864[a] 100%[b] 270,620[a] 79.6%[b]	38,176[a] 96.3%[b] 77,544[a] 100%[b] 526,058[a] 87.4%[b]	20,727[a] 90.4%[b] 471,541[a] 100%[b] 741,888[a] 100%[b]
TSV: NC_003844.1,	-	-	488,593[a] 99.8%[b]	1,983[a] 52.3%[b]	-	-	-

NC_003842.1, NC_003845.1			397,623[a] 78%[b] 1,428,425[a] 100%[b]	54,465[a] 77.4%[b] 10,793[a] 85.9%[b]			
GRSV: NC_043503.1, NC_043502.1, NC_043504.1	23,767[a] 100%[b] 20,308[a] 100%[b] 150,485[a] 100%[b]	-	33,405[a] 100%[b] 77,224[a] 100%[b] 135,025[a] 100%[b]	-	26[a] 83%[b] 23[a] 41.4%[b] 4,019[a] 56.6%[b]	669,918[a] 100%[b] 510,703[a] 100%[b] 2,806,220[a] 100%[b]	597,712[a] 100%[b] 626,610[a] 100%[b] 2,756,858[a] 100%[b]
BCMV: NC_003397.1	-	91,770[a] 100%[b]	-	-	-	-	-
PaVY: MW165064	-	-	-	-	-	-	2,231,517[a] 100%[b]

[a] number of reads mapped to reference genome sequences; [b] percentage of genome coverage; -virus absent in the sample.

Samples contigs that were submitted to BLASTn analyses against the NCBI virus genome database demonstrate that no other soybean viruses, in addition to those previously detected, were detected in the samples.

2.4 DISCUSSION

Here we successfully obtained e-probes for the detection of 46 viruses reported to infect soybean in the world, contributing to the rapid diagnosis of these different viruses using the MiFi® platform. Usually, bioinformatics knowledge is required to analyze the HTS data. MiFi® platform has a simple interface that eliminates advanced code skills. Furthermore, compared with other bioinformatics tools such as Metaphlan and Kraken (Wood and Salzberg 2014; Beghini et al. 2021), MiFi® was more sensitive to detecting some pathogens, and results were obtained within 30 minutes compared to an average time of 1h 12 min. and 2h 42 min, required by Meaphlan and Kraken, respectively (Espindola and Cardwell 2021).

MiFi® platform has already been used to detect cucurbit viruses (Zuniga et al. 2017), water-borne plant viruses (Zuniga et al. 2018), viruses associated with Citrus leprosis syndrome (Espindola et al. 2019), grapevine pathogens (Espindola et al. 2019; Espindola and Cardwell 2021), blueberries pathogens (Bocsanczy et al. 2018), fungi and oomycetes (Espindola et al. 2015; Proano et al. 2018) and aflatoxin production in the soil (Espindola et al. 2018).

The pipeline proposed by Espindola and Cardwell (2021) recommends *in silico, in vitro,* and validation with field samples. We are reporting the validation of nine out of the 46 viruses because it was not possible to obtain positive controls or field samples for all the viruses, since some of them are exotic pests to EUA and

Brazil. The samples collected in Brazil were infected with nine virus species: BGMV, SiMV, CPCMV, CPMMV, CMV, TSV, GRSV, BCMMV, and PaVY. For these viruses, the e-probes were validated *in vitro*, demonstrating their specificity to detect the viruses. Espindola and Cardwell (2021) point out that the level of validation achieved by EDNA depends on the availability of positive controls, funding, laboratory materials, field samples, and others. Since the validated soybean e-probes developed here are available to other researchers and diagnosticians on MiDetect, it needs to be tested using vegetable samples composed of different mixed infections of viruses. This can contribute to continually validating MiFi and will help to improve the method.

The level of sensitivity (LoD) and specificity (when the e-probe sequence matches only with the target) also depend on the goal. Biosecurity laboratories require both to detect exotic or high-risk pathogen species (Espindola et al. 2022). In our study, the limit of detection in silico ranged from 0.0001% to 0.5% when using one million reads.

Finally, we can conclude that with the global exchanges of seeds and seedlings, quick detection of pathogens is a key factor to prevent the dissemination of important pests worldwide and that the tool developed here for the detection of the 46 viruses infecting soybean can be very helpful and faster in preventing the entrance of exotic viruses in a country or region of the world.

REFERENCES

Beghini F, McIver LJ, Blanco-Míguez A, et al (2021) Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife 10:. https://doi.org/10.7554/ELIFE.65088

Bocsanczy AM, Espindola A, Norman DJ, Cardwell KF (2018) E-probes development for rapid, sensitive and specific pathogen detection in blueberries. In: International Congress of Plant Pathology (ICPP) 2018: Plant Health in A Global Economy. APSNET

Espindola A, Roy A, Mavrodieva VA, Cardwell KF (2019) E-probe: A new diagnostic tool for detection of Dichorhaviruses associated with Citrus leprosis syndrome. Plant Health 2019

Espindola AS, Cardwell KF (2021) Microbe finder (Mifi®): Implementation of an interactive pathogen detection tool in metagenomic sequence data. Plants 10:1–14. https://doi.org/10.3390/plants10020250

Espindola AS, Cardwell KF, Martin FN, et al (2022) A step towards validation of high-throughput sequencing for the identification of plant pathogenic oomycetes. Phytopathology®. https://doi.org/10.1094/PHYTO-11-21-0454-R

Espindola AS, Schneider W, Cardwell KF, et al (2018) Inferring the presence of aflatoxin-producing Aspergillus flavus strains using RNA sequencing and electronic probes as a transcriptomic screening tool. PLoS ONE 13:. https://doi.org/10.1371/JOURNAL.PONE.0198575

Espindola AS, Schneider WL, Marek S, et al (2015) Metagenome data mining for eukaryotic plant pathogens. In: PHYTOPATHOLOGY. AMER PHYTOPATHOLOGICAL SOC 3340 PILOT KNOB ROAD, ST PAUL, MN 55121 USA, p 41

Hajimorad MR, Domier LL, Tolin SA, et al (2018) Soybean mosaic virus: a successful potyvirus with a wide distribution but restricted natural host range. Mol Plant Pathol 19:1563–1579. https://doi.org/10.1111/mpp.12644

Hill JH, Whitham SA (2014) Control of Virus Diseases in Soybeans. Elsevier Inc.

Kutnjak D, Tamisier L, Adams I, et al (2021) A Primer on the Analysis of High-Throughput Sequencing Data for Detection of Plant Viruses. Microorganisms 9:841. https://doi.org/10.3390/microorganisms9040841

Maree HJ, Fox A, Al Rwahnih M, et al (2018) Application of hts for routine plant virus diagnostics: state of the art and challenges. Front Plant Sci 9:. https://doi.org/10.3389/fpls.2018.01082

Oliver R (2022) Globalizing plant health. Plant Pathol 71:226–235. https://doi.org/10.1111/ppa.13480

Proano MF, Espindola A, Garzon CD (2018) Detection of multiple oomycetes in metagenomic data by Using E-probe Detection of Nucleic Analysis (EDNA). In: International Congress of Plant Pathology (ICPP) 2018: Plant Health in A Global Economy. APSNET

Rahman SU, McCoy E, Raza G, et al (2022) Improvement of Soybean; A Way Forward Transition from Genetic Engineering to New Plant Breeding Technologies. Mol Biotechnol. https://doi.org/10.1007/s12033-022-00456-6

Ribeiro-Junior MR, Barreto da Silva F, Marubayashi JM, et al (2022) Molecular and biological characterization of an isolate of the potyvirus passiflora virus Y naturally infecting soybean (*Glycine max*) in Brazil. Arch Virol. https://doi.org/10.1007/s00705-022-05605-5

Richter DC, Ott F, Auch AF, et al (2008) MetaSim - A sequencing simulator for genomics and metagenomics. PLoS ONE 3:. https://doi.org/10.1371/JOURNAL.PONE.0003373

Stobbe AH, Daniels J, Espindola AS, et al (2013) E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. J Microbiol Methods 94:356–366. https://doi.org/10.1016/j.mimet.2013.07.002

USDA (2022) World Agricultural Production. Circ. Ser.

Wood DE, Salzberg SL (2014) Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:1–12. https://doi.org/10.1186/GB-2014-15-3-R46/FIGURES/5

Zuniga LP, Daniels J, Espindola A, Corona FO (2018) EDNA-Water, using deep sequencing and bioinformatics approach for water-borne plant virus detection. In: International Congress of Plant Pathology (ICPP) 2018: Plant Health in A Global Economy. APSNET

Author Contributions

MRRJ, FMOC, AE and RKS initiated the project; MRRJ, FMOC, AE, DMN, and RKS analyzed the data; all authors contributed to writing the paper. All authors have read and agreed to the published version of the manuscript.

Funding

MRRJ was a recipient of CNPq and CAPES-PrInt Ph.D. fellowships (grant numbers: 140130/2019-5 and 88887.570292/2020- 00); FMOC hold a Sarkeys Foundations Professorship III; RKS hold CNPq research fellowship (grant number: 303411/2018-0). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Data availability

The nucleotide sequences reported the current study are available from the corresponding author on reasonable request.

Corresponding author designation

Marcos Roberto Ribeiro-Junior

Declarations

Ethics approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate: The authors declare their consent to publish the data presented in this research.

Conflict of interest: The authors have no relevant financial or non-financial interests to disclose.

CHAPTER 3

MOLECULAR IDENTIFICATION AND PHYLOGENETIC ANALYSES OF THRIPS SPECIES COLLECTED IN BRAZIL*

Marcos Roberto Ribeiro-Junior[1,2,3], Felipe Barreto da Silva[1], Julio Massaharu Marubayashi[1], Angelica Maria Nogueira[1], Juliana Uzan[1], Cristiane Muller[4], Élison Fabrício Bezerra Lima[5], Daniele Maria Nascimento[2,3], Francisco Manuel Ochoa-Corona[2,3], Renate Krause-Sakate[1].

[1]Department of Plant Protection, São Paulo State University, School of Agriculture, Botucatu, SP, Brazil.
[2]Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, USA.
[3]Institute for Biosecurity and Microbial Forensics, Oklahoma State University, Stillwater, OK, USA.
[4]Corteva Agriscience, Mogi Mirim, SP, Brazil.
[5]Universidade Federal do Piauí, Floriano, Piauí, Brazil.
Corresponding author (e-mail: maribei@okstate.edu) https://orcid.org/0000-0002-5024-1801

Abstract

Thrips are among the most important crop pests worldwide. They are polyphagous key pests of various ornamentals, vegetables, and fruit crops. Thrips are difficult to identify by visual inspection, and molecular methods represent an alternative for situations where a quick and precise identification is necessary. Nine thrips species, collected from 20 different plant species in Brazil, were identified by sequencing partial COI gene. Bayesian phylogenetic analyses demonstrated that

* Chapter written according to the journal **Phytoparasitica**

specimens of the same species were clustered under their respective clade, but not necessarily under the same genus clade. Pairwise analyses indicated that *F. schultzei* is the species that vary the most among its individuals, but no morphological trait could be correlated to this molecular variation. Specific primers based on COI and 28S sequences were obtained for *C. phaseoli* and *F. occidentalis* species, respectively. Our results demonstrate that COX1 PCR followed by sequencing can be used for the identification of described species, but traditional taxonomy must be combined with molecular techniques for the identification of new species since the sequence database available on GenBank is still lacking information for most of the thrips species. Here we provided the first COI sequence for *C. phaseoli* and *F. brevicaulis*.

