



Research

Development and Statistical Validation of E-Probe Diagnostic Nucleic Acid Analysis (EDNA) Assays for the Detection of Citrus Pathogens from Raw High-Throughput Sequencing Data

Tyler Dang¹ | Huizi Wang² | Andres S. Espindola^{3,4,*} | Josh Habiger² | Georgios Vidalakis¹ | Kitty Cardwell^{3,4}

¹ Department of Microbiology and Plant Pathology, University of California, Riverside, CA 92521

² Department of Statistics, Oklahoma State University, Stillwater, OK 74078

³ Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078

⁴ Institute for Biosecurity and Microbial Forensics (IBMF), Oklahoma State University, Stillwater, OK 74078

* Corresponding author: A. S. Espindola; andres.espindola@okstate.edu

Accepted for publication 7 October 2022.

Author contributions: A.E., G.V., J.H., and K.C. conceived the work; G.V. and T.D. identified samples; T.D. and A.E. generated high-throughput sequencing data; T.D., H.W., and A.E. developed pathogen-specific and internal control e-probes; H.W. and J.H. developed the models and performed the statistics; T.D. and H.W. prepared the manuscript. All authors edited the final manuscript.

Data availability: Data used in this study can be provided upon request.

The findings and conclusions in this publication are those of the author(s) and should not be construed to represent any official USDA, U.S., or state government determination or policy. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the UC or OSU. UC and OSU are equal opportunity providers and employers.

Funding

This research was funded by the Citrus Research Board (CRB) projects 5300-179 and 5300-205 and the NSF Partnership for Innovation grant # G20004052. Additional support was also provided by the CRB project 6100; USDA National Institute of Food and Agriculture, Hatch project 1020106; and the National Clean Plant Network-USDA Animal and Plant Health Inspection Service (AP17PPQS&T00C118, AP18PPQS&T00C107, AP19PPQS&T00C148, & AP20PPQS&T00C049) awarded to G. Vidalakis.

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

Abstract

The cost for high-throughput sequencing (HTS) has decreased significantly and has made it possible for the application of this technology for routine plant diagnostics. There are constraints to the use of HTS as a diagnostic tool, including the need for dedicated personnel with a bioinformatic background for data analysis and the lack of a standardized analysis pipeline that makes evaluating and validating results generated at different HTS laboratories difficult. E-probe diagnostic nucleic acid analysis (EDNA) is an in-silico bioinformatic tool that utilizes short curated electronic probes (e-probes) designed from pathogen-specific sequences that allow users to detect and identify single or multiple pathogens of interest in raw HTS data sets. This platform streamlines the bioinformatic data analysis into a graphical user interface as a plant diagnostic tool used by diagnosticians. In this study, we describe the process for the development, validation, and use of e-probes for detection and identification of a wide range of taxonomically unique citrus pathogens that include citrus exocortis viroid, citrus tristeza virus, '*Candidatus* Liberibacter asiaticus', and *Spiroplasma citri*. We demonstrate the process for evaluating the analytical and diagnostic sensitivity and specificity metrics of the in-silico EDNA assays. In addition, we show the importance of including background noise (internal controls) to generate variance in noninfected samples for a valid statistical test using the quadratic discriminant analysis. The fully validated EDNA assays from this study can be readily integrated into existing citrus testing programs that utilize HTS.

Keywords: assay validation, high-throughput sequence (HTS) diagnostics, limit of detection, MiFi, next-generation sequencing (NGS)

Citrus is an iconic high-value crop in California (estimated at \$3.6 billion with an economic impact of \$7.6 billion) that is under constant threat of endemic and exotic citrus pathogens (Babcock 2022). State and federal phytosanitary regulatory quarantine programs for variety introductions and frequent testing of propagative materials at citrus



nurseries have helped prevent pathogens from establishing and spreading in California (Bostock et al. 2014; Fuchs et al. 2021; Gergerich et al. 2015). Many diagnostic methods have been developed for the detection of citrus pathogens. Traditional methods such as biological indexing have proved reliable for detecting pathogens; however, the method is low throughput, is time consuming, demands highly trained and experienced personnel in the identification of symptoms, and requires large greenhouse space with a range of environmental conditions to maintain different species of plant indicators to detect a wide range of citrus pathogens (Krueger and Vidalakis 2022; Roistacher 1991; Vidalakis et al. 2004). Laboratory-based methods such as the enzyme-linked immunosorbent assay (ELISA), sequential polyacrylamide gel electrophoresis (sPAGE), and polymerase chain reaction (PCR) have evolved over time for citrus diagnostics and now lay the foundation for the transition toward high-throughput sequencing (HTS) workflows (Cambra et al. 2000; Duran-Vila et al. 1993; Osman et al. 2015). Citrus diagnostics technologies have been continuously evolving, and the addition of HTS would be the next major development (Adams et al. 2009; Shendure and Ji 2008).

HTS is a powerful technology that combines molecular biology and computer science. It has been used in various applications such as to evaluate differential gene expression, microbiome community analysis, microRNA sequencing, and de novo discovery and characterization of pathogens (Adams et al. 2009; Villamor et al. 2019). HTS has gained traction as a potential tool for routine plant pathogen diagnostics (Al Rwahnih et al. 2015; Dang et al. 2022; Espindola and Cardwell 2021; Espindola et al. 2015, 2021, 2022; Rott et al. 2017; Soltani et al. 2021; Villamor et al. 2019; Visser et al. 2016). In the case of citrus, the proper implementation of HTS-based diagnostics can streamline laboratory processes and progressively complement or phase out more than 20 individual laboratory tests currently required for the detection of all known graft-transmissible pathogens of citrus. HTS can generate data with enough resolution to discern between different isolates of the same pathogen (Bocsanczy et al. 2019; Villamor et al. 2019). HTS reduces the need for plant indicators for biological indexing that will free up valuable greenhouse space and time to new germplasm release. The declining cost of HTS has made the technology more accessible for laboratories to implement (Adams et al. 2009; Shendure and Ji 2008).

One of the primary difficulties with implementing HTS-based diagnostics is the data analysis. HTS data analysis is time consuming and laborious, and it requires dedicated personnel with high-level knowledge in bioinformatics and computer programming, as well as access to expensive high-performance computing. Cutoff values for diagnostic calls using a traditional bioinformatic workflow (i.e., aligning, assembling, and BLASTn reads) can vary from laboratory to laboratory and in some cases can be arbitrary (Massart et al. 2019; Rott et al. 2017). Online platforms such as Virfind provide an accessible bioinformatic pipeline that can be used for virus pathogen discovery and detection (Ho and Tzanetakis 2014). However, such analysis can be overcomplicated because of the different parameters that the user needs to define for the statistical cutoff values, as well as the excess information provided by the software. The results include the global database queries that require postprocessing by the user to discern the relevant pathogens of regulatory importance (Villamor et al. 2019).

To overcome the challenges with HTS data analysis, e-probe diagnostic nucleic acid analysis (EDNA) was developed by the Oklahoma State University (OSU) Institute for Biosecurity and Microbial Forensic in collaboration with the United States Department of Agriculture-Agricultural Research Service (USDA-

ARS) (Stobbe et al. 2013). MiFi, a graphical user interface that contains EDNA, was later created to be a user-friendly online HTS data analysis tool for diagnostic applications of specific targeted pathogens (Espindola and Cardwell 2021). The system is not designed for the de novo discovery of all pathogens and pathogen-like sequences in an HTS data set but to make HTS analysis accessible to laboratory diagnosticians.

