PhytoFrontiers™
An Open Access Journal from The American Phytopathological Society

Research

An Attempt Toward the Global Screening of Soybean Viruses Using EDNA-MiFi-Based Electronic Probes

Marcos R. Ribeiro-Junior[1,2,3,†] | Andres Espindola[1,2] | Daniele M. Nascimento[1,2] | Felipe Barreto da Silva[3,4] | Renate Krause-Sakate[3] | Francisco M. Ochoa-Corona[1,2,†]

[1] Institute for Biosecurity and Microbial Forensics, Oklahoma State University, Stillwater, OK, U.S.A.

[2] Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK, U.S.A.

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

[4] Gulf Coast Research and Education Center, University of Florida, Wimauma, FL, U.S.A.

Accepted for publication 20 May 2025.

† Corresponding authors: F. M. Ochoa-Corona; ochoaco@okstate.edu, and M. R. Ribeiro-Junior; marcos.ribeiro@okstate.edu.

Author contributions: M.R.R.-J., F.M.O.-C., and R.K.-S. initiated the project. F.B.d.S. performed the validation PCR analyses. M.R.R.-J., F.M.O.-C., A.E., D.M.N., and R.K.-S. analyzed the data. M.R.R.-J., D.M.N., R.K.-S., and F.M.O.-C. contributed to writing the manuscript. All authors have read and agreed to the published version of the manuscript.

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.

Data availability: The nucleotide sequences reported in this study are available from the corresponding author upon reasonable request.

Funding: M. R. Ribeiro-Junior received Ph.D. fellowships from CNPq and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)-PrInt (grants 140130/2019-5 and 88887.570292/2020-00). F. M. Ochoa-Corona holds a Sarkeys Foundation Professorship III. R. Krause-Sakate holds a CNPq research fellowship (grant 303411/2018-0). This study was financed in part by CAPES, finance code 001.

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

Abstract

Soybean (*Glycine max*) is a globally significant crop renowned for its oilseed and protein content. Brazil and the United States are the leading producers worldwide. However, viral diseases are a persistent threat to the soybean industry, resulting in annual yield losses. Accurate and rapid diagnosis is essential for effective disease management. This study presents the application of Electronic-probe Diagnostic Nucleic-acid Analysis and Microbe Finder (EDNA-MiFi), a web-based software designed for rapid identification of predetermined pathogen species in high-throughput sequencing (HTS) datasets. MiFi uses curated electronic probes (e-probes) specific to target sequences, enabling screening of raw, unassembled HTS data. E-probes (20 to 60 nt) for 46 known soybean-infecting viruses were designed using MiProbe, a MiFi platform component. E-probe curation and validation involved BLASTn analyses and the creation of mock HTS data consisting of positive controls of reference virus and host genome sequences. The *in silico* analysis confirmed the efficacy of the e-probes in detecting all 46 viruses in the metatranscriptomic soybean data. The *in vitro* analysis was conducted with Illumina HTS data from soybean collected in Brazil. Results showed infections with nine virus species, which were further validated through PCR, RT-PCR, and mapping of reads to virus reference genomes. This study demonstrates that MiFi enables rapid (less than an hour) detection of targeted virus sequences in raw HTS outputs based on the screening of all currently reported soybean-infecting viruses. The curated e-probes specific to soybean viruses are accessible for diagnostic purposes via MiDetect, another component of the MiFi platform.

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

The relevance of soybean (*Glycine max* (L.) Merr.) is described by a combination of attributes and adjectives, such as golden bean, super legume, the crop of the planet, or just the miracle bean. Soybean is a staple crop worldwide used for food, feed, and fuel (Rahman et al. 2023). In 2023 and 2024, global soybean production reached 394.73 million metric tons, with Brazil accounting for 39% (153.00 million tons) of the total,

making it the world's leading producer and exporter. Brazil also led soybean exports, contributing 105 million tons of the globally traded soybean production, whereas the United States contributed 49 million tons (USDA 2024).

Plant viruses significantly impact soybean production, with more than 100 virus species known to infect soybeans. Of these, at least 46 have been detected in naturally occurring field infections (Elmore et al. 2022; Hill and Whitham 2014). Among these viruses, soybean mosaic virus (SMV), a cosmopolitan aphid- and seed-transmitted virus, represents a substantial risk by causing yield crop losses of up to 86% (Hajimorad et al. 2018).

Accurate and rapid detection is crucial for delimiting and managing the spread of new viruses in a region, particularly if the presence of the targeted viruses has not been previously reported (Espindola and Cardwell 2021). In 2022, a new potyvirus, passiflora virus Y (PaVY), was identified in soybean crops in Brazil, highlighting the need to improve existing detection methods (Ribeiro-Junior et al. 2022).

Biosecurity measures are crucial in preventing disease spread by minimizing the risks of pathogen entry and infection in susceptible crops. Seed biosecurity involves monitoring, restricting, and controlling pathogen movement across national and regional borders (Oliver 2022).

Sequence-based detection technology offers a promising alternative for routine detection by plant quarantine agencies, enabling simultaneous screening for multiple viruses (Maree et al. 2018). High-throughput sequencing (HTS) has become an indispensable tool for virus diagnosis due to its capacity for comprehensive virus detection in samples (Kutnjak et al. 2021). However, the large volume of data generated by HTS presents a challenge, extending the time required to reach a definitive diagnostic decision. Electronic-probe diagnostic nucleic-acid analysis (EDNA) has been applied to address this gap between HTS and pathogen detection (Stobbe et al. 2013).