Keywords: mtCOI, thysanoptera, DNA sequencing

3.1 INTRODUCTION

Thrips are among the main agricultural pests worldwide. Almost all of the thrips species that are considered pests are members of a single subfamily of *Thripidae*, the *Thripinae*, a group that represents around 30% of the species in the insect order Thysanoptera. At least 49 Thripidae species are associated with some damage to cultivated plants in Brazil. These insects are highly polyphagous and key pests of various ornamentals, vegetables, and fruitful crops (Mound and Kibby 1998).

In addition to direct feeding damage, thrips can be orthotospovirus vectors, transmitting them in a circulative-propagative manner (German et al. 1992; Riley et al. 2011). The family *Tospoviridae* (order *Bunyavirales,* class *Ellioviricetes,*

realm *Riboviria*, subphylum *Polyploviricotina*, phylum *Negarnaviricota,* and kingdom *Orthornavirais*) is constituted of tri-segmented negative-sense single-stranded RNA viruses that infect plants and are also able to replicate in their insect vectors in a persistent manner. The family is composed of a single genus, *Orthotospovirus*, whose type species is *Tomato spotted wilt virus* (Butković et al. 2021). A total of seven *Orthotospovirus* species are reported in Brazil, of which for six, the thrips vector is known (Kitajima 2020; Mound et al. 2022; Lima et al. 2022). Seven thrips species with the capability of vectoring this group of viruses are recorded in Brazil, of which five species have confirmed transmission in the national territory (Lima et al. 2022).

Problems with orthotospovirus are increasing worldwide (Pappu et al. 2009; Oliver and Whitfield 2016; Zhang et al. 2021). Recently it was observed that the host range of these orthotospoviruses has expanded from vegetables, tobacco, groundnuts, and ornamental plants to different species of weeds and even woody plants (Qi et al. 2013; Meng et al. 2015; Wang et al. 2016). This is probably related to high densities of thrips observed in different crops.

Given the increasing importance of thrips on various crops, rapid means of accurate identification are crucial for implementing control strategies. Conversely, due to their small size, thrips are difficult to identify by a visual inspection (Mehle and Trdan 2012), and thrips species identification is traditionally based on external morphology (Moritz 1994; Mound and Kibby 1998; Arnett Jr 2000) that requires time-consuming slide mounting and examination of specimens under a microscope, as well as an expert that demands years of practice (Mehle and Trdan 2012).

Molecular methods represent an alternative for situations where a quick and precise identification is necessary. Furthermore, molecular methods allow reliable identification at all growth stages, including immature thrips (Wang et al. 2018), whose traditional taxonomy is still poorly studied and includes only a small fraction of the known thrips richness (Vierbergen et al. 2010; Lima et al. 2022).

The database of thrips nucleotide sequences is lagged since just a few sequences are currently available. Molecular and morphological identification can be used together to identify species, and experts in both are needed to improve the current diagnosis. Thrips genomic sequences could be used to improve the species-level identification if available in public-access databases and blast mechanisms.

Given the lack of information about the molecular identification of thrips species occurring in Brazil, this study aimed to develop a molecular tool to identify the main species of thrips collected from crops, where thrips often constitute a phytosanitary problem in Brazil, and to contribute to generating sequence data of some species not studied yet at the molecular level.

3.2 MATERIAL AND METHODS

3.2.1 Thrips sampling

Thrips were collected from soybean, cotton, tomato, common bean, eggplant, sweet pepper, and others plant species. Insects were collected using a hand-held aspirator and immediately placed in 2 mL tubes containing absolute ethanol. Samples were stored at -20°C until processing. Samples were also sent from farmers and researchers in Brazil. However, in some cases, it was not

possible to certify the integrity of the specimens due to different procedures of collection and storage.

3.2.2 Morphological identification of thrips species

Individual insects were identified using morphological keys (Cavalleri and Mound 2012; Lima et al. 2018; Mound et al. 2022), prior to molecular analysis and sequencing. The species identified during the collection were: *Caliothrips phaseoli, F. schultzei, F. occidentalis, F. williamsi, F. brevicaulis, F. tritici, F. insularis, Arorathtips mexicanus,* and *Thrips tabaci.*

3.2.3 Molecular identification

DNA from individual insects was extracted according to the protocol of Chelex (Leão et al. 2017) or using the PureLink Viral RNA/DNA Mini Kit (Invitrogen).

Chelex method: One individual thrips per sample was crushed and homogenized in a 50 uL of Chelex 5% solution in a 1.5 mL Eppendorf tube and then immediately incubated at 94 °C for 5 min. After centrifugation at 14.000 rpm for 5 min, the supernatant was collected and stored at -20 C until further PCR reaction (Leão et al. 2017).

Commercial Kit: Modifications to the PureLink Viral RNA/DNA Mini Kit were made to obtain a suitable protocol for extraction. The final procedure was set as follows: one insect per sample was crushed with a plastic pestle in a 1.5 ml tube containing 50 µl of Potassium Phosphate Buffer (pH 7,0). Were added in the following order, 25 µl of Proteinase K, 210 µl of Lysis Buffer and 5.88 µl of Carrier RNA. Tube was vortexed and incubated at 56°C for 15 minutes. 250 µl of absolute

ethanol was added to the tube, followed by vortex for 15 seconds and incubating at room temperature for 5 minutes. The sample was transferred to the column and centrifuged at 12,000 rpm for 1 minute. The precipitate was discarded and 500 µl of Wash Buffer (WII) with ethanol was added into the column, followed by centrifugation at 12,000 rpm for 1 minute. The washing steps were repeated 2 times, followed by an extra spin to dry the column. The column was transferred to a new tube and the DNA/RNA was resuspended in 30 µl of sterile, RNase-free water and centrifuged at 12,000 rpm for 1 minute. The DNA/RNA was stored at -80°C until further PCR reaction. The extraction by kit was preferentially used for DNA sequencing.

3.2.4 PCR and Sequencing

Primers mtD-7.2F and mtD-9.2R described by Brunner et al. (2002) were used. Amplification was performed in 25 µl total reaction volumes, containing 3 µl of DNA, 12.5 µl GoTaq® Master Mix (Promega), and 0.25 µl of each primer. The PCR amplification parameters were set for initial denaturation at 95°C for 5 min, and 40 cycles of 95°C for 30 sec, 55°C for 30 sec, and 72°C for 45 sec, followed by a final extension at 72°C for 10 min. After amplification, 6 µl of the PCR products were subjected to electrophoresis on a 1 % agarose gel stained with Ethidium bromide.

The obtained PCR amplicons were purified using PureLink PCR Purification Kit (Invitrogen) and directly sequenced using the primer mtD-7.2F. Sequence chromatograms were manually checked and edited using Geneious software 11.5.1. DNA sequences obtained were assigned according to the previously traditional taxonomically identified species. Sequences were also compared with

the corresponding sequences of other thrips species previously deposited in GenBank, when available.

3.2.5 Obtaining specific primers for *C. phaseoli, F. occidentalis and F. shultzei*

Since *Caliothrips phaseoli* and *Frankliniella occidentalis* thrips species were very frequently found on our samplings, specific primers were designed by aligning the COI sequences of thrips obtained from the different samples previously sequenced. Geneious software was used to obtain a consensus sequence for each species, and the Primer 3 software within Geneoius was used to design the primers. Sequences of partial 28S ribosomal RNA retrieved from public databases were also considered for primer designing.

Primer validation was performed by setting ideal primer TM through a temperature gradient ranging from 53°C to 60°C. Amplification was performed in 25 µl total reaction volumes, containing 3 µl of DNA, 12.5 µl GoTaq® Master Mix (Promega), and 0.25 µl of each primer. Water was used as a negative control. To test primer specificity, each sample consisted of the specie of interest, plus all other species found in this work, non-predicted to be detected.

Published specifics primers described to *F. schultzei*, AG87F/AG88R, and FS_for/FS_rev described by Leão et al. (2017) and Jangra et al. (2020) were also included in our tests.

3.2.6 Bayesian phylogenetic analysis

To better visualize the genetic diversity between the thrips species collected, a Bayesian phylogenetic tree was constructed using the sequences

obtained using the primers mtD-7.2F and mtD-9.2R. The sequences were aligned using Muscle v.3.8.425 within the Geneious v.11.1.5 software, and phylogenetic analysis was performed using Mr. Bayes 3.2.2 (Ronquist and Huelsenbeck 2003). The program MrModeltest v.2.2 (Nylander 2004) was used to select the nucleotide substitution model using the Akaike information criterion (AIC). Two independent runs were conducted simultaneously using 10 million generations and excluding 25% from the resulting trees as burn-in. Phylogenetic trees were visualized using FigTree v1.4.4. Pairwise comparisons were performed with SDT v1.2 (Muhire et al. 2014) using the same dataset used before in the phylogenetic tree.

3.3 RESULTS

3.3.1 Molecular identification

For all samples collected from 20 different plant species, a total of 72 individual thrips were sequenced (Table 1; Figure 1). The sequences were deposited in Genbank, and the accession numbers are indicated in Table 1. A total of nine thrips species were found. The sequence obtained from *F. brevicaulis* was the first to be deposited for this species in the NCBI database.

Table 1 – Thrips samples collected and sequenced.