EDNA is a bioinformatic tool that utilizes short curated electronic probes (e-probes) designed from pathogen-specific sequences. The e-probes allow users to detect and identify single or multiple pathogens of interest from raw HTS data sets and ignore irrelevant sequences such as the host or other nontargeted microbes present in the sample, similar to pathogen-specific primers used in a PCR reaction (Espindola and Cardwell 2021; Stobbe et al. 2013, 2014). However, where PCR relies on a single known oligo for pathogen detection, EDNA probes include many unique oligos from along the length of the pathogen genome. EDNA can be utilized on raw HTS data generated from different sequencing platforms such as Illumina (San Diego, CA) and MinIon (Oxford Nanopore, Oxford, U.K.). This technology has been previously used for the detection of foodborne pathogens such as *Escherichia coli* O157: H7 (Blagden et al. 2016), plant pathogens such as oomycetes (*Phytophthora ramorum* and *Pythium ultimum*), fungi (*Phakopsora pachyrhizi* and *Puccinia graminis*) (Espindola et al. 2015, 2022), and viruses (plum pox virus) (Stobbe et al. 2014).

In this study, we evaluated the EDNA technology as a routine diagnostic tool for the detection of viroid, virus, and bacterial graft-transmissible pathogens of citrus. E-probes were developed and validated for the detection of citrus exocortis viroid (CEVd), citrus tristeza virus (CTV), '*Candidatus Liberibacter asiaticus*' (CLas), and *Spiroplasma citri* (*S. citri*). For the EDNA technology to be adopted for mainstream citrus diagnostics, the performance of each set of e-probes has to be evaluated for sensitivity and specificity and compared with results from existing molecular diagnostic technologies.

MATERIALS AND METHODS

Plant material and sample collection

Infected citrus plants were collected from screenhouse and greenhouse sources of the Citrus Clonal Protection Program (CCPP) Disease Bank at the Rubidoux Quarantine Facility of the University of California (UC), Riverside, and the USDA-ARS U.S. Horticultural Research Laboratory in Fort Pierce, Florida. Additional infected plant samples were collected from field sources by the Central California Tristeza Eradication Agency (CCTEA) in Tulare, California, and intercepted as citrus variety introductions from China, Pakistan, and Puerto Rico by the CCPP (USDA permits PCIP-14-00356, PCIP-16-00029, and PCIP-17-00613) (Supplementary Table S1). All noninfected plant material was collected from screenhouse sources of the CCPP Lindcove Foundation Facility at the UC Agriculture and Natural Resources (ANR) Lindcove Research and Extension Center (LREC) (Supplementary Table S2).

Stem samples (i.e., shoots with leaves and thorns removed) were collected from the last mature vegetative flush (approximately 12 to 18 months old) and around the tree canopy to account for any unequal distribution of the pathogen in the plant. To avoid cross-contamination, pruners were sanitized with 10% household bleach solution (0.5% sodium hypochlorite) and dried with a paper towel between sampling of each tree. All samples were packaged into separate resealable bags, placed in an ice chest, transported to the CCPP, and immediately stored at 4°C until further processing.

Sample preparation

For stem samples, a clean razor blade was used to peel and separate the phloem-rich bark tissues from the stems. The peeled bark was finely chopped into 0.5-mm pieces, transferred to a separate mortar filled with liquid nitrogen, and pulverized with a pestle into a fine powder. One hundred milligrams of the powdered sample was transferred to a 2.0-ml microcentrifuge tube (Eppendorf, Hamburg, Germany) using a disposable spatula and immediately transferred on dry ice until all samples were processed.

RNA and DNA extraction

RNA was extracted from the pulverized bark tissues using TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's recommended protocol adjusted for citrus stem tissues previously described in Dang et al. (2022). The final RNA pellet was resuspended in 100 μ l of UltraPure DNase/RNase-free distilled water (Thermo Fisher Scientific). The concentration, quality, and purity of the extracted RNA was evaluated with a spectrophotometer and bioanalyzer (Agilent, Santa Clara, CA). The eluted RNA was aliquoted into three 1.5-ml DNA LoBind microcentrifuge tubes (Eppendorf) and stored at -80°C until further use.

DNA extraction was performed using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA) with the manufacturer's recommended protocol. All samples were mixed by flicking the tube in lieu of a vortex to prevent DNA from shearing. The final DNeasy Mini spin column was transferred to a 1.5-ml microcentrifuge tube, and 100 μ l of Buffer AE was added and incubated at 5 min at room temperature. The samples were centrifuged for 1 min at $6,000 \times g$ at room temperature. The concentration and purity of the extracted DNA was determined with a spectrophotometer. The DNA was subsequently aliquoted into three 1.5-ml DNA LoBind microcentrifuge tubes and stored at -20°C until further use.

HTS library preparation

The RNA-seq library was prepared for the detection of viral and viroid citrus pathogens. Plant ribosomal RNA depletion was performed with the Illumina Ribo-Zero rRNA Removal Kit for Plants, followed by library preparation using the Illumina TruSeq Stranded Total RNA kit as previously described in Dang et al. (2022).

The DNA-seq were enriched using the NEBNext Microbiome DNA Enrichment Kit (New England Biolabs, Ipswich, MA). All treatments were performed using the manufacturer's recommended protocol and the ProFlex thermal cycler (Thermo Fisher Scientific). AMPure XP Beads (Beckman Coulter, Brea, CA) were used for all cleaning and size selection steps. The DNA-seq library was prepared using NEBNext Ultra II FS DNA Library Prep Kit (New England Biolabs). All library preparation and sample pooling was performed based on the manufacturer's recommended protocol. The quality of the final pooled RNA-seq and DNA-seq libraries was evaluated with a spectrophotometer, qPCR, and bioanalyzer. The libraries were sequenced at the UC Riverside IIGB Core Facility on the Illumina NextSeq500 platform with paired-end reads (2×75 bp).

HTS data processing and analysis with a standard bioinformatic pipeline

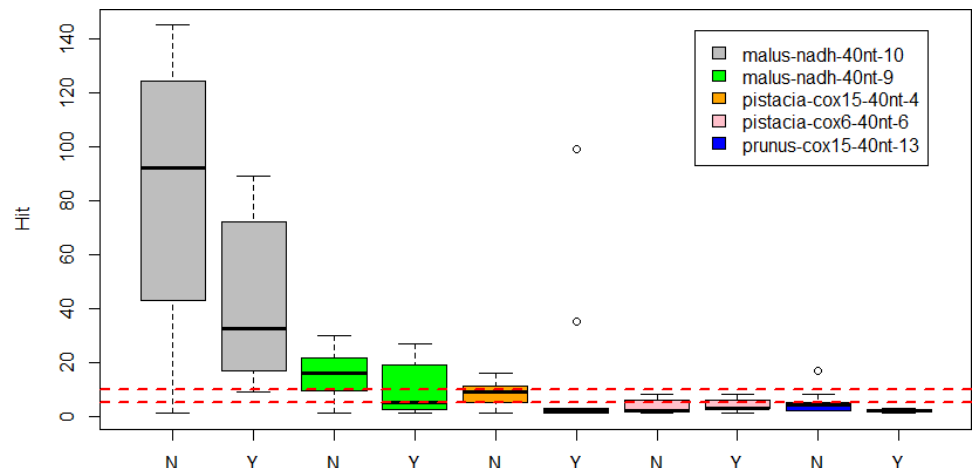
Low-quality Illumina reads were filtered with Fastqc (Andrews 2010) using the default settings. After quality control, citrus host reads were removed using Bowtie2 version 2.3.4.1 (Langmead et al. 2009) by mapping the reads to the reference citrus genome (*Citrus sinensis*, GCA_000317415). Reads that mapped to the citrus reference genome were discarded, and the unmapped reads were assembled de novo using Trinity version 2.8.5 (Grabherr et al. 2011). BLASTn (Altschul et al. 1990; Madden et al. 1996) was performed to confirm the identity of the assembled contigs. A custom Bash shell and R script was used to filter out BLASTn results that were unique only to citrus pathogens.