The EDNA pipeline is accessible online through two web-based platforms. The Legacy-MiFi software (https://bioinfo.okstate.edu), hosted on the Oklahoma State University (OSU) High Performance Computing Center's research cloud, is available to researchers and includes the MiProbe component for designing e-probes—short, pathogen-specific nucleic acid sequences used to detect target organisms in metagenomic data. The MiFi.Tech software (https://app.mifi.tech), hosted on Amazon Web Services (AWS), incorporates the MiDetect component, which utilizes curated e-probes to query raw HTS data and identify pathogen sequences.

This study aimed to develop a bioinformatics tool that integrates HTS for detection of multiple viruses previously reported to infect soybeans worldwide through the MiFi platform. The tool rapidly detects known viral species in unassembled HTS metagenomic data. For this purpose, the study focused on the design, curation, and *in silico* and *in vitro* validation of e-probes. Specifically, *in silico*-curated e-probes were developed for 46 virus species reported to infect soybean crops worldwide to enable global detection of soybean viruses.

Materials and Methods

To develop and validate e-probes capable of detecting all previously reported soybean-infecting viruses worldwide as of September 2021, we used the MiProbe pipeline described by Stobbe et al. (2013), Ochoa-Corona et al. (2019), and Espindola and Cardwell (2021). Complete and partial genome sequences of each virus species, including their respective isolates, were obtained from the public database of GenBank (Supplementary Table S1). These sequences formed the basis for designing and curating e-probes specific to each virus species using the MiProbe pipeline (Fig. 1) to ensure specificity and broad coverage across virus strains and isolates for accurate detection.

Raw e-probe design

A total of 46 viruses (Table 1; Supplementary Table S2) were selected as target genomes for generating e-probes. For each target virus, a set of taxonomically related near neighbors—closely related species or strains not expected to be detected—was selected to assess specificity. Near-neighbor genomes were retrieved from GenBank, focusing on viruses within the same genus or family as the target virus, with priority given to those known to infect legumes or commonly found in similar agroecosystems. This approach ensured that the e-probes yielded positive detection signals only for their intended targets while minimizing false positives.

The complete or partial genomes of all available isolates of each virus species were concatenated into a single multi-fasta file. Similarly, the genomes of the near-neighbor virus species were concatenated into a separate multi-fasta file. The two files were uploaded to the Legacy-MiFi platform, where the MiProbe function was used to generate short e-probe sequences. The e-probes ranged from 20 to 60 nucleotides in length, which is optimal for virus detection (Espindola and Cardwell 2021). The generated e-probes were designed to match target genomes specifically while minimizing cross-reactivity with near-neighbor sequences.

E-probe curation: Specificity and cross-reactivity

The curation process involved eliminating e-probe sequences that could match organisms other than the intended viral target and thus cause false positive results (Dang et al. 2022; Espindola and Cardwell 2021). An initial, extensive set of raw e-probe sequences was generated for each target virus. These e-probes were then evaluated for specificity and cross-reactivity through BLASTn searches against the GenBank nucleotide database. Any e-probes that matched nontarget taxa—defined as sequences outside the intended virus's NCBI taxonomy ID and its subtaxa—

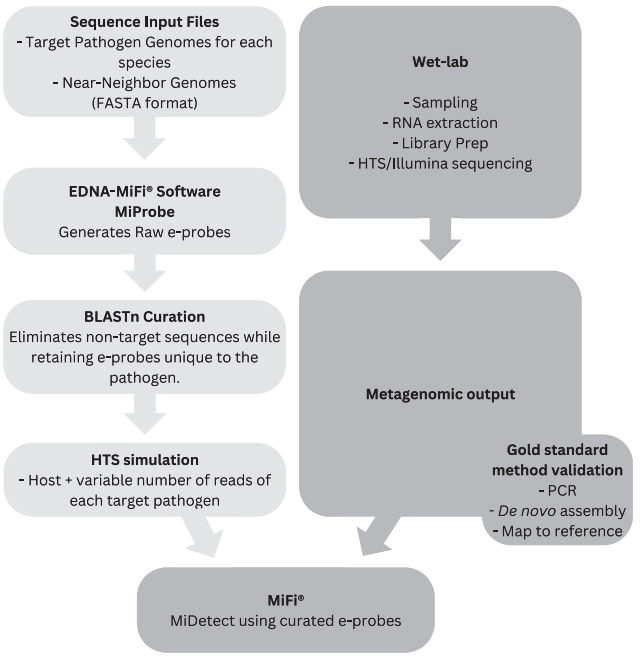

FIGURE 1
Overview of the MiFi pipeline for generating and validating e-probes. HTS, high-throughput sequencing.

with ≥90% identity and ≥90% query coverage were removed using a custom R script. This taxonomy-aware filtering ensured that only highly specific e-probes were retained, minimizing the risk of cross-reactivity. Additionally, duplicate e-probes were identified and removed before curation using a Perl script to ensure uniqueness across each set.

Limit of detection and validation *in silico*

To determine the *in silico* limit of detection (LoD)—defined as the minimum number of pathogen-derived reads in a metagenomic dataset that can be consistently identified as positive—simulated HTS data were generated using MetaSim (Richter et al. 2008) with 100-bp single-end reads. The soybean genome (*Glycine max* Wm82.a2.v1; NCBI accession GCF_000004515.6) was incorporated alongside varying proportions of each of the 46 virus species. The pathogen concentrations ranged from 0 to 5,000 pathogen reads to represent different infection levels.