Sample ID	Collection Date	City/State	Country	Plant	DNA extraction method	Sequencing results	GenBank accession No.
C1	Feb 2019	Estiva Gerbi, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556424
C3	Feb 2019	Mogi Guaçu, SP	Brazil	*Solanum melongena* (Eggplant)	Chelex	*Frankliniella schultzei*	ON556425
C4	Feb 2019	Mogi Guaçu, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556426
C5	Feb 2019	Mogi Guaçu, SP	Brazil	*Solanum aethiopicum* (scarlet eggplant)	Chelex	*Frankliniella schultzei*	ON556427
C6	Feb 2019	Aguaí, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556428
C7	Feb 2019	São João da Boa Vista, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556429
C8	Feb 2019	Vargem Grande do Sul, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556430
C9	Feb 2019	Vargem Grande do Sul, SP	Brazil	*Momordica charantia* (Bitter melon)	Chelex	*Frankliniella schultzei*	ON556431
C13	Feb 2019	Arujá, SP	Brazil	*Begonia maculata* (begonia)	Chelex	*Frankliniella schultzei*	ON556432
C14	Feb 2019	Presidente Prudente, SP	Brazil	*Punica granatum* (pomegranate)	Chelex	*Frankliniella schultzei*	ON556433
C15	Feb 2019	Presidente Prudente, SP	Brazil	*Cucumis melo* (melon)	Chelex	*Frankliniella schultzei*	ON556434

C17	Feb 2019	Álvarez Machado, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556435
C18	Feb 2019	Anhumas, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556436
C19	Feb 2019	Campinas, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556437
C20	Feb 2019	Campinas, SP	Brazil	*Phaseolus vulgaris* (common bean)	Chelex	*Frankliniella occidentalis*	ON556438
C21	Feb 2019	Campinas, SP	Brazil	*Glycine max* (soybean)	Chelex	*Frankliniella schultzei*	ON556439
C22	Dec 2019	São Miguel do Arcanjo, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556440
C23	Dec 2019	Ribeirão Branco, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556441
C24	Dec 2019	Campinas, SP	Brazil	*Phaseolus vulgaris* (common bean)	Chelex	*Caliothrips phaseoli*	ON556442
C25	Jan 2020	Mogi Mirim, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556443
C26	Jan 2020	Mogi Mirim, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556444
C27	Jan 2020	Mogi Mirim, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556445
C28	Jan 2020	Mogi Guaçu, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556446
C28.1	Jan 2020	Mogi Guaçu, SP	Brazil	*Glycine max* (soybean)	Chelex	*Frankliniella schultzei*	ON556447
C29	Jan 2020	Aguaí, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556448

C30	Jan 2020	São Joao da Boa Vista, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556449
C31	Jan 2020	Paranapanema, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556450
C31.1	Jan 2020	Paranapanema, SP	Brazil	*Glycine max* (soybean)	Chelex	*Frankliniella schultzei*	ON556451
C31.2	Jan 2020	Paranapanema, SP	Brazil	*Glycine max* (soybean)	Chelex	*Frankliniella schultzei*	ON556452
C32	Jan 2020	Itaí, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556453
C33	Jan 2020	Pirajú, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556454
C34	Jan 2020	Ipaussú, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556455
C34.1	Jan 2020	Ipaussú, SP	Brazil	*Glycine max* (soybean)	Chelex	*Frankliniella schultzei*	ON556456
C34.2	Jan 2020	Ipaussú, SP	Brazil	*Glycine max* (soybean)	Chelex	*Frankliniella schultzei*	ON556457
C35	Jan 2020	Canitar, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556458
C36	Jan 2020	Assis, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556459
C37	Jan 2020	Assis, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON614711
C38	Jan 2020	Candido Mota, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556461
C39	Jan 2020	Palmital, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556462

C40	Jan 2020	Salto Grande, SP	Brazil	*Glycine max* (soybean)	Chelex	*Caliothrips phaseoli*	ON556463
R3	Jun 2019	Iporanga, SP	Brazil	*Allium cepa* (onion)	Kit	*Thrips tabaci*	ON614712
R8	Nov 2018	Botucatu, SP	Brazil	*Musa acuminata* (banana)	Kit	*Frankliniella brevicaulis*	ON614713
R10	Jun 2019	Uberlandia, MG	Brazil	*Phaseolus vulgaris* (common bean)	Kit	*Frankliniella schultzei*	ON556464
R11	Jun 2019	Uberlandia, MG	Brazil	*Citrullus lanatus* (watermelon)	Kit	*Frankliniella schultzei*	ON556465
R13	Jun 2019	Mogi Mirim, SP	Brazil	*Glycine max* (soybean)	Kit	*Frankliniella schultzei*	ON556466
R14	Jun 2019	Andradas, MG	Brazil	*Capsicum annuum* (sweet pepper)	Kit	*Frankliniella occidentalis*	ON556467
R15	Jun 2019	Andradas, MG	Brazil	*Capsicum annuum* (sweet pepper)	Kit	*Frankliniella occidentalis*	ON556468
R17	Nov 2019	Goixim, PR	Brazil	*Glycine max* (soybean)	Kit	*Frankliniella schultzei*	ON556469
R18	Nov 2019	Candói, PR	Brazil	*Glycine max* (soybean)	Kit	*Frankliniella schultzei*	ON556470
R21	Nov 2019	Pinhão, PR	Brazil	*Glycine max* (soybean)	Kit	*Frankliniella occidentalis*	ON556471
R22	Nov 2019	Mangueirinha, PR	Brazil	*Glycine max* (soybean)	Kit	*Frankliniella occidentalis*	ON556472
R23	Nov 2019	Guarapuava, PR	Brazil	*Glycine max* (soybean)	Kit	*Frankliniella schultzei*	ON556473
R24	Nov 2020	Guarapuava, PR	Brazil	*Zea mays* (corn)	Kit	*Frankliniella williamsi*	ON614714

R27	Feb 2019	Botucatu, SP	Brazil	*Musa acuminata* (banana)	Kit	*Frankliniella brevicaulis*	ON614716
R28	Mar 2020	Correntina, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556474
R29	Mar 2020	Rosário, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556475
R30	Mar 2020	Formosa do Rio Preto, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556476
R31	Mar 2020	Placas, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556477
R32	Mar 2020	Roda Velha, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556478
R33	Mar 2020	Luis Eduardo Magalhães, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556479
R34	Mar 2020	Roda Velha, BA	Brazil	*Gossypium hirsutum* (cotton)	Chelex	*Frankliniella schultzei*	ON556480
R36	May 2022	Botucatu, SP	Brazil	*Musa acuminata* (banana)	Kit	*Frankliniella brevicaulis*	ON638979
R37	May 2022	Bahia State	Brazil	*Mangifera indica* (mango)	Kit	*Frankliniella tritici*	ON638980
R38	May 2022	Botucatu, SP	Brazil	*Hibiscus esculentus*	Kit	*Frankliniella insularis*	ON638981
R39	May 2022	Botucatu, SP	Brazil	*Zea mays* (corn)	Kit	*Arorathtips mexicanus*	ON638982
R40	Aug 2019	Rio Parnaiba, MG	Brazil	*Bidens pilosa*	Kit	*Caliothrips phaseoli*	ON638983
R41	Jan 2020	Uberaba, MG	Brazil	*Solanum tuberosum* (potato)	Kit	*Caliothrips phaseoli*	ON638984
R42	May 2021	Goiás, GO	Brazil	*Solanum*	Kit	*Caliothrips phaseoli*	ON638985

				lycopersicum (tomato)			
R43	Jun 2018	Rio Verde, GO	Brazil	*Phaseolus vulgaris* (common bean)	Kit	*Caliothrips phaseoli*	ON638986
R44	Apr 2015	Candiota, RS	Brazil	*Eucalyptus* spp.	Kit	*Caliothrips phaseoli*	ON638987
R45	Jan 2020	Luis Eduardo Magalhães, BA	Brazil	*Glycine max* (soybean)	Kit	*Caliothrips phaseoli*	ON638988
R46	Aug 2019	Timbiras, MA	Brazil	*Glycine max* (soybean)	Kit	*Caliothrips phaseoli*	ON638989

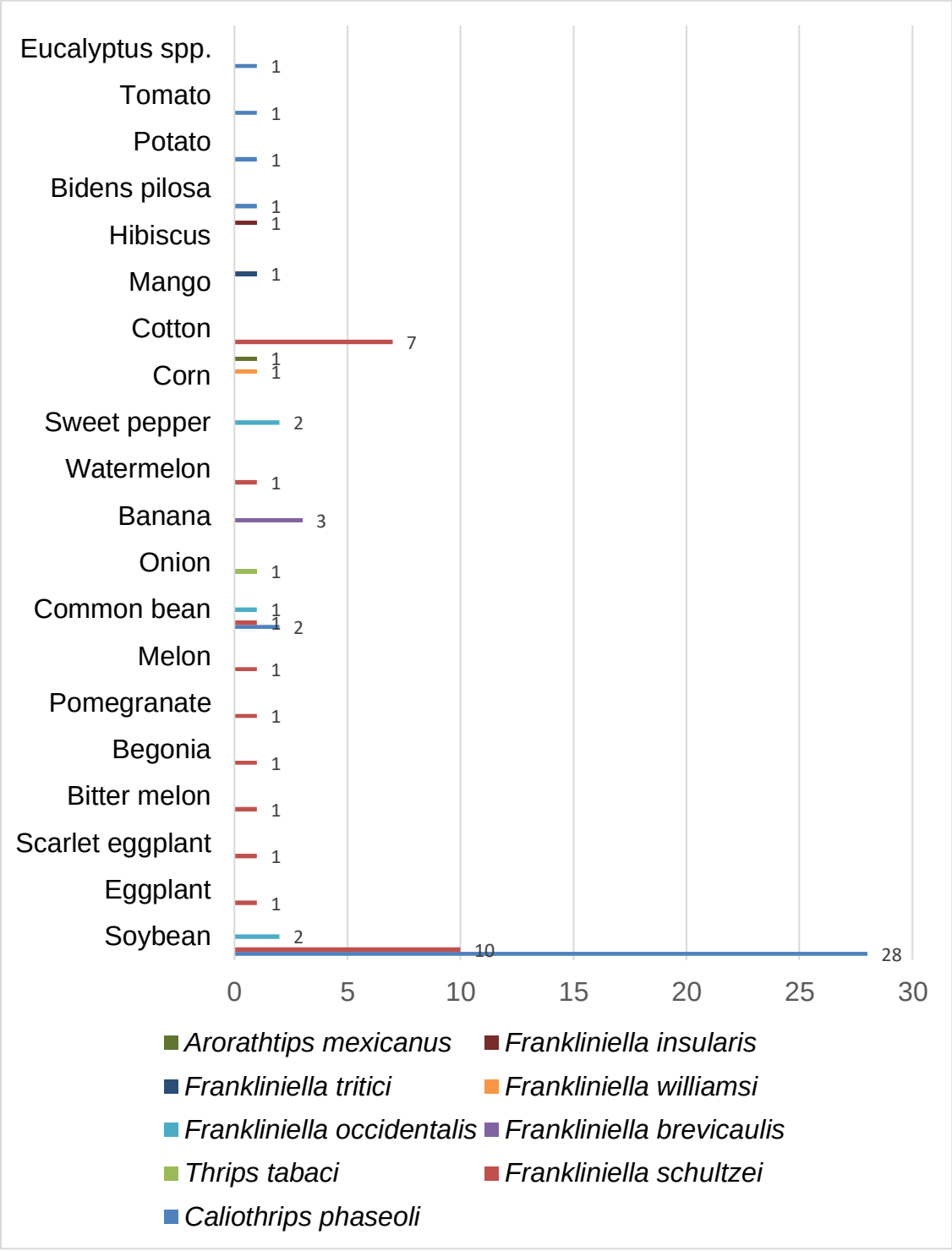


Figure 1 - Number of thrips species collected in different plants and individually sequenced.