Verification of the presence of citrus pathogens with PCR-based methods

PCR was performed for the validation of HTS and EDNA results. The primers and probes used in this study are listed in Supplementary Table S3. All multiplex reverse transcription (RT) quantitative PCRs (qPCR) for the detection of CTV and citrus viroids (namely "Nonapsca", which include pospi-, hostu-, and cocad-viroids) were performed in 12- μ l reactions using the QuantiFast Multiplex RT-PCR kit (Qiagen), 0.045 μ l of nuclease-free water, 6.25 μ l of $2\times$ QuantiFast RT Master Mix, 0.58 μ l of primer and probe mix, 0.125 μ l of QuantiFast RT mix, and 5 μ l of total RNA (Osman et al. 2015, 2017).

FIGURE 1

Box plot of the five internal controls shows the distribution of hits for the internal controls across the noninfected (N) and infected samples (Y).



qPCR for detection of CLAs was performed in 20 µl consisting of 7.7 µl of nuclease-free water, 0.5 µl of forward and reverse primer, 10 µl of iQ Universal Probes Supermix (Bio-Rad, Hercules, CA), 0.3 µl of probe, and 1 µl of template (Li et al. 2007). The samples were loaded into a CFX96 Real-Time PCR machine (Bio-Rad) with the following conditions: 95°C for 5 min, 95°C for 10 s, and 58°C for 30 s for 40 cycles. qPCR for the detection of *S. citri* was performed in 25-µl reactions consisting of 12.5 µl of iQ SYBR Green Supermix (Bio-Rad), 0.5 µl of forward and reverse primer, 11 µl of nuclease-free water, and 1 µl of template (Wang et al. 2015). The *S. citri* qPCR was performed in the CFX96 Real-Time PCR machine with the following conditions: 95°C for 3 min, 95°C for 10 s, and 59°C for 30 s for 35 cycles. Amplicons were visualized on 2% agarose gel stained with ethidium bromide under UV light. All PCRs performed had the appropriate positive and negative controls in ordinance with the MIQE guidelines (Bustin et al. 2009).

Target citrus pathogen e-probe design and curation for specificity

E-probes, ranging from 20 to 40 nt, were designed to target CEVd, CTV, CLAs, and *S. citri*. The pathogen target and the taxonomically near neighbors or other citrus pathogen genomes were retrieved from the NCBI GenBank database (Supplementary Table S4). Pathogen-specific sequences were identified using the EDNA built-in sequence alignment program MUMmer by comparing the target against a file of concatenated near neighbor sequences (Delcher et al. 2003). All similar or homologous sequences were removed, and sequences unique only to the target pathogen were selected as e-probes. The next step during the e-probe design phase is curation. The in-silico probes go through a bioinformatic pipeline to remove any probes that could bind to targets other than the pathogen of interest. The raw e-probes were curated by uploading to NCBI for BLASTn analysis. Any sequence matches not from the target pathogen (e-value of 1×10^{-10} or lower) were removed from the final e-probe set to ensure diagnostic specificity (Espindola et al. 2022; Stobbe et al. 2013, 2014).

Decoy e-probes were generated from the curated e-probes by using the reverse sequence of the target e-probes. The analysis was based on a *t* test comparing the scores of the target and decoy e-probes. No significant differences between the target and decoy scores indicates the absence of the targeted pathogen ($P > 0.05$), whereas a significant difference signals the presence of the targeted pathogen ($P < 0.05$) (Stobbe et al. 2013).

Evaluation of EDNA probes using simulated HTS data for determining the analytical sensitivity

MetaSim was used to generate simulated HTS data from the citrus host (*Citrus sinensis*, GCA_000317415) and the targeted pathogen (e.g., CTV and CEVd) reference genomes for e-probe analytical sensitivity validation (Richter et al. 2008). The simulated HTS data were generated with approximately 10 million reads per sample and spiked with a percentage of pathogen sequence reads that ranged from 0.01 to 0.0001%. Simulated HTS data with no pathogen reads were also generated. The “empirical” option was utilized to generate simulated data consistent with Illumina sequencing lengths and error rates.

EDNA data analysis for diagnostic sensitivity and limit of detection

The diagnostic sensitivity and limit of detection (LoD) classifications were generated based on known PCR-confirmed positive

and negative sample data. For EDNA diagnostic sensitivity analysis, the compressed raw HTS fastq files of the known positive and negative sample metagenome sequences were evaluated for false positive and negative results by uploading into the Microbe Finder (MiFi) software platform (<https://bioinfo.okstate.edu/>).

MiFi generates a score for each e-probe in a sample. The score is the sum of all high-quality matches in the metagenome for each target e-probe. An index called Total Score (TS), which sums scores over e-probes (and scaled and presented by dividing by 10,000), was used to train a classifier algorithm for determining whether a pathogen nucleic acid is present or absent from the sample. In other words, a large TS value indicates a high likelihood that a pathogen exists in the sample. The classifier is based on quadratic discriminant analysis (QDA), which facilitates a simple estimate of LoD accounting for unequal variances between the TS of positive and negative samples. The LoD is the TS value at which the probability of a MiFi match to the PCR-positive test is equal to the probability of a MiFi match to a PCR-negative test. This corresponds to the reliability of the LoD of the assay at which point the chance of a false negative is 50/50. The LoDs of four genera of pathogens, a virus, two bacteria, and a viroid, were calculated.

Statistical methods

The QDA assumes that TSs are generated from a normal distribution with mean denoted as μ_1 and variance denoted as σ_1^2 when a pathogen is present and mean denoted as μ_0 and variance denoted as σ_0^2 if a pathogen is not present in $N = 10$ known positive and 10 known negative samples. To compute a posterior probability of pathogen presence, assume a prior probability that $Pr(\text{Pathogen}) = Pr(\text{Non Pathogen}) = \frac{1}{2}$. Then, the standard Bayes classifier (Hastie et al. 2001) computes the posterior probability of a pathogen given $TS = z$ as follows:

$$Pr(\text{Pathogen} | TS = z) = \frac{\frac{1}{\sigma_1} \exp\left(-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sigma_0} \exp\left(-\frac{(z-\mu_0)^2}{2\sigma_0^2}\right) + \frac{1}{\sigma_1} \exp\left(-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right)} \quad (1)$$

Setting equation 1 equal to 0.5 and solving for z gives

$$\text{LoD} = \frac{\frac{\mu_1 - \mu_0}{\sigma_1^2} - \frac{(\mu_1 - \mu_0)^2}{2\sigma_1^2\sigma_0^2} - \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}}{\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}} \times 2\log\frac{\sigma_1}{\sigma_0} \quad (2)$$

TABLE 1

E-probes designed for citrus exocortis viroid (CEVd), citrus tristeza virus (CTV), ‘*Candidatus Liberibacter asiaticus*’ (CLAs), and *Spiroplasma citri* and the limit of detection (LoD) for each probe

Target pathogen	Genome size (bp)	E-probe length (nt)	Number of e-probes generated	Number of internal control e-probes	LoD ^z
CEVd	~372	20	12	5	475.3
		30	6		138.5
CTV	~20,000	20	1,146	5	105.0
		30	818		116.7
CLAs	~1,200,000	40	9,004	5	424.2
<i>S. citri</i>	~1,600,000	40	1,076	1	31.5

^z The LoD is defined as the lowest reliable total score for a diagnostic positive result.

