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

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”.

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.