C. phaseoli was the species with the largest number of samples analyzed, representing 47.2% of the individuals sequenced. *F. schultzei* was the second most common species representing 34.7%, followed by *F. occidentalis* (6.9%). The other five species represent only 11.1% of the thrips samples sequenced (Figure 1).

3.3.2 Specific primers

Specific primes based on COI sequences were obtained for *C. phaseoli* (Table 2). *In silico* analysis also reveals that the primers obtained in this work are highly specific and do not match with *C. fasciatus*, an exotic pest in Brazil, but a primary pest in U.S. (Bailey 1933). Primers for *F. occidentalis* were successfully obtained from 28S sequences available in the NCBI database (Table 2). The optimal melting temperature for the primers Cp COI 156F/ Cp COI 388R and Fo 28S 145F/Fo 28S 491R was set at 55°C and fragments of 233 bp and 349 bp were obtained and sequenced for *C. phaseoli* and *F. occidentalis*, respectively. Negative controls using DNA of all other species found in this study and nuclease-free water confirmed the specificity of the primers.

Specific primers were not obtained for *F. schultzei* yet and testing is still in progress. We faced issues with non-specificity of primers for this species, where, in some cases, primers that amplify *F. schultzei* also cause amplification on other species. The specific primer AG87F-AG88R (Jangra et al. 2020) showed amplification for *F. schultzei* and *C. phaseoli* while FS_for/FS_rev (Leão et al. 2017) showed amplification for *F. schultzei* and *F. occidentalis.*

Table 2 - Primers designed for *Caliothrips phaseoli, Frankliniella schultzei,* and *Frankliniella occidentalis*.

Thrips specie	Primer name	Nucleotide Sequence (5' - 3')	Product size	TM	Gene	Results
Caliothrips phaseoli	Cp COI 156F	ACCTGGCTGGGATCTCTTCT	233 bp	55°C	COI	Positive specific results
	Cp COI 388R	TGGGTCTCCACCACCTCTAG				
Caliothrips phaseoli	Cp COI 156F	ACCTGGCTGGGATCTCTTCT	163 bp	55°C	COI	No amplification
	Cp COI 388R	CCACCTGCTAGAACGGGAAG				
Frankliniella schultzei	Fs COI 13 F	AGGTTTAAGAAAAGAGGGAGCAGGA	355 bp	55°C	COI	Nonspecific amplification
	Fs COI 368 R	AGGTTTAAGAAAAGAGGGAGCAGGA				
Frankliniella schultzei	Fs COI 23 F	AAAGAGGGAGCAGGAACAGG	332 bp	55°C	COI	Nonspecific amplification
	Fs COI 355 R	AGAATTGGATCACCTCCTCCTC				
Frankliniella schultzei	Fs 28S 451 F	TGGGCTCTGCGAAACCTAAA	295 bp	55°C	28S	No amplification
	Fs 28S 746 R	GGAACCAGCTACTAGATGGTTC				
Frankliniella schultzei	Fs 28S 690 F	GATCGTCGGAGCTGGGTATAG	504 bp	55°C	28S	No amplification
	Fs 28S 955 R	CGCCAGTTCTGCTTACCAAAAT				

Frankliniella schultzei	Fs 28S 451 F	TGGGCTCTGCGAAACCTAAA	277 bp	55°C	28S	Nonspecific amplification
	Fs 28S 728 R	TCGATTAGTCTTTCGCCCCTATAC				
Frankiliniela occidentalis	Fo 28S 145F	CACCGCTTCCCCCGTAAAG	349 bp	55°C	28S	Positive specific results
	Fo 28S 491R	CGCTGCTACTAGGGGAATCC				

3.3.3 Bayesian phylogenetic analysis

The phylogenetic trees generated using Bayesian Inference resulted in two major clusters, the first one including individuals from *F. schultzei*, *F. brevicaulis, F. wiliamsi, F. tritici, F. occidentalis, A. mexicanus,* and *T. tabaci* species, and the second cluster composed of insects of the *C. phaseoli* and *F. brevicaulis* species (Figure 1). Tree clustering revealed that species under the genus *Frankliniella* were not clustered under the respective genus clade but under the species clade. Specimens of *F. schultzei* could be differentiated in different subclades that could not be correlated with a different morphological trait. The specimens of *C. phaseoli* were genetically highly homogeneous and formed a unique clade.

According to SDT analysis (Figure 2), the identity between individuals of the same species ranged between 97.5% and 100%. *F. schultzei* presented more variation between individuals (>97.5%), followed by *C. phaseoli* (>99.3%), *F. brevicaulis* (>99.8%), and *F. occidentalis*.

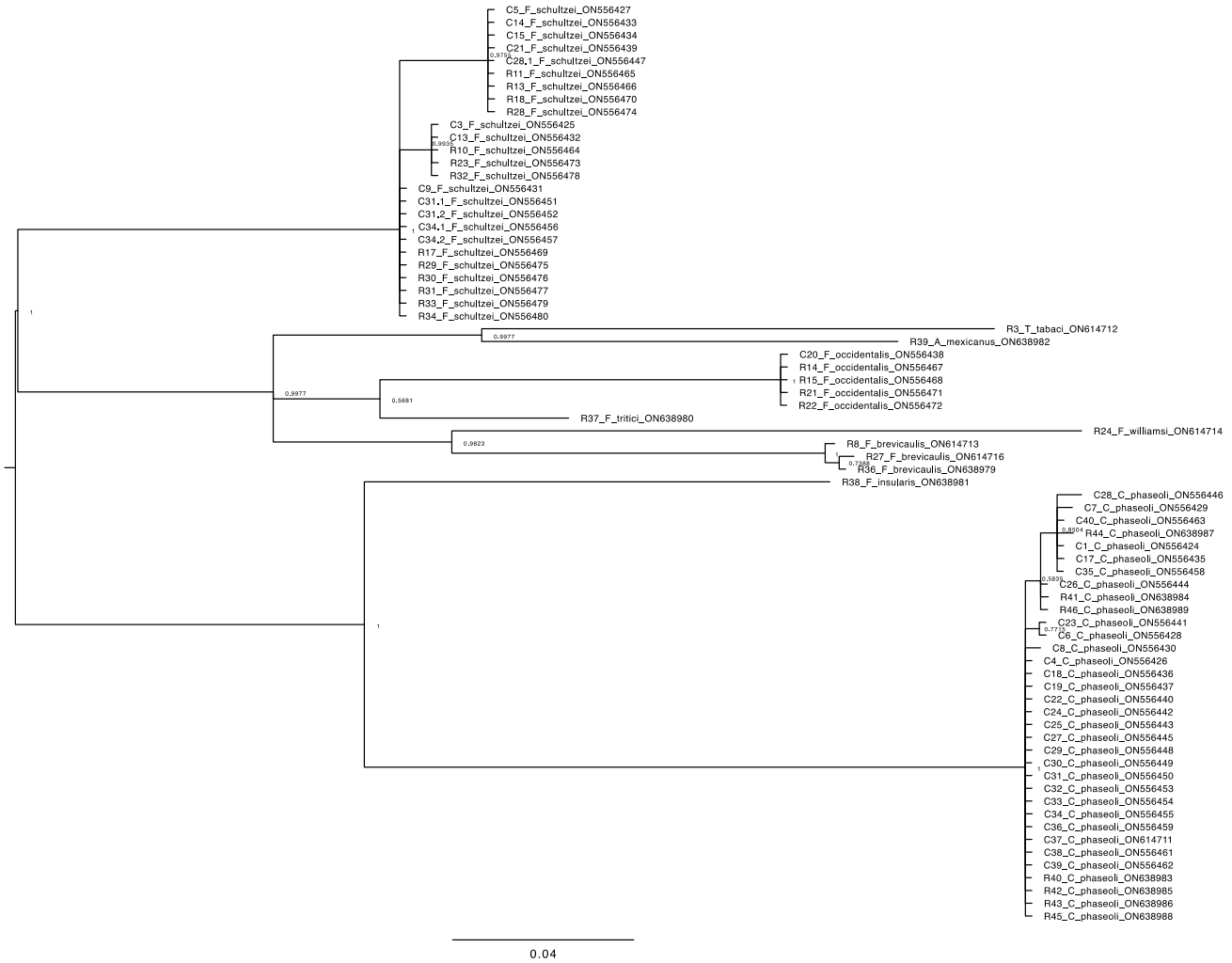


Figure 2 - Bayesian phylogenetic tree of partial COI sequences of thrips species collected.

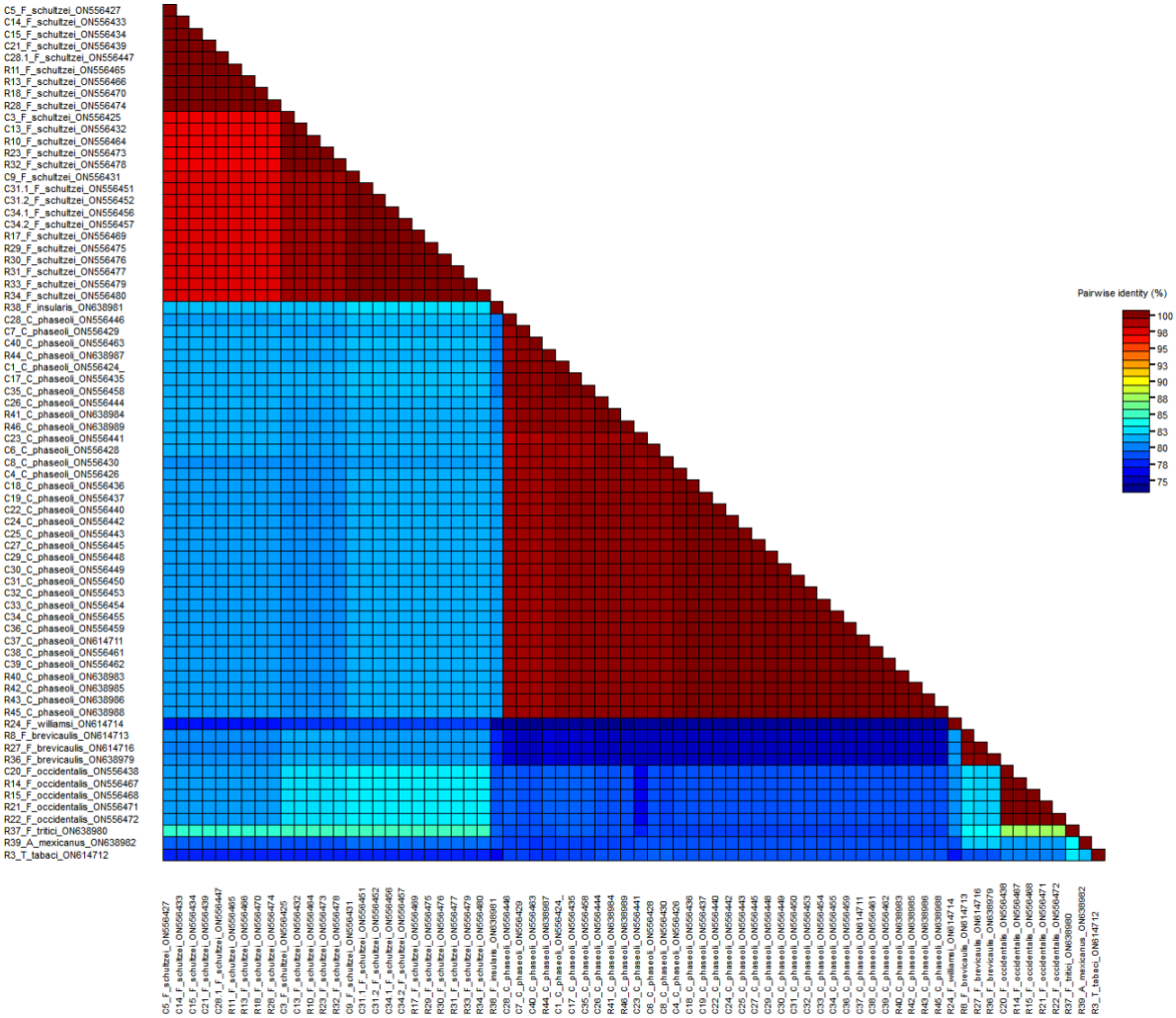


Figure 3 - Percentage of identity of nucleotide sequence of partial COI gene of thrips species collected on different crops in Brazil.