Ten replicates were performed for each virus concentration to ensure statistical reliability. Each replicate consisted of a unique set of independently simulated reads. While the total number of reads and the proportion of pathogen reads were consistent across replicates, the sequences varied to mimic natural sampling variation. The LoD was calculated based on a threshold where at least 9 out of the 10 replicates had to yield a positive detection result. Simulations were conducted with a maximum of 1 million total reads per dataset, allowing for a robust evaluation of pathogen detectability. The LoD for each virus was defined as the lowest number of pathogen reads that resulted in a positive detection in at least nine out of ten replicates.

The results were expressed as the minimum relative abundance of pathogen reads to detect (MRAD), representing the proportion of pathogen reads within the simulated dataset. For instance, a LoD of 10 reads in a 1 million-read dataset corresponds to a relative abundance of 0.001%.

Validation with field samples

Leaf samples exhibiting virus-like symptoms were collected from soybean fields in São Paulo State, Brazil, over three growing seasons (2018 to 2020), comprising one sample in 2018, three in 2019, and three in 2020. Total RNA was extracted from these

TABLE 1

Genus, virus species, acronym/e-probe ID, genome size (kb), and number of raw and curated e-probes for viruses infecting soybean worldwide

Genus	Viral species	Acronym	Genome size (kb)	Raw e-probes	Curated e-probes
Alfamovirus	Alfalfa mosaic virus	AMV	3.64, 2.59, 2.04	1,197	213
Begomovirus	Abutilon mosaic virus	AbMV	2.63, 2.59	19	13
	Bean golden mosaic virus	BGMV	2.62, 2.58	414	38
	Euphorbia mosaic virus	EuMV	2.61, 2.6	73	20
	Horsegram yellow mosaic virus	HgYMV	2.73, 2.68	29	14
	Mungbean yellow mosaic virus	MYMV	2.72, 2.66	220	54
	Okra mottle virus	OMV	2.66, 2.55	14	7
	Rhynchosia golden mosaic virus	RhGMV	2.55, 2.6	60	15
	Sida micrantha mosaic virus	SimMV	2.68, 2.66	75	63
	Soybean chlorotic spot virus	SoCSV	2.62, 2.59	4	1
	Soybean crinkle leaf virus	SCLV	2.74	12	3
Bromovirus	Cowpea chlorotic mottle virus	CCMV	3.17, 2.77, 2.17	87	43
Carlavirus	Cowpea mild mottle virus	CPMMV	8.13	506	190
Comovirus	Bean pod mottle virus	BPMV	6, 3.66	284	92
	Broad bean true mosaic virus	BBTMV	6.07, 3.53	64	31
	Cowpea mosaic virus	CPMV	5.89, 3.48	42	19
	Cowpea severe mosaic virus	CPSMV	5.96, 3.73	99	36
Cucumovirus	Cucumber mosaic virus	CMV	3.36, 3.05, 2.22	7,472	1,017
	Peanut stunt virus	PSV	3.36, 2.95, 2.19	252	115
Enamovirus	Pea enation mosaic virus	PEMV	5.71	305	136
Fabavirus	Broad bean wilt virus	BBWV	5.82, 3.45	1,282	779
Gammacarmovirus	Soybean yellow mottle mosaic virus	SYMMV	4.01	98	9
Ilarvirus	Tobacco streak virus	TSV	3.49, 2.93, 2.21	444	143
Luteovirus	Bean leaf roll virus	BLRV	5.96	125	45
	Soybean dwarf virus	SbDV	5.85	663	213
Nepovirus	Tobacco ringspot virus	TRSV	7.51, 3.93	145	87
	Tomato ringspot virus	ToRSV	8.21, 7.27	22	13
Orthotospovirus	Groundnut ringspot virus	GRSV	8.63, 3.4, 0.78	83	24
	Soybean vein necrosis virus	SVNV	9.01, 4.96, 2.6	105	34
	Tomato spotted wilt virus	TSWV	8.9, 4.82, 2.92	1,092	213
Potyvirus	Azuki bean mosaic virus	AzMV	No full genome	4	4
	Bean common mosaic virus	BCMV	9.99	15	15
	Bean yellow mosaic virus	BYMV	9.53	601	183
	Blackeye cowpea mosaic virus	BlCMV	9.92	15	15
	Passiflora virus Y	PaVY	9.67	9	9
	Passion fruit woodiness virus	PWV	9.62	58	24
	Peanut mottle virus	PMotV	9.71	107	41
	Peanut stripe virus	PStV	10.09	25	25
	Soybean mosaic virus	SMV	9.59	199	59
	Soybean yellow shoot virus	SoyYSV	9.05	14	14
Sobemovirus	Southern bean mosaic virus	SBMV	4.13	73	30
	Soybean yellow common mosaic virus	SYCMV	4.15	170	50
Soymovirus	Soybean chlorotic mottle virus	SbCMV	8.17	48	32
Tobamovirus	Sunn-hemp mosaic virus	SHMV	6.43	30	12
	Tobacco mosaic virus	TMV	6.39	442	58
Tobravirus	Tobacco rattle virus	TRV	6.79, 3.85	767	284

symptomatic leaves using the PureLink Viral RNA/DNA Mini Kit (Invitrogen, 12280050) following the manufacturer's instructions. Seven samples were treated for ribosomal depletion using the RiboMinus Plant Kit for RNA-Seq (Invitrogen, A1083808), and RNA quality was assessed using the Agilent 2100 Bioanalyzer. Complementary DNA (cDNA) libraries were prepared using the ScriptSeq Complete Kit (Epicenter, Illumina). Transcriptome sequencing was performed on the Illumina HiSeq 2500 platform using the HiSeq Flow Cell v4 and HiSeq SBS v4 kits (Illumina) in paired-end mode (2 × 100 bp) at the Centro de Genômica Funcional–ESALQ/USP.