The estimate of LoD is $\hat{L}oD$ and found by inputting known PCR-positive and -negative sample TS means and variances for parameters in equation 2. Here, $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ are the sample mean and variance of healthy samples, respectively, and $\hat{\mu}_1$ and $\hat{\sigma}_1^2$ are the sample mean and variance of diseased samples. The R code for generating the $\hat{L}oD$ can be found at <https://github.com/microbefinder/LoD-Calculator>. The tool also includes essential plots for verifying that the TSs are roughly normally distributed with unequal variances and means.

The performance of the LoD is assessed with the misclassification rate and the area under (AU) the receiver operating characteristic (ROC) curve (Gareth et al. 2013). For any given sample, the possible diagnostic result could be classified with the LoD into true negative (TN), false negative (FN), false positive (FP), and true positive (TP). The number of healthy samples is denoted by $N = TN + FP$ and diseased samples by $P = FN + TP$. Then, the misclassification rate (MR) for an LoD is defined as $MR = (FN + FP)/(N + P)$. We also computed the AUROC curve for each data set. The ROC curve plots (FP/N) versus (TP/P). These are also called the 1 - Specificity and Sensitivity, respectively. Ideally, the AUROC is 1, which indicates that the LoD was greater than all healthy TS values and less than all diseased TS values.

Development and curation of internal controls

Due to the highly curated nature of e-probe sets, little to no variance was identified in noninfected control samples because all e-probes with possible spurious FP matches with the nontarget were removed. Doing so results in very specific e-probe sets. Zero variance in the negative control sample renders both the t test and the QDA inaccurate. Therefore, e-probe sets to serve as internal

controls were developed to generate the background necessary to generate a proper P value from the t test (one-tailed) between the target and decoy scores. The e-probes were generated targeting plant housekeeping genes such as nicotinamide adenine dinucleotide hydrogen (NADH) dehydrogenase, cytochrome oxidase factor (COX) 6, and COX 15 from noncitrus hosts such as *Pistacia vera*, *Prunus persica*, and *Malus domestica* (Supplementary Table S5). E-probes were designed manually or by using the MiProbe function within the MiFi platform. The internal control e-probes were subsequently added to the pathogen-specific e-probes. The internal control e-probes were curated to remove nonspecific e-probes by testing against 10 noninfected and 10 infected tree samples with the specific target pathogen. Internal control e-probes that had approximately equal but low scores in both the noninfected and infected samples were retained. The average of the total score was calculated, and the five lowest scored e-probes were retained (Fig. 1).

RESULTS

E-probe development

A total of 12 CTV, 6 CEVd, 10 CLas, and 15 *S. citri* genome sequences from different isolates were used to generate the pathogen-targeted e-probes (Table 1; Supplementary Table S4). E-probes with lengths of 20 and 30 nt were designed for CEVd and CTV, and e-probes with lengths of 40 nt were designed for CLas and *S. citri*. The e-probes were curated by comparing genome sequences against other citrus pathogens, and BLASTn analysis was performed to identify and retain the pathogen-specific e-probes of interest. Pathogen genome size directly cor-

TABLE 2

Evaluation of analytical sensitivity of citrus tristeza virus (CTV)-specific e-probes combined with internal control e-probes using in silico simulated Illumina data sets

Simulated prevalence of CTV	CTV with 30-nt e-probes and internal control e-probes			CTV with 20-nt e-probes and internal control e-probes		
	Total score	P value	Diagnostic results	Total score	P value	Diagnostic results
784	2,611.3	3.31E-206	Positive	2,454.4	3.52E-165	Positive
786	2,564.4	5.06E-197	Positive	2,359.8	2.21E-155	Positive
97	525.1	7.13E-69	Positive	441.3	4.53E-46	Positive
90	442.7	2.22E-62	Positive	426.7	3.68E-58	Positive
4	29.6	0.241	Negative	69.4	2.06E-03	Positive
9	72.1	4.32E-04	Positive	72.3	2.72E-04	Positive
8	70.9	1.65E-03	Positive	80.1	1.15E-06	Positive
0	25.0	0.820	Negative	49.0	0.383	Negative

TABLE 3

Evaluation of analytical sensitivity of citrus exocortis viroid (CEVd)-specific e-probes combined with internal control e-probes using in silico simulated Illumina data sets

Simulated prevalence of CEVd	CEVd with 30-nt e-probes and internal control e-probes			CEVd with 20-nt e-probes and internal control e-probes		
	Total score	P value	Diagnostic results	Total score	P value	Diagnostic results
1,183	46.3	1.74E-07	Positive	46.3	6.97E-09	Positive
1,167	45.4	2.08E-07	Positive	45.4	9.74E-08	Positive
122	47.3	1.61E-08	Positive	47.3	4.44E-09	Positive
129	41.9	8.40E-06	Positive	41.9	6.98E-09	Positive
10	43.4	8.00E-06	Positive	43.4	9.40E-07	Positive
10	40.8	1.10E-08	Positive	40.8	7.42E-07	Positive
9	45.1	1.05E-07	Positive	45.1	7.42E-07	Positive
0	28.5	0.024	Positive	11.0	0.063	Negative

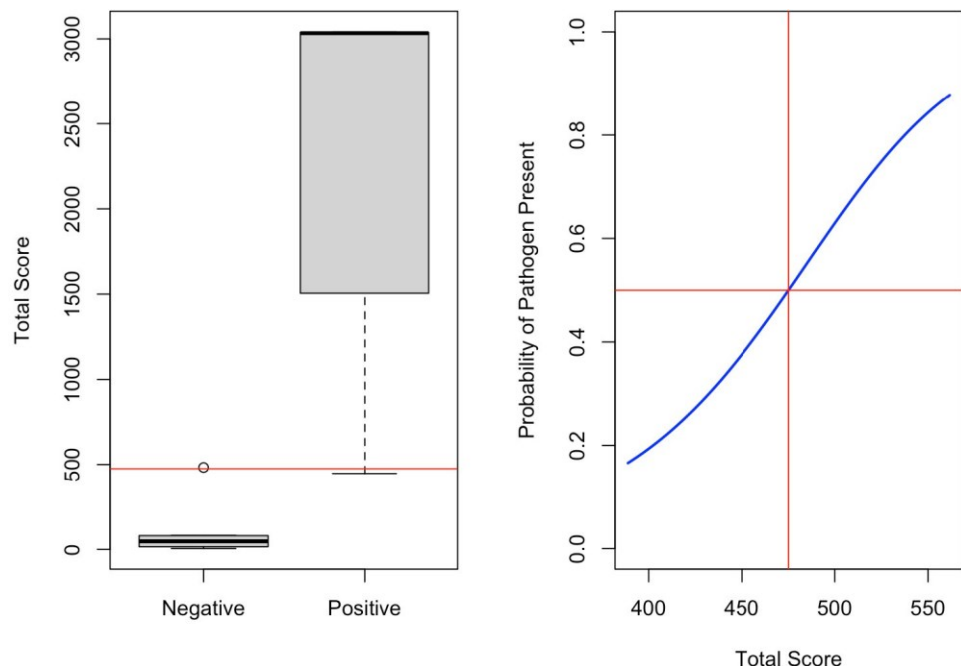
related with the total number of e-probes generated. Bacterial pathogens (CLas and *S. citri*) with larger genomes generated more e-probes than smaller pathogens (CEVd) (Table 1). In addition, the target pathogen e-probe length directly influenced the total number of e-probes generated. Shorter, 20-nt-long e-probes targeting CEVd and CTV generated more e-probes than the longer 30-nt versions (Table 1).