3.4 DISCUSSION

Thrips are an emergent pest in some crops and species identification is crucial to the correct management of the insect. The traditional taxonomy is laborious, time-consuming, and requires expert knowledge, so here we confirmed that the mtCOI DNA amplification of thrips specimens, followed by sequencing and blast analyses can contribute to the correct identification of the insects. Some thrips species, such as *C. phaseoli* and *F. brevicaulis,* could not be identified directly by BLAST searches, because sequences were not available. Now, our generated database could help further researchers to easily discriminate between these species by BLAST searches.

Mitochondrial DNA cytochrome oxidase (COI) (Kox et al. 2005; Zhang et al. 2012) and the 28S ribosomal RNA (Bravo-Pérez et al. 2018) are the two most used DNA markers for thrips identification. COI proved useful in identifying *T. tabaci* populations in onion crops and plants from *Allium* genus (Li et al. 2020; Khan et al. 2022). We were able to identify nine thrips species collected from different hosts in Brazil: *C. phaseoli, F. schultzei, F. occidentalis, F. williamsi, F. brevicaulis, F. tritici, F. insularis, A. mexicanus,* and *T. tabaci*. For two of them, *C. phaseoli* and *F. occidentalis*, primer pairs derived from COI and 28S, respectively, were considered suitable to discriminate these species. These species are very frequently found infesting crops in Brazil and PCR will be used for a rapid identification of these insects. Total DNA can also be easily and rapidly extracted with the Chelex protocol here described, providing good quality for PCR and further analysis.

Designing a new specific primer for *F. schultzei* is still a dilemma since problems with nonspecificity and dimerization were observed in our attempts. The primers obtained here presented problems to discriminate *F. schultzei* and *F. occidentalis*. Some primers previously described as specific for *F. schultzei* (Leão et al. 2017;

Jangra et al. 2020), presented the same specificity problem observed for our primers, probable, during the validation the authors did not have such a wide variety of thrips species to analyze the cross reactivity.

The phylogenetic tree of the COI partial region discriminates species into different clades and SDT analysis indicated that *F. schultzei* is the species that vary the most between its individuals. This variability could not be correlated with biological properties like color, host collection, or any other morphological parameter. In contrast, insects belonging to *C. phaseoli* species, were highly homogenous in COI sequence, even when collected from different hosts.

Here we conclude that thrips species identification can be a big challenge and classic morphological analysis is still essential mainly for new or poorly understood thrips species. Therefore, combining taxonomic identification with molecular techniques can contribute to the enrichment of thrips databases that will permit a more accessible and easy identification even for non-experts in thrips.

3.5 CONCLUSION

Thrips sampling in Brazil revealed a total of 11 different species found. PCR using generic primers followed by direct sequencing is suitable for identifying this species using our new sequences in BLAST mechanisms or by Bayesian analyses. Bayesian phylogenetic analyses demonstrated that species in the genus *Frankliniella* were not clustered under the respective genus clade. SDT analysis indicated that *F. schultzei* is the specie that varies the most among its individuals. Specific primers were successfully developed for the precise identification of *C. phaseoli* and *F. occidentalis.* Specific primers for *F. schultzei* are still in development. Traditional taxonomy is still

necessary for the identification of new species and must be combined with molecular techniques.

REFERENCES

Arnett Jr RH (2000) American insects: a handbook of the insects of America north of Mexico. Crc Press

Bailey SF (1933) The biology of the bean thrips. University of California Printing Office

Bravo-Pérez D, Santillán-Galicia MT, Johansen-Naime RM, et al (2018) Species diversity of thrips (Thysanoptera) in selected avocado orchards from Mexico based on morphology and molecular data. J Integr Agric 17:2509–2517. https://doi.org/10.1016/S2095-3119(18)62044-1

Brunner PC, Fleming C, Frey JE (2002) A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach. Agric For Entomol 4:127–136. https://doi.org/10.1046/J.1461-9563.2002.00132.X

Butković A, González R, Elena SF (2021) Revisiting Orthotospovirus phylogeny using full-genome data and testing the contribution of selection, recombination and segment reassortment in the origin of members of new species. Arch Virol 166:491–499. https://doi.org/10.1007/s00705-020-04902-1

German TL, Ullman DE, Moyer JW (1992) Tospoviruses: diagnosis, molecular biology, phylogeny, and vector relationships. Annu Rev Phytopathol 30:315–348

Jangra S, Mittal A, Dhall H, et al (2020) A multiplex PCR assay for rapid identification of major tospovirus vectors reported in India. BMC Genomics 21:. https://doi.org/10.1186/s12864-020-6560-x

Khan F, Roy MC, Kim Y (2022) Thelytokous Reproduction of Onion Thrips, Thrips tabaci Lindeman 1889, Infesting Welsh Onion and Genetic Variation among Their Subpopulations. Insects 13:. https://doi.org/10.3390/insects13010078

Kitajima EW (2020) An annotated list of plant viruses and viroids described in Brazil (1926-2018). Biota Neotropica 20:. https://doi.org/10.1590/1676-0611-bn-2019-0932

Kox LFF, van den Beld HE, Zijlstra C, Vierbergen G (2005) Real-time PCR assay for the identification of Thrips palmi. EPPO Bull 35:141–148. https://doi.org/10.1111/j.1365-2338.2005.00797.x

Leão EU, de Almeida Spadotti DM, Rocha KCG, et al (2017) Efficient detection of *Frankliniella schultzei* (Thysanoptera, Thripidae) by cytochrome oxidase I gene (mtCOI) direct sequencing and real-time PCR. Braz Arch Biol Technol 60:. https://doi.org/10.1590/1678-4324-2017160425

Li X, Zhang Z, Zhang J, et al (2020) Population Genetic Diversity and Structure of Thrips tabaci (Thysanoptera: Thripidae) on Allium Hosts in China, Inferred from Mitochondrial COI Gene Sequences. J Econ Entomol 113:1426–1435. https://doi.org/10.1093/jee/toaa001

Lima ÉFB, Silva L de MR, Fontes L da S, Borbón CM (2022) Identification of second instar larvae of thrips (Thysanoptera: Thripidae) vectors of *Orthotospovirus* (Tospoviridae) in South America. Austral Entomol 61:199–208. https://doi.org/10.1111/aen.12597

Mehle N, Trdan S (2012) Traditional and modern methods for the identification of thrips (Thysanoptera) species. J Pest Sci 85:179–190. https://doi.org/10.1007/s10340-012-0423-4

Meng J, Liu P, Zhu L, et al (2015) Complete Genome Sequence of Mulberry Vein Banding Associated Virus, a New Tospovirus Infecting Mulberry. PLOS ONE 10:e0136196. https://doi.org/10.1371/journal.pone.0136196

Moritz G (1994) Pictorial key to the economically important species of Thysanoptera in Central Europe. EPPO Bull 24:181–208

Mound LA, Kibby G (1998) Thysanoptera: an identification guide, 2nd edn. CABI

Mound LA, Wang Z, Lima ÉFB, Marullo R (2022) Problems with the Concept of “Pest” among the Diversity of Pestiferous Thrips. Insects 13:61. https://doi.org/10.3390/insects13010061

Muhire BM, Varsani A, Martin DP (2014) SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. PLoS ONE 9:e108277. https://doi.org/10.1371/journal.pone.0108277

Qi F, Ming D, Jia-Hong D, et al (2013) Preliminary report of Tospovirus infecting Macadamia seedling in Yunnan, China. Acta Hortic Sin 40:350

Riley DG, Joseph SV, Srinivasan R, Diffie S (2011) Thrips vectors of tospoviruses. J Integr Pest Manag 2:I1–I10

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Vierbergen G, Kucharczyk H, Kirk WDJ (2010) A key to the second instar larvae of the Thripidae of the Western Palaearctic region (Thysanoptera). Tijdschr Voor Entomol 153:99–160. https://doi.org/10.1163/22119434-900000294

Wang H, Kennedy GG, Reay-Jones FPF, et al (2018) Molecular Identification of Thrips Species Infesting Cotton in the Southeastern United States. J Econ Entomol 111:892–898. https://doi.org/10.1093/jee/toy036

Wang YX, Hong N, Wang GP, et al (2016) First Report of the Tospovirus Tomato necrotic spot associated virus Infecting Kiwifruit (Actinidia sp.) in China. Plant Dis 100:2539. https://doi.org/10.1094/PDIS-05-16-0629-PDN

Zhang GF, Meng XQ, Min L, et al (2012) Rapid diagnosis of the invasive species, Frankliniella occidentalis (Pergande): a species-specific COI marker. J Appl Entomol 136:410–420. https://doi.org/10.1111/J.1439-0418.2011.01661.X

Author Contributions

MRRJ, FBS, JMM and RKS initiated the project; MRRJ, FBS, JMM, AMN, JU, CM, EFBL, DMN, FMOC and RKS analyzed the data; MRRJ, EFBL, DMN and RKS contributed to writing the paper. All authors have read and agreed to the published version of the manuscript.

Funding

MRRJ was a recipient of CNPq and CAPES-PrInt Ph.D. fellowships (grant numbers: 140130/2019-5 and 88887.570292/2020-00); FBS was a recipient of a FAPESP Ph.D. fellowship (grant number: 2019/24642-8); AMG is a recipient of a FAPESP post-doctoral fellowship (grant number: 2019/18639-4); RKS hold CNPq research fellowship (grant number: 303411/2018-0). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Data availability

The nucleotide sequences reported the current study are available from the corresponding author on reasonable request.

Corresponding author designation

Marcos Roberto Ribeiro-Junior

Declarations

Ethics approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate: The authors declare their consent to publish the data presented in this research.