The resulting seven Illumina output datasets were analyzed using the newly curated e-probes. In addition, positive virus matches were confirmed by PCR and RT-PCR using GoTaq Green Master Mix (Promega, M712), AMV Reverse Transcriptase (Promega, M510A), and the primers listed in Supplementary Table S3. All reactions were performed on a Mastercycler Nexus Thermal Cycler (Eppendorf, EP6334000026).

The Illumina metatranscriptomic datasets were uploaded to the MiFi platform and screened using the MiDetect function, which was set with the recommended parameters for virus detection as outlined by Espindola and Cardwell (2021). The e-value threshold was set at 1e-1, and the minimum number of reads required for detection was designated as "sensitive," requiring a minimum of 10 hits.

Reference genomes of all viruses detected using MiFi were obtained from GenBank to validate the effectiveness of MiDetect. The HTS reads were then mapped to these reference virus genomes utilizing Geneious software v11.1.6, confirming the presence of virus reads in the soybean Illumina HTS samples.

The generated metatranscriptomic reads were quality-trimmed and *de novo*-assembled using CLC Genomics Workbench v9.0.3 with the NGS Core Tools module to ensure no other viruses were present in the sample. The resulting contigs were then aligned against the GenBank viral database using BLASTn to identify potential viral sequences in the sample.

Results

E-probe design and curation

The number of generated e-probe sequences was significantly reduced for many viruses after curation. For example, the number of e-probes for cucumber mosaic virus (CMV) decreased from 7,472 raw e-probes to 1,017 curated e-probes, representing an 86.4% reduction (Table 1). However, for some viruses, such as azuki bean mosaic virus (AzMV), PaVY, and soybean yellow shoot virus (SoyYSV), the number of e-probes remained unchanged before and after curation, indicating that the initial set already met the specificity and coverage criteria, likely due to limited sequence similarity between these viruses and their respective near-neighbor genomes.

Limit of detection and validation *in silico*

The *in silico* validation of the e-probes showed that the minimum number of pathogen reads required for detection ranged from 1 to 5,000, corresponding to a relative abundance of 0.0001 to 0.5% in a metagenome containing 1 million host reads (Table 2). While some e-probe sets, such as that for CMV, could detect as few as a single pathogen read *in silico*, such minimal detections are not recommended as diagnostic confirmation due to the possibility of read carryover or cross-contamination in HTS datasets. Therefore, although these simulations indicate high sensitivity, especially with larger e-probe sets (e.g., CMV with 1,017 e-probes), detection thresholds should be interpreted conservatively in real-world diagnostic contexts. This correlation between the number of e-probes and LoD suggests that larger e-probe sets improve detection sensitivity for low-abundance viruses, enhancing diagnostic potential in metagenomic data. The MiFi platform allows users to define the minimum number of hits per e-probe, ranging from highly sensitive (1 hit) to more stringent thresholds, such as 10 or 250 hits, to accommodate varying diagnostic requirements, which can reduce the likelihood of false positives in complex datasets.

Validation with field samples

The set of curated e-probes loaded into MiDetect successfully detected nine virus species across the soybean field samples collected in Brazil (Table 3). Multiple viruses were detected in several samples, indicating the presence of co-infections. Although the samples were collected from plants showing virus-like symptoms, no correlation was made between specific symptoms and the detected viruses. This validation confirms the effectiveness of the respective e-probe sets, demonstrating their potential for use as a reliable and rapid diagnostic method.

TABLE 2

Minimum relative abundance of pathogen reads to detect (MRAD), calculated after *in silico* simulations over a million total reads

Viral species	MRAD (limit of detection)
Alfalfa mosaic virus	0.0003%
Abutilon mosaic virus	0.002%
Bean golden mosaic virus	0.001%
Euphorbia mosaic virus	0.001%
Horsegram yellow mosaic virus	0.0011%
Mungbean yellow mosaic virus	0.001%
Okra mottle virus	0.0025%
Rhynchosia golden mosaic virus	0.001%
Sida micrantha mosaic virus	0.001%
Soybean chlorotic spot virus	0.5%
Soybean crinkle leaf virus	0.45%
Cowpea chlorotic mottle virus	0.001%
Cowpea mild mottle virus	0.0007%
Bean pod mottle virus	0.001%
Broad bean true mosaic virus	0.0006%
Cowpea mosaic virus	0.001%
Cowpea severe mosaic virus	0.001%
Cucumber mosaic virus	0.0001%
Peanut stunt virus	0.001%
Pea enation mosaic virus	0.001%
Broad bean wilt virus	0.0009%
Soybean yellow mottle mosaic virus	0.0017%
Tobacco streak virus	0.0008%
Bean leaf roll virus	0.001%
Soybean dwarf virus	0.0006%
Tobacco ringspot virus	0.0009%
Tomato ringspot virus	0.013%
Groundnut ringspot virus	0.0025%
Soybean vein necrosis virus	0.0015%
Tomato spotted wilt virus	0.0009%
Azuki bean mosaic virus	0.08%
Bean common mosaic virus	0.003%
Bean yellow mosaic virus	0.001%
Blackeye cowpea mosaic virus	0.005%
Passiflora virus Y	0.004%
Passion fruit woodiness virus	0.0014%
Peanut mottle virus	0.0011%
Peanut stripe virus	0.0040%
Soybean mosaic virus	0.0015%
Soybean yellow shoot virus	0.0035%
Southern bean mosaic virus	0.0007%
Soybean yellow common mosaic virus	0.0006%
Soybean chlorotic mottle virus	0.0015%
Sunn-hemp mosaic virus	0.0025%
Tobacco mosaic virus	0.001%
Tobacco rattle virus	0.0008%