Verification of EDNA with PCR and HTS with infected plant sample HTS data

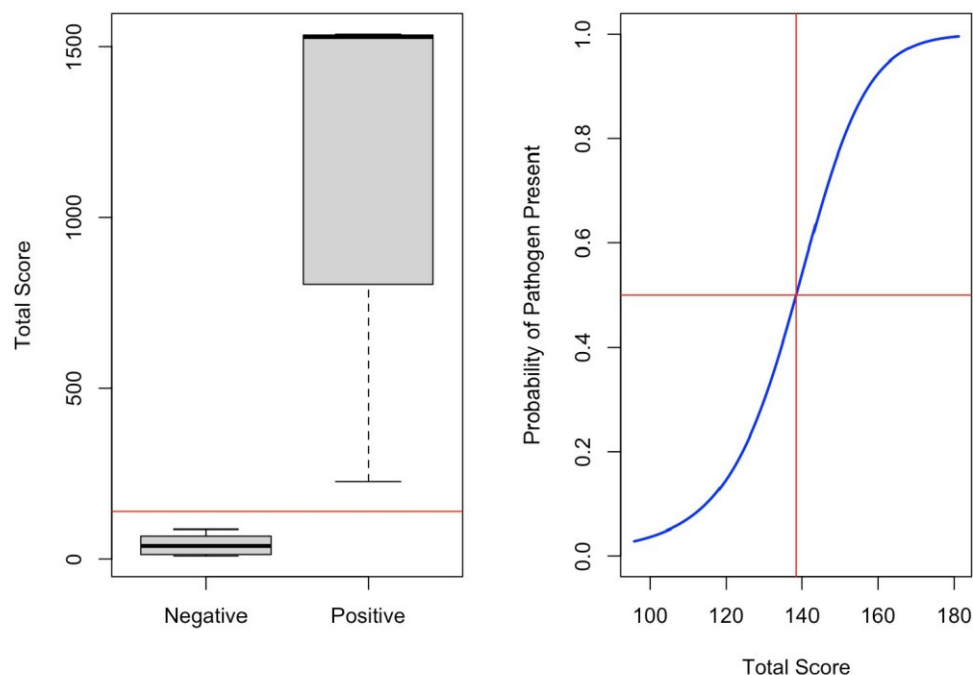
All samples were confirmed for the presences of the targeted pathogen (CEVd, CTV, CLas, and *S. citri*) using PCR-based methods. Results between the traditional HTS analysis, EDNA, and RT-qPCR were the same between the different methods,

FIGURE 2

Box plot and pathogen-probability curves versus total score (TS) for e-probes (lengths 20, 30, or 40 nt) designed for different graft transmissible pathogens of citrus. **A**, Citrus exocortis viroid (CEVd) with 20-nt e-probes, **B**, CEVd with 30-nt e-probes, **C**, citrus tristeza virus (CTV) with 20-nt e-probes, **D**, CTV with 30-nt e-probes, **E**, '*Candidatus* Liberibacter asiaticus' (CLas) with 40-nt e-probes, and **F**, *Spiroplasma citri* with 40-nt e-probes. The limit of detection (LoD), misclassification rate (MR), and area under (AU) the receiver operating characteristic (ROC) curve were calculated for each target pathogen e-probe. For the box plot, the x axis is the infected (positive) and noninfected (negative) samples, and the y axis is the TS. The box plot shows differences in TS between infected and noninfected samples. For the pathogen-probability curves, the intersection of the red line represents the TS required for the probability of infection at 0.5.



A CEVd-20 nt LoD=475.3, MR=0.1 and AUROC=0.9.



B CEVd-30 nt LoD=138.5, MR=0 and AUROC=1.

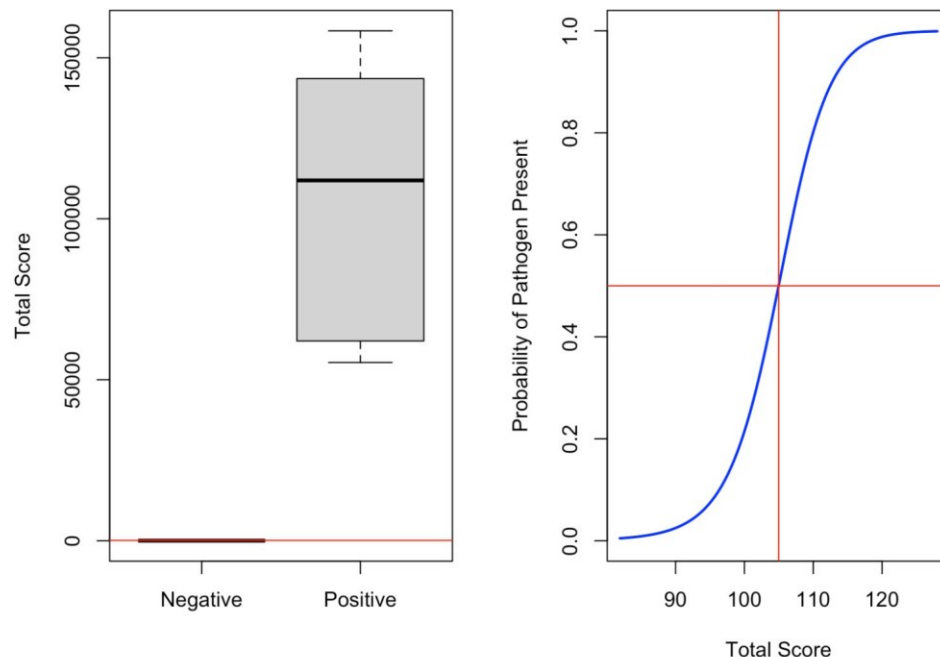
(Continued)

indicating that EDNA is reliable and repeatable (Supplementary Table S1). Healthy controls, from different varieties, were also tested between the different diagnostic methods. EDNA for the healthy controls was negative, indicating that pathogen e-probes were not cross-reacting with the citrus host genome sequences and that the designed e-probes are highly specific to the pathogen (Supplementary Table S2).