Conflict of interest: The authors have no relevant financial or non-financial interests to disclose.

FINAL CONSIDERATIONS

Complete genome sequence of passiflora virus Y (PaVY) was obtained for the first time in Brazil. PaVY-Br induced chlorotic spots and systemic mosaic on soybean and chlorotic local lesions on passion fruit (*Passiflora edulis*) and sesame (*Sesamum indicum*) and was successfully transmitted by *Myzus persicae.*

E-probes ranging from 19-60 nucleotides in length were designed for the 46 viruses that infect soybean using MiProbe within MiFi® platform. *In silico* results validate the usage of e-probes to detect up to 46 viruses in metagenomic soybean data. Curated soybean virus e-probes were made available to the public for diagnostics purposes through MiDetect within MiFi® platform.

A total of nine different species of thrips were detected on sampling in Brazil. PCR using generic primers followed by direct sequencing is now suitable for identifying this species using our new sequences in BLAST mechanisms or by Bayesian analyses. Specific primers were successfully developed for precise identification of *C. phaseoli* and *F. occidentalis*. Traditional taxonomy is still necessary for the identification of new species and must be combined with molecular techniques.

REFERENCES

BANDARA AY, WEERASOORIYA DK, BRADLEY CA, ALLEN TW, ESKER PD. Dissecting the economic impact of soybean diseases in the United States over two decades. **PLoS ONE**. 2020;15(4). doi:10.1371/journal.pone.0231141

BOCSANCZY AM, ESPINDOLA A, NORMAN DJ, CARDWELL KF. E-probes development for rapid, sensitive and specific pathogen detection in blueberries. In: **INTERNATIONAL CONGRESS OF PLANT PATHOLOGY (ICPP) 2018**: Plant Health in A Global Economy. APSNET; 2018.

ESPINDOLA A, CARDWELL KF. Microbe Finder (MiFi): Pathogen detection in metagenomic sequence data. **Plant Health 2019**. Published online 2019.

ESPINDOLA A, ROY A, MAVRODIEVA VA, CARDWELL KF. E-probe: A new diagnostic tool for detection of Dichorhaviruses associated with Citrus leprosis syndrome. **Plant Health 2019**. Published online 2019.

ESPINDOLA AS, CARDWELL KF. Microbe finder (Mifi®): Implementation of an interactive pathogen detection tool in metagenomic sequence data. **Plants**. 2021;10(2):1-14. doi:10.3390/plants10020250

ESPINDOLA AS, SCHNEIDER W, CARDWELL KF, et al. Inferring the presence of aflatoxin-producing Aspergillus flavus strains using RNA sequencing and electronic probes as a transcriptomic screening tool. **PLoS ONE**. 2018;13(10). doi:10.1371/JOURNAL.PONE.0198575

ESPINDOLA AS, SCHNEIDER WL, MAREK S, HOYT P, GARZON CD. Metagenome data mining for eukaryotic plant pathogens. In: **PHYTOPATHOLOGY**. Vol 105. AMER PHYTOPATHOLOGICAL; 2015:41.

FONTES MG, ANDREWS DA SILVA GF, LIMA MF, et al. First Report of Groundnut Ringspot Orthotospovirus Infecting Soybeans in Brazil. **Plant Disease**. 2019;103(4):777. doi:10.1094/PDIS-07-18-1246-PDN

HARTMAN GL, WEST ED, HERMAN TK. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. **Food Security**. 2011;3(1):5-17.

HILL JH, WHITHAM SA. Control of Virus Diseases in Soybeans. Vol 90. 1st ed. Elsevier Inc.; 2014. doi:10.1016/B978-0-12-801246-8.00007-X

JONES WJ. High-throughput sequencing and metagenomics. **Estuaries and coasts**. 2010;33(4):944-952.

KITAJIMA EW. An annotated list of plant viruses and viroids described in Brazil (1926-2018). **Biota Neotropica.** 2020;20(2). doi:10.1590/1676-0611-bn-2019-0932

KREUZE JF, PEREZ A, UNTIVEROS M, et al. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. **Virology**. 2009;388(1):1-7.

OLIVER R. Globalizing plant health. **Plant Pathology**. 2022;71(1):226-235. doi:10.1111/ppa.13480

PAPPU HR, JONES RAC, JAIN RK. Global status of tospovirus epidemics in diverse cropping systems: successes achieved and challenges ahead. **Virus Res**. 2009;141(2):219-236. doi:10.1016/J.VIRUSRES.2009.01.009

POSTNIKOVA E, BALDWIN C, WHITEHOUSE CA, et al. Identification of bacterial plant pathogens using multilocus polymerase chain reaction/electrospray ionization-mass spectrometry. **Phytopathology**. 2008;98(11):1156-1164.

PROANO MF, ESPINDOLA A, GARZON CD. Detection of multiple oomycetes in metagenomic data by Using E-probe Detection of Nucleic Analysis (EDNA). In: **International Congress of Plant Pathology (ICPP) 2018**: Plant Health in A Global Economy. APSNET; 2018.

PROINDE. Loading Soya Beans in Brazil-Practical Guidance Loading Soya Beans in Brazil.; 2020. www.proinde.com.br

RIBEIRO-JUNIOR MR, BARRETO DA SILVA F, MARUBAYASHI JM, et al. Molecular and biological characterization of an isolate of the potyvirus passiflora virus Y naturally infecting soybean (Glycine max) in Brazil. **Arch Virol**. (2022) https://doi.org/10.1007/s00705-022-05605-5

SAVARY S, WILLOCQUET L, PETHYBRIDGE SJ, ESKER P, MCROBERTS N, NELSON A. The global burden of pathogens and pests on major food crops. **Nat Ecol Evol**. 2019;3(3):430-439.

STOBBE AH, DANIELS J, ESPINDOLA AS, et al. E-probe Diagnostic Nucleic acid Analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. **Journal of Microbiological Methods**. 2013;94(3):356-366. doi:10.1016/j.mimet.2013.07.002

TYSON GW, CHAPMAN J, HUGENHOLTZ P, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. **Nature**. 2004;428(6978):37-43.

USDA. World Agricultural Production. Circular Series. Published online May 2022.

ZUNIGA LP, DANIELS J, ESPINDOLA A, CORONA FO. EDNA-Water, using deep sequencing and bioinformatics approach for water-borne plant virus detection. In: International Congress of Plant Pathology (ICPP) 2018: **Plant Health in A Global Economy**. APSNET; 2018.

ZUNIGA LP, ESPINDOLA A, MELOUK HA, ALI A, CARDWELL KF, CORONA FO. Detection of cucurbit viruses in Oklahoma combining EDNA with Multiplex RT-PCR

coupled with High Resolution Melting. In: 2017 **APS Annual Meeting**. APSNET; 2017.

APPENDIX A – Published article unrelated to the main thesis subject

Ribeiro-Junior, M.R., Gorayeb, E.S., Nogueira, A.M. et al. Molecular and biological characterization of apium virus Y infecting parsley in Brazil. Trop. plant pathol. 46, 565–570 (2021).

Received: 22 February 2021

Accepted: 31 May 2021

Published: 22 June 2021

Issue Date: October 2021

DOI: https://doi.org/10.1007/s40858-021-00444-2

Molecular and biological characterization of apium virus Y infecting parsley in Brazil

Marcos Roberto Ribeiro-Junior[1], Eduardo Silva Gorayeb[1], Angélica Maria Nogueira[1], Giovana Carolina Dourado Cruciol[1], Daniele Maria Nascimento[1], Marcelo Agenor Pavan[1] & Renate Krause-Sakate[1].

[1]Department of Plant Protection, São Paulo State University, UNESP, School of Agriculture, Botucatu, SP, Brazil.

*Corresponding author (e-mail: marcos.ribeiro@unesp.br) https://orcid.org/0000-0002-5024-1801

Tropical Plant Pathology (2021) 46:565–570

https://doi.org/10.1007/s40858-021-00444-2

Abstract

Parsley is a biannual herbaceous plant, frequently used as a condiment in Brazilian gastronomy, also having medicinal properties. This work reports the molecular identification, partial host range, and complete genome sequencing of an apium virus Y (*Potyvirus*) isolate (ApVY-Br) infecting parsley in Brazil. Complete genome sequencing revealed a 9913nt long virus with high identity with previously reported ApVY isolates. The isolate was capable of infecting parsley, coriander, *Chenopodium amaranticolor*, and *Chenopodium quinoa* plants. Bayesian phylogenetic

analyses based on the complete genome and coat protein sequences clustered the ApVY isolates based on sampling locations.

Keywords: Potyvirus. ApVY. *Petroselinum crispum*

Parsley (*Petroselinum crispum* (Mill.) Nyman ex AW Hill) is a biennial herbaceous plant that belongs to Apiaceae family originally from the central Mediterranean region (southern Italy, Algeria, and Tunisia). This plant stands out for its importance in gastronomy as a condiment and also for its medicinal properties (Pedroso et al. 2010). Besides being widely used in Brazilian gastronomy (where it is known as “salsa”), this crop is still poorly studied, with a lack of updated data related to production and a few numbers of studies regarding diseases detection and management.

Among all apiaceous potyviruses, celery mosaic virus (CeMV) is the oldest reported and most widely spread virus, being also the only one infecting parsley in Brazil (Novaes et al. 2000). Recently, new apiaceous-infecting potyviruses were recognized, such as apium virus Y (ApVY), arracacha mottle virus (AMoV), carrot virus Y (CarVY), carrot thin leaf virus (CTLV), and panax virus Y (PanVY) (Moran et al. 2002; Orílio et al. 2009; Yan et al. 2010; Xu et al. 2014; Kinoti et al. 2020). Apium virus Y was firstly reported in Australia (Moran et al. 2002) infecting sea celery (*Apium prostraum*), poison hemlock (*Conium maculatum*), and parsley (*Petroselinum crispum*). Since then, this virus was found infecting several apiaceous crops, such as cilantro, celery, and parsley, in the USA, New Zealand, and Slovenia (Tang et al. 2007; Tian et al. 2008; Xu et al. 2011; Mehle et al. 2019).

During 2018, parsley plants were found in a private garden in the municipality of Assis, São Paulo, Brazil (22° 39′ 40″ S, 50° 23′ 58″ W), showing symptoms of chlorotic spots and mild leaf crinkle (Fig. 1). Symptomatic leaves were collected and initially submitted to an indirect ELISA test, using a generalist antiserum for members of the genus Potyvirus (Agdia), which resulted in positive result.



Fig 1. Symptoms observed in Parsley (*Petroselinum crispum*) plants collected in Assis, São Paulo state, Brazil

For virus characterization and identification, total RNA was extracted from symptomatic parsley leaves using the Purelink Viral RNA/DNA Kit (Thermo Fisher Scientific) following manufacturer instructions. A single-step RT-PCR reaction was performed using the potyvirus universal primers W-CIEN (5′-ATG GTT TGG TGY ATY GAR AAT-3′) and PV-1 (5′-GAT TTA GGT GAC ACTATA GT(17)-3′) (Gibbs and Mackenzie 1997; Mota et al. 2004), which amplify a fragment of approximately 850 bp corresponding to part of the capsid protein-coding region and part of the 3′ UTR. The reaction conditions were identical to those recommended by Mota et al. (2004). The amplified fragment was sequenced (GenBank Accession No. MT152163) and shared 92.91% identity (616 nt of 663), in BLASTn analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi), with an apium virus Y isolate from coriander (EU515125).