De novo assembly of the metatranscriptomic reads followed by BLASTn searches against the NCBI viral database confirmed that no viral sequences outside the target list were present in the datasets.

For each virus-positive sample, the corresponding reads were mapped to reference genomes using Geneious software to further verify the presence of viral sequences (Table 4). The number of mapped reads varied significantly across the samples, ranging from 4,068 reads for groundnut ringspot virus (GRSV) in sample 5 to 9,391,113 reads for cowpea mild mottle virus (CPMMV) in sample 1.

Discussion

In this study, e-probes for the *in silico* detection of 46 soybean viruses were successfully designed, enabling rapid and efficient detection of viruses from HTS using MiFi software. These results were confirmed *in vitro* for nine of these viruses. The operational simplicity of the MiFi interface eliminates the need for advanced bioinformatics skills typically required for HTS data analysis. Compared with conventional HTS-based diagnostic pipelines that require trimming, host read removal, *de novo* assembly, and extensive bioinformatics expertise, the EDNA-MiFi platform offers a rapid and user-friendly alternative. The platform is compatible with Illumina and Oxford Nanopore reads (Espindola and Cardwell 2021), and results can be obtained without data preprocessing or command-line steps. These features reduce the need for specialized infrastructure, making the method suitable for implementation in routine diagnostic labs. Moreover, validation with naturally infected soybean field samples demonstrated the method's ability to detect target viruses even in complex microbial backgrounds, supporting its relevance for real-world agricultural applications.

On average, MiFi provided results within 15 min for a 5 GB FASTQ file (approximately 10 to 13 million 100-bp single-end reads). For context, Espindola and Cardwell (2021) reported that Metaphlan3 (Beghini et al. 2021) required an average of 1 h and 12 min, whereas Kraken2 (Wood and Salzberg 2014) took approximately 2 h and 42 min for similar analyses. MiFi runs on Amazon Web Services, which enables scalable, cloud-based processing without relying on local hardware. This allows users to perform analyses quickly and reliably regardless of their computing resources, making MiFi especially suitable for diagnostic and regulatory labs with limited IT infrastructure or bioinformatics support. The significantly shorter detection time of MiFi makes this platform a valuable tool for faster and more efficient pathogen detection and identification, especially in scenarios in which timely responses are critical. The speed of analysis is influenced by factors such as the number of e-probes used, the size of the target genome, and the size of HTS data being processed. The consistent advantage in detection time demonstrated by MiFi highlights its efficiency as a bioinformatics tool for detection of a large set of predetermined viruses and reinforces its potential for real-time diagnostics and surveillance applications.

The versatility of the MiFi platform has been previously demonstrated in various studies involving the detection of different pathogens, including cucurbit viruses, water-borne plant viruses, citrus leprosis syndrome-associated viruses, grapevine pathogens, blueberry pathogens, fungi, oomycetes, and aflatoxin-producing organisms in soil (Bocsanczy et al. 2023; Espindola and Cardwell 2021; Espindola et al. 2018, 2019; Proaño-Cuenca et al. 2023; Zuniga et al. 2017, 2018).

The validation process for the developed e-probes involves *in silico* curation, *in vitro* validation using HTS datasets from field samples, and confirmation by PCR/RT-PCR. The curation of the e-probes using the GenBank nucleotide (nt) database is essential to eliminate sequences that could align with other pathogens not initially included in the near-neighbor list. Notably, some e-probe sets—such as AzMV, PaVY, and SoyYSV—required no filtering during curation, suggesting high initial specificity due to limited similarity with near-neighbor genomes.

Although field samples were available for only 9 of the 46 viruses, the e-probes tested *in silico* demonstrated specificity and provided insights into their potential use for *in vitro* diagnostics, which remains an ongoing area of research. Moreover, several viruses are exotic pests in the United States and Brazil and should be tested overseas.

The nine viruses confirmed in samples from Brazil included bean golden mosaic virus (BGMV), Sida micrantha mosaic virus (SiMV), cowpea chlorotic mottle virus (CCMV), CPMMV, CMV, tobacco streak virus (TSV), GRSV, bean common mosaic virus (BCMV), and PaVY. All viruses identified in Brazil had been previously reported in the country (Kitajima 2020; Ribeiro-Junior et al. 2022). Among these, CMV, TSV, GRSV, and BCMV have been reported in the United States, impacting various crops, including soybean (Davis and Hampton 1986; Mtonga and Maruthi 2024; Rabedeaux et al. 2005; Rizzo and Palukaitis 1989; Webster et al. 2015).