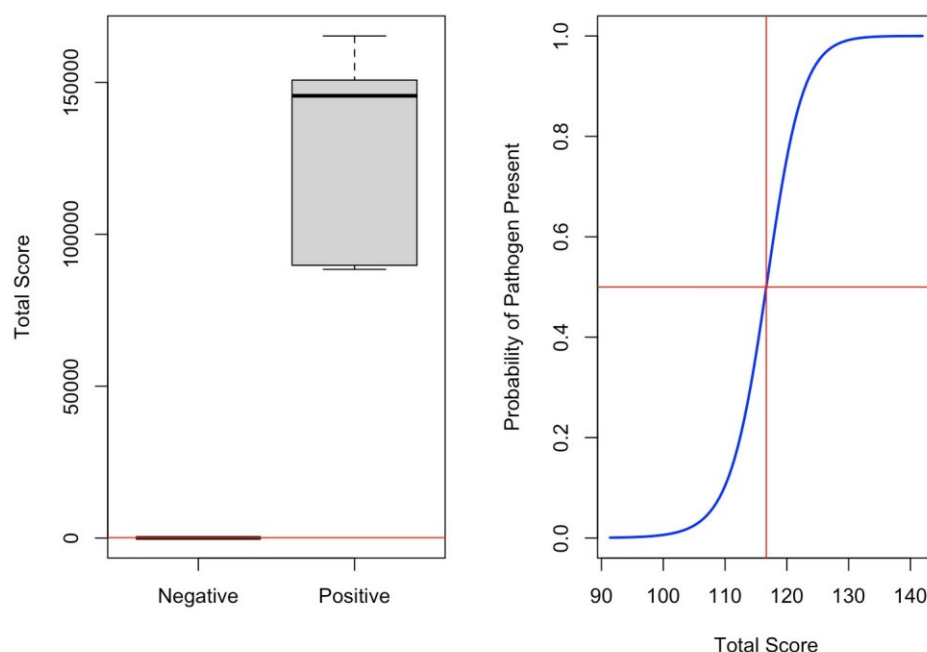
Determination of analytical sensitivity of pathogen e-probes with simulated HTS data

Simulated data from the pathogen and citrus host were used to test the analytical sensitivity and theoretical LoD for the e-probes designed for selected pathogens CTV and CEVd (Tables 2 and 3). Both the 20- and 30-nt e-probes for CTV showed decreasing

FIGURE 2
(Continued from previous page)



C CTV-20 nt LoD=105.0, MR= 0 and AUROC=1.



D CTV-30 nt LoD=116.7, MR= 0 and AUROC=1.

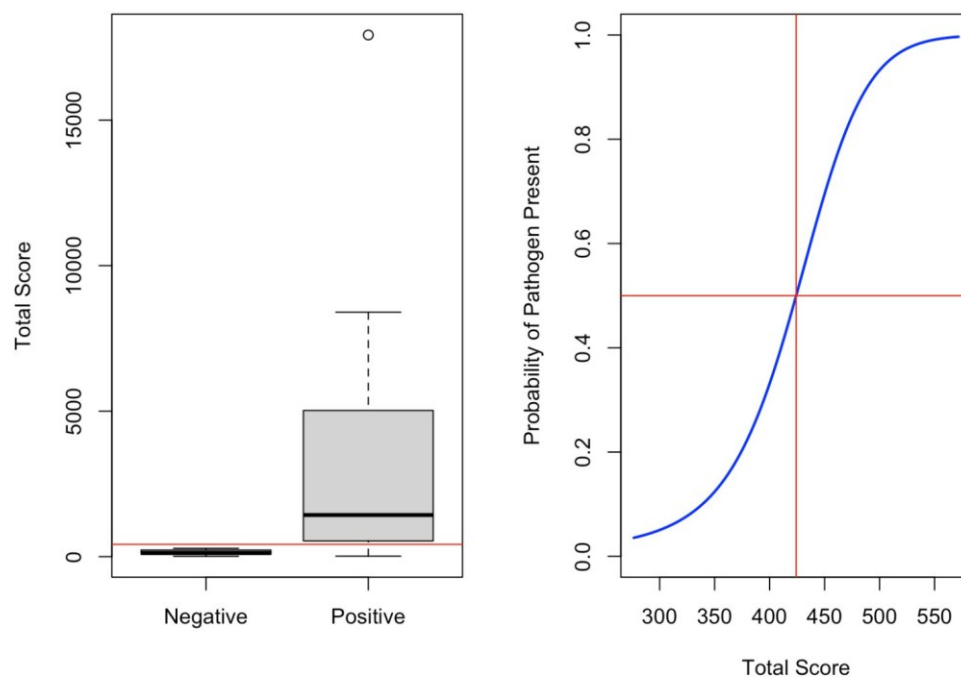
(Continued)

total scores and an increase in P value as the prevalence of CTV reads were reduced. The pattern was not present in the CEVd e-probes. This is due to the signal from the internal control e-probes oversaturating the CEVd e-probe signals because of the high-quality hits that resulted in FP diagnostic results with CEVd 30-nt e-probes from simulated HTS data without the targeted pathogen (Table 3).

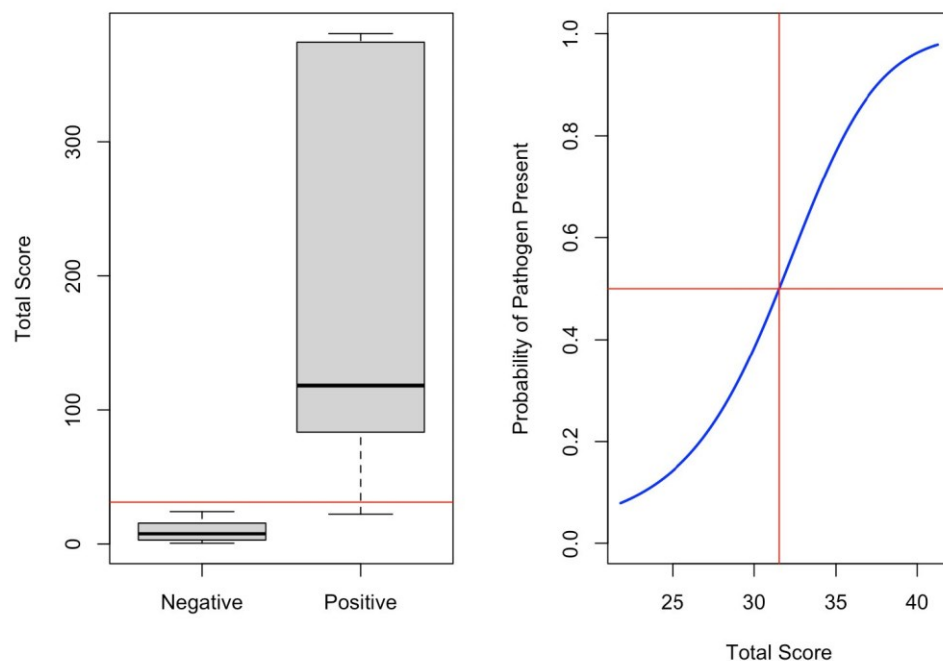
Generation of LoDs, diagnostic sensitivity, and analysis of LoDs

A total of six LoDs (four graft-transmissible pathogens of citrus and two different lengths for CEVd and CTV) were generated by equation 2 using the data sets presented in Supplementary Tables S6 to S10 and S12. As previously stated, the interpretation

FIGURE 2
(Continued from previous page)



E *Clas* LoD=424.2, MR=0.1 and AUROC=0.9.



F *S. citri* LoD=31.5, MR=0.1 and AUROC=0.9.

of LoDs is that a sample has a high likelihood of confirming a target disease if the observed total score (TS) is greater than the LoD for that target pathogen. For example, the LoD for the CLAs data set was 424.2, and the observed TS of IPPN-743 for CLAs was 3,131.4 (Supplementary Table S10). Thus, the IPPN-743 sample was positive for CLAs because the sample TS was much higher than the LoD for CLAs. We can be confident in our prediction because the probability of pathogen presence when TS is 3,131.4 is in fact much greater than 0.5; it is nearly 1. In fact, any TS greater than 550 gives a pathogen probability of nearly 1 (Fig. 2). The LoDs and corresponding pathogen-probability curves versus TS were generated for the remaining target pathogens (Table 1 and Fig. 2A to F).