Because of the similarity shared by all apiaceous potyviruses, and to understand more deeply the characteristics of the Brazilian potyvirus isolate, five samples previously confirmed by RT-PCR were pooled and submitted to high-throughput sequencing (HTS) and sequence analysis. The total RNA was submitted to quality control using the 2100 Bioanalyzer (Agilent Technologies) and used for the construction of a cDNA library utilizing the Complete ScriptSeq Kit (Epicenter, Illumina). The transcriptome sequencing was performed with Illumina HiSeq2500 platform at the Center of Functional Genomics (ESALQ/USP, Piracicaba, Brazil) using HiSeq Flow Cell v4 and HiSeq SBS v4 (Illumina) kits in paired-end mode (2×100 bp).

Adapter sequence removal and quality trimming were performed with CLC Genomics Workbench software version 9.0.3 and Geneious software version 11.1.5 was used for genome assembly. A total of 31,336,949 reads were generated, of which 459,742 were mapped to the apium virus Y (HM363516) genome, covering the entire

sequence with a mean coverage depth of 4346.6×. No other potyvirus sequences were identified in HTS sequence data. The final sequence obtained (MT152162) was 9913 nt long, with a 5′ UTR of 114 nt and a 3′ UTR of 243 nt (excluding the poly(A) tail) (Fig. 2).

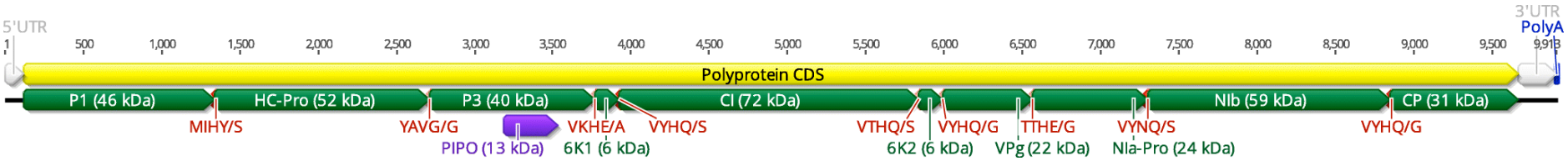


Fig 2. Genome organization of parsley isolate of apium virus Y (ApVY-Br). White boxes indicate the 5′- and 3′- untranslated regions, the yellow box indicate the large open reading frame (ORF), the ten putative mature proteins and sizes are indicated by green boxes and the putative protein PIPO is indicated by a purple box. The location of each putative cleavage site on the polyprotein is red marked. Created using Geneious software version 11.1.5

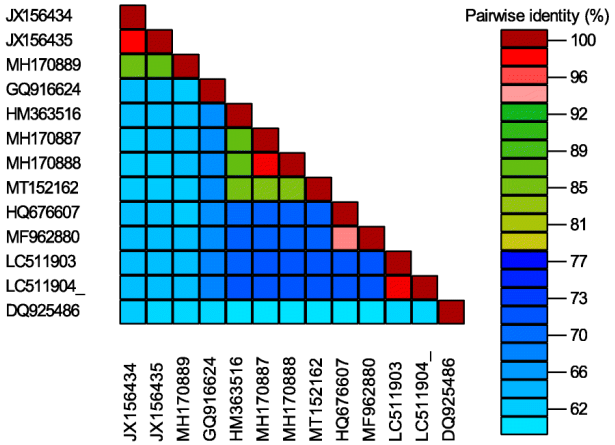
The putative ORF starts at an AUG codon at nucleotide position 115–118 in the sequence and ends with an UAA codon at nt 9667–9669. This gene encodes a polyprotein of 3184 amino acids (aa) with a molecular weight of 362.42 kDa. Nine putative protease cleavage sites were identified in the ApVY-Br polyprotein, giving rise to ten putative mature proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP) (Fig. 2). The putative protein PIPO was identified within the P3 region protein in the +2 reading frame, starting at position 3186 and ends with a UAG termination codon at position 3528–3530.

Pairwise comparisons were performed with SDT v1.2 (Muhire et al. 2014) using the complete coding region of the polyprotein sequence obtained with all ApVY complete polyprotein sequences (MH170887, MH170888, HM363516) and other apiaceous potyviruses, such as CeMV (HQ676607, MF962880), CarVY (LC511903, LC511904), PanVY (GQ916624), CTLV (MH170889, JX156434, JX156435), and AMoV (DQ925486) retrieved from public data bases. Pairwise nucleotide and amino acid identities of the polyprotein ORF, cleaved products, and nucleotide of 5′- end and the 3′-terminus of the ApVY Brazilian isolate with other ApVY sequences were also performed using SDT v1.2.

The Brazilian potyvirus isolate shared 84.3–85.0% nucleotide and 92.5–92.6% amino acid identities with the ApVY isolates HM363516, MH170888, and MH170887 in polyprotein ORF (Fig. 3A). Based on the potyvirus species demarcation criterion established by the International Committee on Taxonomy of Viruses (ICTV) (Wylie et al. 2017), members of the same species share higher than 76% nucleotide and 82% amino acid identities to complete ORF. Therefore, the isolate obtained here should be considered an isolate of ApVY (ApVY-Br). The 5′ UTR and the 3′UTR regions were the most variables and conserved regions, respectively. Regarding the coding regions, P1

was the most variable, whereas VPg, PIPO, and HC-Pro regions were the most conserved (Fig. 3B). Almost all motifs and domains in the polyprotein were identical to those previously described by Xu et al. (2011), except for a substitution of an isoleucine for a valine at amino acid 382 (F-V-V-R-G - aa 381-385), changing the aa sequence in a motif related with autoproteolytic activity.

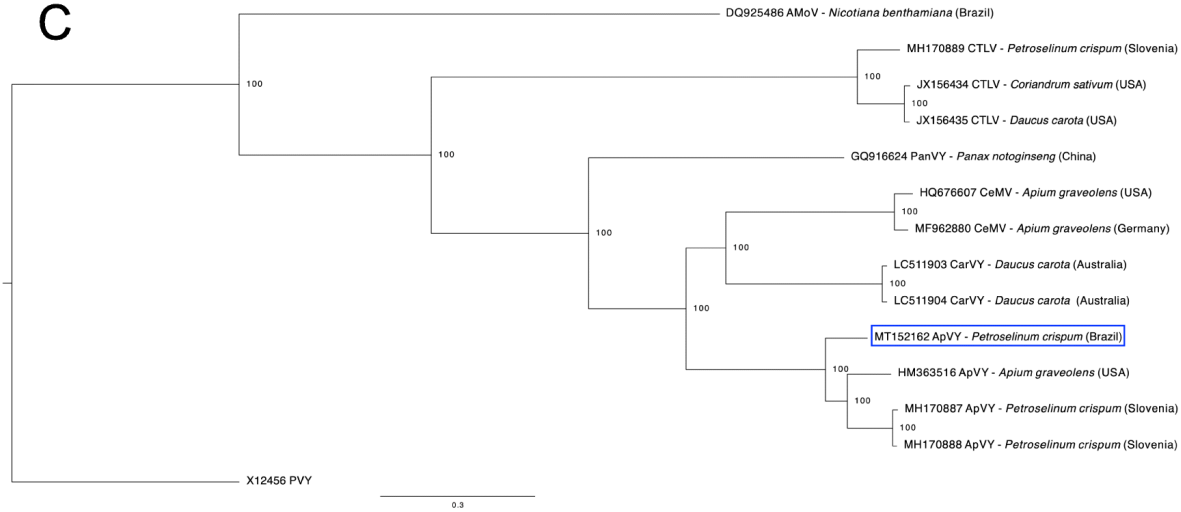
A



B

Proteins	Isolates					
	MH170887		MH170888		HM363516	
	Nt	Aa	Nt	Aa	Nt	Aa
5' UTR	73.1	-	71.4	-	74.8	-
P1	79.2	81.2	79.4	80.9	80.9	79.0
HC-pro	86.0	96.1	86.1	96.1	91.2	97.6
P3	85.4	88.6	85.3	88.4	84.2	88.9
6k1	86.0	98.1	85.0	98.1	84.0	98.1
CI	85.6	96.7	85.7	96.5	85.1	96.5
6k2	84.9	94.3	85.5	94.3	87.4	96.2
VPg	90.4	91.1	89.9	91.1	89.0	92.2
Nla-Pro	83.7	96.7	83.8	96.7	82.3	95.1
NIB	84.4	93.4	84.2	93.8	83.9	94.2
CP	88.2	94.1	88.2	93.8	88.2	92.7
PIPO	90.4	80.7	89.9	79.0	89.0	79.8
Polyprotein	84.3	92.6	84.4	92.6	85.0	92.5
3'UTR	91.73	-	92.12	-	92.91	-

C



D

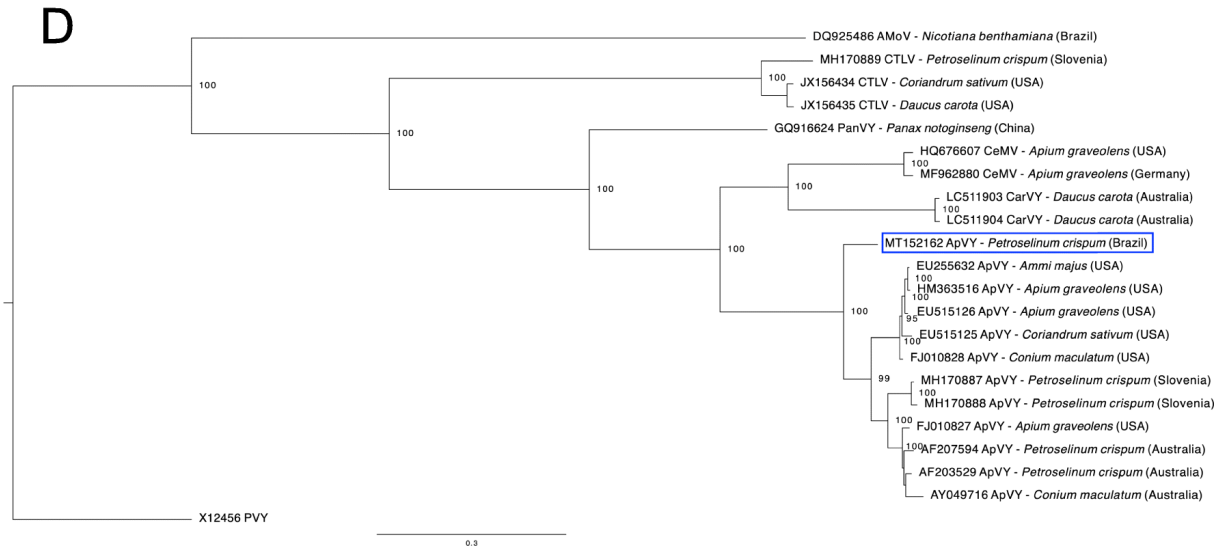


Fig 3. Percentage of nucleotide sequence identity of the Brazilian isolate of apium virus Y (ApVY-Br) (MT152162) and all others apiaceous potyviruses, CeMV (HQ676607, MF962880), CarVY (LC511903, LC511904), PanVY (GQ916624), CTLV (MH170889,