CPMMV was the most frequently identified virus among field samples. This virus has been increasingly reported in South America and is associated with chlorosis, stunting, and reduced soybean seed quality in soybeans. Previous studies have shown that CPMMV infections can result in yield losses ranging from 10 to 40%, depending on the cultivar and environmental conditions (Barreto da Silva et al. 2020, 2024). The high prevalence of CPMMV is likely linked to the widespread occurrence of whiteflies (*Bemisia tabaci*) in Brazilian agro-ecosystems, which serve as the primary vector for its transmission. This underscores the importance of continued monitoring and implementation of effective management strategies in affected regions.

The level of validation achieved by EDNA methods depends on several factors, including the availability of positive controls, laboratory resources, and field samples (Espindola and Cardwell 2021).

The validated e-probes for soybean viruses provide a valuable resource for researchers and diagnosticians and are made publicly available through two online platforms. The Legacy-MiFi software (accessible at https://bioinfo.okstate.edu) offers free access with a storage capacity limited to 200 GB. Alternatively, the MiFi.Tech software (available at https://app.mifi.tech) provides

TABLE 3

Viruses detected in high-throughput sequencing metafiles from samples collected in Brazil with MiFi-MiDetect using curated e-probes

E-probe ID[a]	MiFi diagnostic	Samples
BGMV	Positive	3
SimMV	Positive	5
CCMV	Positive	1
CPMMV	Positive	1, 2, 3, 4, 5, 6, 7
CMV	Positive	1, 3, 4, 5, 6, 7
TSV	Positive	3, 4
GRSV	Positive	1, 3, 5, 6, 7
BCMV	Positive	2
PaVY	Positive	7

[a] BGMV, bean golden mosaic virus; SimMV, Sida micrantha mosaic virus; CCMV, cowpea chlorotic mottle virus; CPMMV, cowpea mild mottle virus; CMV, cucumber mosaic virus; TSV, tobacco streak virus; GRSV, groundnut ringspot virus; BCMV, bean common mosaic virus; PaVY, Passiflora virus Y.

a modernized interface and supports unlimited storage capacity upon user registration, ensuring broader accessibility and scalability for data analysis. The EDNA-MiFi platform offers new opportunities for additional experimentation toward validation and improvement, testing of new virus combinations, and field samples infected with mixed viral infections, which will contribute to the refinement of the method in diverse agricultural and diagnostic settings.

Determining the LoD and specificity of the e-probes is crucial, particularly if the EDNA-MiFi method being developed will be applied in diagnostics laboratories focusing on detecting of exotic or high-risk pathogen species (Espindola and Cardwell 2021). In our study, the *in silico* LoD ranged from 0.0001 to 0.5% for 1 million reads, demonstrating the tool's ability to detect even low-abundance viral pathogens with high accuracy. Nonetheless, single-read detections should be interpreted with caution due to the risk of noise or contamination.

The developed e-probe tool shows significant potential for preventing the global spread of major soybean viruses, especially considering the international trade of seeds and seedlings. EDNA-MiFi implementation can enable rapid and precise pathogen detection, which is essential for timely interventions. EDNA-MiFi contributes to strengthening biosecurity measures and protecting global agricultural systems, preventing the introduction and

TABLE 4

Number of read matches with reference virus genomes and percentage of genome coverage

Positive e-probe matches and their reference virus accession number[a]	Samples (total reads per high-throughput sequencing sample)						
	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	–	–	3,213[b] (92.4%[c])	–	–	–	–
NC_004043.1	–	–	194[b] (99.3%[c])	–	–	–	–
SimMV							
NC_005330.1	–	–	–	–	938[b] (100%[c])	–	–
NC_005331.1	–	–	–	–	2,446[b] (100%[c])	–	–
CCMV							
NC_003543.1	1,365[b] (98.3%[c])	–	–	–	–	–	–
NC_003541.1	2,785[b] (98.1%[c])	–	–	–	–	–	–
NC_003542.1	15,068[b] (97,8%[c])	–	–	–	–	–	–
CPMMV							
NC_014730.1	9,391,113[b] (100%[c])	6,394,631[b] (100%[c])	3,161,251[b] (100%[c])	1,1686,900[b] (100%[c])	1,152,927[b] (96.4%[c])	2,922,714[b] (100%[c])	4,815,528[b] (100%[c])
CMV							
NC_002034.1	320[b] (97.3%[c])	–	36,065[b] (57%[c])	36,065[b] (75.2%[c])	51,855[b] (93.7%[c])	38,176[b] (96.3%[c])	20,727[b] (90.4%[c])
NC_002035.1	86[b] (86.5%[c])	–	100,380[b] (100%[c])	108,628[b] (100%[c])	87,864[b] (100%[c])	77,544[b] (100%[c])	471,541[b] (100%[c])
NC_001440.1	16,175[b] (98.3%[c])	–	14,197[b] (100%[c])	102,457[b] (100%[c])	270,620[b] (79.6%[c])	526,058[b] (87.4%[c])	741,888[b] (100%[c])
TSV							
NC_003844.1	–	–	488,593[b] (99.8%[c])	1,983[b] (52.3%[c])	–	–	–
NC_003842.1	–	–	397,623[b] (78%[c])	54,465[b] (77.4%[c])	–	–	–
NC_003845.1	–	–	1,428,425[b] (100%[c])	10,793[b] (85.9%[c])	–	–	–
GRSV							
NC_043503.1	23,767[b] (100%[c])	–	33,405[b] (100%[c])	–	26[b] (83%[c])	669,918[b] (100%[c])	597,712[b] (100%[c])
NC_043502.1	20,308[b] (100%[c])	–	77,224[b] (100%[c])	–	23[b] (41.4%[c])	510,703[b] (100%[c])	626,610[b] (100%[c])
NC_043504.1	150,485[b] (100%[c])	–	135,025[b] (100%[c])	–	4,019[b] (56.6%[c])	2,806,220[b] (100%[c])	2,756,858[a] (100%[c])
BCMV							
NC_003397.1	–	91,770[b] (100%[c])	–	–	–	–	–
PaVY							
MW165064	–	–	–	–	–	–	2,231,517[b] (100%[c])