The misclassification rate (MR) for each data set was computed and is listed in Figures 2A to F and 3A to F. For example, the MR is 0 in the CTV data with an e-probe length of 20 nt. Hence, all the predictions for this classifier were correct (Fig. 2C). All diseased samples had a TS greater than the LoD (depicted by the red line), and all healthy samples had a TS less than the LoD. Hence, there were no FPs and no FNs, and $MR = 0$. However, some classifiers did not perform as well. For example, some FPs and FNs occurred for the CLAs data because some TS values from negative samples were greater than the LoD and vice versa. Here, $MR = 0.1$ (Fig. 2E).

Recall that an ideal AUROC value is 1 (Gareth et al. 2013). In short, this means that the threshold depicted by the red line in Figure 2A to F could be moved and the classifier would still have $FP/N = 0$, $TP/P = 1$, and $MR = 0$. Observe that all of the AUROCs are at least 0.9, and many are 1 (Fig. 2A to F).

Curation and the removal of FP internal control e-probes

The five curated internal control e-probes were able to generate the necessary background noise required for the EDNA *t* test for most of the target pathogens (Tables 2 and 3; Supplementary Tables S6 to S11). For the initial predicted LoD for *S. citri*, the *t* test results of the EDNA resulted in three FP samples (Supplementary Table S11). As mentioned, the internal controls generated excessive high-quality hits in the healthy samples, which increased the total score in healthy samples, and in some positive samples, internal controls gave few high-quality hits. These effects combined to yield FPs. To mitigate the FP results, the five internal control e-probes were further curated to select for internal controls that contribute equally and minimally to both positive and negative samples (Fig. 1). Based on the results, the internal control pistacia-cox6-40nt-6 indeed contributes equally and minimally across groups, and others do not (Fig. 1). The LoD was recalculated using only the single internal control that resulted in no FPs (Supplementary Table S12).

DISCUSSION

HTS has become a popular molecular tool and has expanded to other applications, such as plant diagnostics. As with any pathogen detection technology, there are limitations to HTS and EDNA, as it is not a “one-size-fits-all” method. For example, HTS is extremely sensitive and can be prone to cross-contamination (Rong et al. 2023). This places a greater importance on the need to implement stricter laboratory sanitation and sample-handling protocols compared with protocols for handling and preparing samples for PCR-based diagnostics (Borst et al. 2004). In addition, HTS is still a developing technology. As a result, there is still a need to validate HTS results with complementary diagnostic tools such as PCR and biological indexing to resolve any questionable results that might arise.

In this study, we showed that EDNA can be used for the detection of five graft-transmissible pathogens of citrus from raw HTS data. We evaluated the sensitivity and specificity of EDNA. The EDNA platform performed well compared with the validated and widely used diagnostics assays, such as PCR (Supplementary Tables S1 and S2). With regard to evaluating the sensitivity of EDNA, a QDA was performed to estimate an LoD value, which sets a baseline score required for positive diagnostic results for all targeted pathogens (Table 1). In addition, we found that non-infected controls did not produce the necessary variance required for proper statistics within EDNA. To address this shortcoming,

A CEVd-20 nt FN = 1, FP = 1, N = 10, P = 10, MR = 0.1

		Classification		
		N	P	
Pathogen	N	9	1	10
	P	1	9	10

B CEVd-30 nt FN = 0, FP = 0, N = 10, P = 8, MR = 0

		Classification		
		N	P	
Pathogen	N	10	0	10
	P	0	8	8

C CTV-20 nt FN = 0, FP = 0, N = 10, P = 6, MR = 0

		Classification		
		N	P	
Pathogen	N	10	0	10
	P	0	6	6

D CTV-30 nt FN = 0, FP = 0, N = 9, P = 6, MR = 0

		Classification		
		N	P	
Pathogen	N	9	0	9
	P	0	6	6

E CLAs FN = 3, FP = 0, N = 11, P = 12, MR = 0.1

		Classification		
		N	P	
Pathogen	N	11	0	11
	P	3	9	12

F *S. citri* FN = 2, FP = 2, N = 10, P = 10, MR = 0.1

		Classification		
		N	P	
Pathogen	N	10	0	10
	P	2	8	10

FIGURE 3

Summary of the classification procedure. The misclassification rate (MR) is computed by $(FN + FP)/(N + P)$ for **A**, citrus exocortis viroid (CEVd) with 20-nt e-probes, **B**, CEVd with 30-nt e-probes, **C**, citrus tristeza virus (CTV) with 20-nt e-probes, **D**, CTV with 30-nt e-probes, **E**, ‘*Candidatus Liberibacter asiaticus*’ (CLAs) with 40-nt e-probes, and **F**, *Spiroplasma citri* with 40-nt e-probes.

we supplement the pathogen-specific e-probes with internal control e-probes designed from conserved housekeeping genes from noncitrus woody hosts (Supplementary Table S5). The number of internal control e-probes required will depend on the biology and genome complexity of the pathogen of interest (Fig. 2 and Supplementary Tables S11 and S12).

EDNA is a dynamic technology that allows for the continued expansion and ongoing validation of the citrus pathogen e-probe libraries as new pathogens are discovered or new isolates and variants of the known ones are characterized and sequenced. EDNA allows for the design of e-probes via a crowdsourced model where experts in their respective fields design e-probes based on the sequence data available not only in public databases but in data available only to their laboratories. EDNA simplifies and streamlines the HTS data analysis process without the need for dedicated highly trained bioinformatic personnel to analyze the millions of data points, resulting in some cases in arbitrary and noncomparable results for the same samples among different laboratories (Rott et al. 2017). The EDNA platform standardizes HTS data analysis and results interpretation. It can be a starting point for quality assurance and accreditation for laboratories using HTS as a diagnostic tool to make pathogen detection results comparable and reliable utilizing the standard approach of proficiency tests (i.e., ring tests) among different EDNA laboratories (Soltani et al. 2021). Finally, EDNA is not limited to Illumina but can be used with other sequencing platforms such as the Oxford Nanopore MinION (Liefting et al. 2021; Phannareth et al. 2021). This level of integration offers greater flexibility for laboratories and does not require dependence on sequencing facilities or one specific sequencing technology. The EDNA platform is packaged with a graphical user interface (GUI) where the user can upload their compressed raw fastq files into MiFi and select the e-probes for the pathogens of interest to be scanned in the data (Dang et al. 2022). Each e-probe set contains metadata regarding the developers and the validation metrics of e-probes.

However, before EDNA becomes mainstream and widely adopted by citrus diagnostic laboratories and quarantine or survey programs, it will require regulatory approval. The regulatory approval process will require more comprehensive studies that will demonstrate that EDNA can work equally as well as or better than the current regulatory-approved diagnostic assays for each of the targeted pathogens. Based on the experience of the current study, a key element for the successful implementation and regulatory approval of the HTS and EDNA technologies is the development of sophisticated statistical models (e.g., QDA) that will constantly fine-tune the EDNA protocols on their capacity to decipher the quantitative results of the multiple e-probe detection hits as they scout the millions of HTS sequence data points for their targeted pathogens.