JX156434, JX156435), and AMoV (DQ925486) retrieved from Genbank (A); Percentage of nucleotide (nt) and amino acid (aa) sequence identity of the ApVY-Br isolate to other ApVY complete sequences present in Genbank (B); Bayesian Phylogenetic trees of apiaceous-infecting potyviruses, based on complete sequences (C), and capsid protein sequences (D). ApVY-Br is highlighted in blue. Potato virus Y (PVY) was used as an outgroup

A Bayesian phylogenetic tree was constructed using the previously complete nucleotide sequences dataset, including potato virus Y (X12456) as an outgroup. The sequences were aligned using Muscle v.3.8.425 within the Geneious v.11.1.5 software, and phylogenetic analysis was performed using Mr. Bayes 3.2.2 (Ronquist and Huelsenbeck 2003). The program MrModeltest v.2.2 (Nylander 2004) was used to select the nucleotide substitution model GTR+I+G using the Akaike information criterion (AIC). Two independent runs were conducted simultaneously using 10 million generations and excluding 25% from the resulting trees as burn-in. Phylogenetic trees were visualized using FigTree v1.4.4.

The phylogenetic tree indicated the formation of four main clades supported by 100% posterior probability values (Fig. 3C), also confirmed by the high similarities of the clades on pairwise comparisons using SDT (Fig. 3A). The first three clades were separated according to the potyvirus species, and the fourth clade included the CeMV, the recently published CarVY (Kinoti et al. 2020) and ApVY. Although the number of available sequences is limited, there is a tendency to form separate intraspecific sub-clusters according to geographical origin and independent of the original host. In the ApVY cluster, the Brazilian isolate clustered in sub-cluster alone and was closely related to isolates from Slovenia (MH170887 and MH170888) and the USA (HM363516).

Since few complete ApVY sequences are available in the databases, another phylogenetic tree was constructed based on the capsid protein nucleotide sequence of the previous dataset including eight ApVY sequences from Australia (AY049716, AF203529, and AF207594) and the USA (EU255632, EU515126, EU515125, FJ010828, and FJ010827). In general, even with the addition of new sequences, CP and the complete genome phylogenetic tree displayed a similar topology (Fig. 3D). The ApVY cluster presented four sub-clusters well-supported and segregated based on geographical proximity except for the migrant isolate FJ010827 from the USA, clustered together with isolates from Australia. The Brazilian isolate remained clustered alone, despite the connection with the other isolates.

For partial host range test, symptomatic leaves of parsley plants were ground in 0.05 M potassium phosphate buffer, pH 8.0, as inoculum source for sap inoculation to Chenopodium quinoa plants for biological purification. C. quinoa leaves developed local chlorotic lesions used to produce an infected plant extract for further inoculations. Eight plants of each 11 indicator hosts tested were used, belonging to Amaranthaceae, Apiaceae, Brassicaceae, Chenopodiaceae, and Solanaceae botanical families, including healthy parsley plants to confirm the isolate infectivity (Table 1). For control, two plants of each indicator host were mock-inoculated. All plants were maintained in a greenhouse with 25±2°C, symptoms were evaluated visually every week, and RT-PCR confirmed results.

Table 1 Reaction of plants inoculated with the Brazilian isolate of apium virus Y obtained from parsley.

Plant species	Sap transmission	
	Symptoms[a]	RT-PCR[b]
Amarantaceae		
Sugar Beet (*Beta vulgaris*)	o	-
Apiaceae		
Carrot (*Daucus carota*)	o	-
Celery (*Apium graveolens*)	o	-
Coriander (*Coriandrum sativum*)	Ch, St	+
Parsley (*Petroselinum crispum*)	Ch, LC, St	+
Brassicaceae		
Raphanus sativus	o	-
Chenopodiaceae		
Chenopodium amaranticolor	ChLL, NLL	+
C. quinoa	ChLL, NLL	+
Solanaceae		
Nicotiana benthamiana	o	-
N. glutinosa	o	-
N. tabacum 'TNN'	o	-

[a]Symptom abbreviations: o = no symptom; Ch = chlorosis; ChLL = chlorotic local lesions; LC = leaf crinkle; NLL = necrotic local lesions; St = Stunting.

[b]Inoculated plants were confirmed the infectivity by RT-PCR test using the primers (W-CIEN 5′-ATG GTT TGG TGY ATY GAR AAT-3′ and PV-1 5′-GAT TTA GGT GAC ACT ATA $GT_{(17)}$-3′): − = negative reaction; + = positive reaction.

ApVY has been identified in apiaceous plants, including celery, coriander, parsley, and carrot (Moran et al. 2002; Tian et al. 2008; Xu et al. 2011; Mehle et al. 2019). However, in this study, celery was not infected by ApVY, as well as carrot, suggesting that ApVY-Br may be an isolate with different biological properties.

Chlorosis and stunting were observed on coriander and parsley leaves. Besides, parsley plants presented leaf crinkles. The virus induces chlorotic and necrotic local lesions on *Chenopodium amaranticolor* and *C. quinoa*. The same symptoms were previously observed on C. quinoa (Xu et al. 2011).

Sugar beet, *Raphanus sativus*, and the solanaceous plants, *Nicotiana benthamiana*, *Nicotiana glutinosa*, and *Nicotiana tabacum*, were not infected by ApVY-Br. Sugar beet and *N. glutinosa* were previously demonstrated to be a non-host of ApVY, but *N. benthamiana* showed chlorotic local lesion and mottling symptoms when inoculated with an isolate from celery (Xu et al. 2011).

Here, we presented the first complete genome of a Brazilian ApVY isolate and the first detection of this virus in Brazil. Partial host range assays confirmed the isolate infectivity and its potential to infect mainly coriander and parsley. Apium virus Y is transmitted by aphids on celery and seed-borne transmission has not been reported yet (Koike et al. 2012). The control of the vector, as well as the eradication of weeds that can harbor both the virus and aphids, is recommended for disease management, but more studies are needed about the potential weed hosts of this virus. Growing parsley or coriander in fields with a history of ApVY must be avoided. More fields of apiaceous plants production need to be tested to evaluate the genetic variability of ApVY and provide information about the introduction and dispersion of this virus in Brazil.

Data availability

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

References

Gibbs A, Mackenzie A (1997) A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. Journal of Virological Methods 63:9–16

Kinoti WM, Moran JR, Gambley C, Rodoni BC, Constable FE (2020) Genome characterization of Two Carrot Virus Y isolates from Australia. Microbiology Resource Announcements 9:96–98

Koike ST, Liu H-Y, Sears J, Tian T, Daugovish O, Dara S (2012) Distribution, cultivar susceptibility, and epidemiology of Apium virus Y on celery in Coastal California. Plant Disease 96:612–617

Mehle N, Kutnjak D, Tušek Žnidarič M, Ravnikar M (2019) First report of Apium virus Y and Carrot thin leaf virus in parsley in Slovenia. Plant Disease 103:592

Moran J, van Rijswijk B, Traicevski V, Kitajima EW, Mackenzie AM, Gibbs AJ (2002) Potyviruses, novel and known, in cultivated and wild species of the family Apiaceae in Australia. Archives of Virology 147:1855–1867

Mota LDC, Della Vecchia MGS, Gioria R, Kitajima EW, Rezende JAM, Camargo LEA, Amorim L (2004) Pfaffia mosaic virus: a new potyvirus found infecting Pfaffia glomerata in Brazil. Plant Pathology 53:368–373

Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9:e108277

Novaes QS, Rezende JAM, Kitajima EW (2000) Brazilian celery yellow mosaic virus is serologically related to celery mosaic virus. Summa Phytopatologica 26:250–252

Nylander JAA (2004) MrModeltest Version 2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala

Orílio AF, Dusi AN, Madeira NR, Inoue-Nagata AK (2009) Characterization of a member of a new Potyvirus species infecting arracacha in Brazil. Archives of Virology 154:181–185

Pedroso DC, Menezes VO, Muniz MFB, Piveta G, Tunes LM, Muller J, Menezes NL (2010) Métodos de inoculação de Alternaria alternata e A. dauci em sementes de salsa e sua influência na qualidade fisiológica. Revista Brasileira de Sementes 32:79–85

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Tang J, Clover GRG, Alexander BJR (2007) First report of Apium virus Y in celery in New Zealand. Plant Disease 91:1682–1682

Tian T, Liu HY, Koike ST (2008) First report of Apium virus Y on cilantro, celery, and parsley in California. Plant Disease 92:1254

Wylie SJ, Adams M, Chalam C, Kreuze J, López-Moya JJ, Ohshima K, Praveen S, Rabenstein F, Stenger D, Wang A, Zerbini FM, ICTV Report Consortium (2017) ICTV virus taxonomy profile: potyviridae. Journal of General Virology 98:352–354

Xu D, Liu H, Koike ST, Li F, Li R (2011) Biological characterization and complete genomic sequence of Apium virus Y infecting celery. Virus Research 155:76–82

Xu D, Liu H-Y, Li F, Tian T, Li R (2014) Characterizations of Carrot thin leaf virus based on host reactions and complete genomic sequences. European Journal of Plant Pathology 138:15–22

Yan ZL, Song LM, Zhou T, Zhang YJ, Li MF, Li HF, Fan ZF (2010) Identification and molecular characterization of a new potyvirus from Panax notoginseng. Archives of Virology 155:949–957

Funding

MRRJ is a recipient of a CNPq PhD fellowship (Grant number: 140130/2019-5). RKS and MAP hold CNPq research fellowships. AMN is a FAPESP post-doctoral fellow. “This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.”

Contributions

Marcos Roberto Ribeiro-Junior: conceptualization, experimental procedures, writing, review and editing. Eduardo Silva Gorayeb, Angélica Maria Nogueira, Giovana Carolina Dourado Cruciol, and Daniele Maria Nascimento: experimental procedures,

writing, review, and editing. Marcelo Agenor Pavan and Renate Krause Sakate: conceptualization, funding acquisition, supervision, review, and editing. All authors have read and agreed to the published version of the manuscript.

Corresponding author

Correspondence to Marcos Roberto Ribeiro-Junior.

Ethics declarations

Ethics approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate

The authors declare their consent to publish the data presented in this research.

Conflict of interest

The authors declare no competing interests.

Additional information

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.