[a] BGMV, bean golden mosaic virus; SimMV, Sida micrantha mosaic virus; CCMV, cowpea chlorotic mottle virus; CPMMV, cowpea mild mottle virus; CMV, cucumber mosaic virus; TSV, tobacco streak virus; GRSV, groundnut ringspot virus; BCMV, bean common mosaic virus; PaVY, Passiflora virus Y.
[b] Number of reads mapped to reference genome sequences.
[c] Percentage of genome coverage; –, virus absent in the sample.

spread of harmful viral species in soybean production. Rapidly screening multiple viruses in a single sample facilitates accurate pathogen detection, thereby contributing to effective biosecurity measures. While this study did not assess virus presence in seeds directly, several of the detected viruses are known to be seed-transmissible, and future work will focus on validating the assay using soybean seed and seedling material to support applications in seed health certification.

Conclusion

The successful development and validation of e-probes for detecting 46 soybean virus species using the MiFi platform is a promising approach to mitigating the global dissemination of these significant pathogens. Through meticulous curation and comprehensive *in silico* testing, the e-probes have shown exceptional specificity and sensitivity, ensuring precise identification of the targeted viruses while minimizing the occurrence of false positive results. The user-friendly interface and superior performance of the MiFi platform render it an invaluable tool for rapid and reliable pathogen detection. The integration of this advanced e-probe technology into biosecurity protocols will enhance the implementation of measures for pathogen surveillance and detection. The ability to promptly and accurately detect soybean viruses will help protect agricultural systems from potential large-scale outbreaks and their devastating consequences. This approach reinforces global biosecurity efforts, safeguarding crop health and the sustainability of soybean production worldwide.

Literature Cited

Barreto da Silva, F., Muller, C., Bello, V. H., Watanabe, L. F. M., Rossitto De Marchi, B., Fusco, L. M., Ribeiro-Junior, M. R., Minozzi, G. B., Vivan, L. M., Tamai, M. A., Farias, J. R., Nogueira, A. M., Sartori, M. M. P., and Krause-Sakate, R. 2020. Effects of cowpea mild mottle virus on soybean cultivars in Brazil. PeerJ 8:e9828.

Barreto da Silva, F., de Sarandy Raposo, R., de Campos, S. F., Uzan, J., Marubayashi, J. M., Ribeiro-Junior, M. R., Nogueira, A. M., da Cruz Martines, C., Bello, V. H., Müller, C., Sartori, M. M. P., and Krause-Sakate, R. 2024. Exploring *Bemisia tabaci* Middle East–Asia Minor I and Mediterranean cryptic species relationship with cowpea mild mottle virus and their dynamics in soybean fields. Insects 15:624.

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., and Segata, N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife 10:e65088.

Bocsanczy, A. M., Espindola, A. S., Cardwell, K., and Norman, D. J. 2023. Development and validation of e-probes with the MiFi system for detection of *Ralstonia solanacearum* species complex in blueberries. PhytoFrontiers 3:137-147

Dang, T., Espindola, A., Vidalakis, G., and Cardwell, K. 2022. An in silico detection of a citrus viroid from raw high-throughput sequencing data. Pages 275-283 in: Viroids: Methods and Protocols. A. L. N. Rao, I. Lavagi-Craddock, and G. Vidalakis, eds. Springer US, New York, NY, U.S.A.

Davis, R. F., and Hampton, R. O. 1986. Cucumber mosaic virus isolates seedborne in *Phaseolus vulgaris*: Serology, host-pathogen relationships, and seed transmission. Phytopathology 76:999-1004.

Elmore, M. G., Groves, C. L., Hajimorad, M. R., Stewart, T. P., Gaskill, M. A., Wise, K. A., Sikora, E., Kleczewski, N. M., Smith, D. L., Mueller, D. S., and Whitham, S. A. 2022. Detection and discovery of plant viruses in soybean by metagenomic sequencing. Virol. J. 19:149.

Espindola, A., Roy, A., Mavrodieva, V. A., and Cardwell, K. F. 2019. E-probe: A new diagnostic tool for detection of Dichorhaviruses associated with Citrus leprosis syndrome. Plant Health 2019. APS Annual Meeting, Cleveland, Ohio, U.S.A.

Espindola, A. S., and Cardwell, K. F. 2021. Microbe finder (MiFi®): Implementation of an interactive pathogen detection tool in metagenomic sequence data. Plants 10:250.

Espindola, A. S., Schneider, W., Cardwell, K. F., Carrillo, Y., Hoyt, P. R., Marek, S. M., Melouk, H. A., and Garzon, C. D. 2018. Inferring the presence of aflatoxin-producing *Aspergillus flavus* strains using RNA sequencing and electronic probes as a transcriptomic screening tool. PLoS One 13:e0198575.

Hajimorad, M. R., Domier, L. L., Tolin, S. A., Whitham, S. A., and Saghai Maroof, M. A. 2018. *Soybean mosaic virus*: A successful potyvirus with a wide distribution but restricted natural host range. Mol. Plant Pathol. 19:1563-1579.

Hill, J. H., and Whitham, S. A. 2014. Control of virus diseases in soybeans. Adv. Virus Res. 90:355-390.

Kitajima, E. W. 2020. An annotated list of plant viruses and viroids described in Brazil (1926-2018). Biota Neotrop. 20:e20190932.

Kutnjak, D., Tamisier, L., Adams, I., Boonham, N., Candresse, T., Chiumenti, M., De Jonghe, K., Kreuze, J. F., Lefebvre, M., Silva, G., Malapi-Wight, M., Margaria, P., Mavrič Pleško, I., McGreig, S., Miozzi, L., Remenant, B., Reynard, J.-S., Rollin, J., Rott, M., Schumpp, O., Massart, S., and Haegeman, A. 2021. A primer on the analysis of high-throughput sequencing data for detection of plant viruses. Microorganisms 9:841.

Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., and Candresse, T. 2018. Application of HTS for routine plant virus diagnostics: State of the art and challenges. Front. Plant Sci. 9:1082.

Mtonga, A., and Maruthi, M. N. 2024. Diseases of common bean. Pages 1-52 in: Handbook of Vegetable and Herb Diseases. W. H. Elmer, M. McGrath, and R. J. McGovern, eds. Springer, Cham, Switzerland.

Ochoa-Corona, F. M., Cardwell, K. F., and Espindola, A. S. 2019. New technologies from the microbial world: Alternatives for biomedical surrogate research. Adv. Biotechnol. Microbiol. 13:555859.

Oliver, R. 2022. Globalizing plant health. Plant Pathol. 71:226-235.

Proaño-Cuenca, F., Espindola, A. S., and Garzon, C. D. 2023. Detection of *Phytophthora*, *Pythium*, *Globisporangium*, *Hyaloperonospora*, and *Plasmopara* species in high-throughput sequencing data by in silico and in vitro analysis using Microbe Finder (MiFi). PhytoFrontiers 3: 124-136.

Rabedeaux, P. F., Gaska, J. M., Kurtzweil, N. C., and Grau, C. R. 2005. Seasonal progression and agronomic impact of *Tobacco streak virus* on soybean in Wisconsin. Plant Dis. 89:391-396.

Rahman, S. U., McCoy, E., Raza, G., Ali, Z., Mansoor, S., and Amin, I. 2023. Improvement of soybean; A way forward transition from genetic engineering to new plant breeding technologies. Mol. Biotechnol. 65:162-180.

Ribeiro-Junior, M. R., Barreto da Silva, F., Marubayashi, J. M., Uzan, J., Nogueira, A. M., Muller, C., Nascimento, D. M., Yuki, V. A., Narita, N., Pavan, M. A., Ochoa-Corona, F. M., and Krause-Sakate, R. 2022. Molecular and biological characterization of an isolate of the potyvirus passiflora virus Y naturally infecting soybean (*Glycine max*) in Brazil. Arch. Virol. 167:2743-2747.

Richter, D. C., Ott, F., Auch, A. F., Schmid, R., and Huson, D. H. 2008. MetaSim—A sequencing simulator for genomics and metagenomics. PLoS One 3:e3373.

Rizzo, T. M., and Palukaitis, P. 1989. Nucleotide sequence and evolutionary relationships of cucumber mosaic virus (CMV) strains: CMV RNA 1. J. Gen. Virol. 70:1-11.

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J., and Schneider, W. 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.

USDA. 2024. Soybeans. USDA Foreign Agricultural Service. https://fas.usda.gov/data/production/commodity/2222000 (accessed 17 December 2024).

Webster, C. G., Frantz, G., Reitz, S. R., Funderburk, J. E., Mellinger, H. C., McAvoy, E., Turechek, W. W., Marshall, S. H., Tantiwanich, Y., McGrath, M. T., Daughtrey, M. L., and Adkins, S. 2015. Emergence of *Groundnut ringspot virus* and *Tomato chlorotic spot virus* in vegetables in Florida and the southeastern United States. Phytopathology 105:388-398.

Wood, D. E., and Salzberg, S. L. 2014. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15:R46.

Zuniga, L. P., Daniels, J., Espindola, A., and Ochoa Corona, F. 2018. EDNA-Water, using deep sequencing and bioinformatics approach for waterborne plant virus detection. International Congress of Plant Pathology (ICPP) 2018: Plant Health in A Global Economy. https://doi.org/10.1094/PHYTO-108-10-S1.240

Zuniga, L. P., Espindola, A., Melouk, H. A., Ali, A., Cardwell, K. F., and Ochoa Corona, F. 2017. Detection of cucurbit viruses in Oklahoma combining EDNA with multiplex RT-PCR coupled with high resolution melting. 2017 APS Annual Meeting. https://doi.org/10.1094/PHYTO-107-12-S5.1