EDNA-based diagnostics is a new technology that will undergo great scrutiny and evaluation from multiple laboratories and scientists in the near future. However, once it is proven to perform accurately and reliably and is adopted by the regulatory agencies, the EDNA technology has the potential to be incorporated seamlessly into existing citrus-testing laboratories and programs (e.g., germplasm and quarantine introductory programs) because of the intuitive MiFi GUI. This technology can help transform the operations of citrus programs because diagnostics is one of their most important functional pillars (Fuchs et al. 2021; Gergerich et al. 2015).

ACKNOWLEDGMENTS

The authors acknowledge the Cahuilla people as the Traditional Custodians of the Land on which the experimental work was completed

at UC Riverside. The authors thank Dr. Subhas Hajeri (Central California Tristeza Eradication Agency) and Dr. Greg McCollum (USDA-ARS Fort Pierce) for providing positive control samples for *S. citri* and CLas. We are grateful to all past and current CCPP personnel for their dedicated work and especially for creating and maintaining the CCPP disease bank and foundation materials.

LITERATURE CITED

- Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., and Boonham, N. 2009. Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10:537-545.
- Al Rwahnih, M., Daubert, S., Golino, D., Islas, C., and Rowhani, A. 2015. Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens in grapevine. *Phytopathology* 105:758-763.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Babcock, B. A. 2022. Economic impact of California's citrus industry in 2020. *J. Citrus Pathol.* 9. <http://dx.doi.org/10.5070/C49156433>
- Blagden, T., Schneider, W., Melcher, U., Daniels, J., and Fletcher, J. 2016. Adaptation and validation of e-probe diagnostic nucleic acid analysis for detection of *Escherichia coli* O157:H7 in metagenomic data from complex food matrices. *J. Food Prot.* 79:574-581.
- Bocsanczy, A. M., Espindola, A. S., and Norman, D. J. 2019. Whole-genome sequences of *Ralstonia solanacearum* strains P816, P822, and P824, emerging pathogens of blueberry in Florida. *Microbiol. Resour. Announc.* 8:e01316-18.
- Borst, A., Box, A. T. A., and Fluit, A. C. 2004. False-positive results and contamination in nucleic acid amplification assays: Suggestions for a prevent and destroy strategy. *Eur. J. Clin. Microbiol. Infect. Dis.* 23: 289-299.
- Bostock, R., Thomas, C., Hoenisch, R., Golino, D., and Vidalakis, G. 2014. Excluding pests and pathogens: Plant health: How diagnostic networks and interagency partnerships protect plant systems from pests and pathogens. *Calif. Agric.* 68:117-124.
- Bustin, S. A., Benes, V., Garson, J. A., Hellems, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., and Wittwer, C. T. 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55:611-622.
- Cambra, M., Gorris, M. T., Román, M. P., Terrada, E., Garnsey, S. M., Camarasa, E., Olmos, A., and Colomer, M. 2000. Routine detection of citrus tristeza virus by direct immunoprinting-ELISA method using specific monoclonal and recombinant antibodies. Pages 34-41 in: *Proc. 14th Conf. IOCV*.
- 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.
- Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. 2003. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics*. 10: Chapter Unit 10.3.
- Duran-Vila, N., Pina, J. A., and Navarro, L. 1993. Improved indexing of citrus viroids. *International Organization of Citrus Virologists Conference Proceedings (1957-2010)*. 12. <https://escholarship.org/uc/item/2gr565m0>
- Espindola, A. S., Cardwell, K., Martin, F. N., Hoyt, P. R., Marek, S. M., Schneider, W., and Garzon, C. D. 2022. A step towards validation of high-throughput sequencing for the identification of plant pathogenic oomycetes. *Phytopathology* 112:1859-1866.
- Espindola, A., Schneider, W., Hoyt, P. R., Marek, S. M., and Garzon, C. 2015. A new approach for detecting fungal and oomycete plant pathogens in next generation sequencing metagenome data tilizing electronic probes. *Int. J. Data Min. Bioinform* 12:115-128.
- 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., Sempertegui-Bayas, D., Bravo-Padilla, D. F., Freire-Zapata, V., Ochoa-Corona, F., and Cardwell, K. F. 2021. TASPERT: Target-Specific Reverse Transcript Pools to improve HTS plant virus diagnostics. *Viruses* 13:1223.
- Fuchs, M., Almeyda, C. V., Al Rwahnih, M., Atallah, S. S., Cieniewicz, E. J., Farrar, K., Foote, W. R., Golino, D. A., Gómez, M. I., Harper, S. J., Kelly, M. K., Martin, R. R., Martinson, T., Osman, F. M., Park, K., Scharlau, V., Smith, R., Tzanetakis, I. E., Vidalakis, G., and Welliver, R. 2021. Economic

- studies reinforce efforts to safeguard specialty crops in the United States. *Plant Dis.* 105:14-26.
- Gareth, J., Daniela, W., Trevor, H., and Robert, T. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, NY, U.S.A.
- Gergerich, R. C., Welliver, R. A., Osterbauer, N. K., Kamenidou, S., Martin, R. R., Golino, D. A., Eastwell, K., Fuchs, M., Vidalakis, G., and Tzanetakis, I. E. 2015. Safeguarding fruit crops in the age of agricultural globalization. *Plant Dis.* 99:176-187.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644-652.
- Hastie, T., Tibshirani, R., and Friedman, J. H. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, U.S.A.
- Ho, T., and Tzanetakis, I. E. 2014. Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* 471-473: 54-60.
- Krueger, R. R., and Vidalakis, G. 2022. Study and detection of citrus viroids in woody hosts. *Methods Mol. Biol.* 2316:3-21.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
- Li, W., Hartung, J. S., and Levy, L. 2007. Evaluation of DNA amplification methods for improved detection of "*Candidatus Liberibacter species*" associated with citrus huanglongbing. *Plant Dis.* 91:51-58.
- Liefting, L. W., Waite, D. W., and Thompson, J. R. 2021. Application of Oxford Nanopore technology to plant virus detection. *Viruses* 13:1424.
- Madden, T. L., Tatusov, R. L., and Zhang, J. 1996. Applications of network BLAST server. *Methods Enzymol.* 266:131-141.
- Massart, S., Chiumenti, M., De Jonghe, K., Glover, R., Haegeman, A., Koloniuk, I., Komiek, P., Kreuze, J., Kutnjak, D., Lotos, L., Maclot, F., Maliogka, V., Maree, H. J., Olivier, T., Olmos, A., Pooggin, M. M., Reynard, J.-S., Ruiz-García, A. B., Safarova, D., Schneeberger, P. H. H., Sela, N., Turco, S., Vainio, E. J., Varallyay, E., Verdin, E., Westenberg, M., Brostaux, Y., and Candresse, T. 2019. Virus detection by high-throughput sequencing of small RNAs: Large-scale performance testing of sequence analysis strategies. *Phytopathology* 109:488-497.
- Osman, F., Dang, T., Bodaghi, S., and Vidalakis, G. 2017. One-step multiplex RT-qPCR detects three citrus viroids from different genera in a wide range of hosts. *J. Virol. Methods* 245:40-52.
- Osman, F., Hodzic, E., Kwon, S.-J., Wang, J., and Vidalakis, G. 2015. Development and validation of a multiplex reverse transcription quantitative PCR (RT-qPCR) assay for the rapid detection of Citrus tristeza virus, Citrus psorosis virus, and Citrus leaf blotch virus. *J. Virol. Methods* 220: 64-75.
- Phannareth, T., Nunziata, S. O., Stulberg, M. J., Galvez, M. E., and Rivera, Y. 2021. Comparison of nanopore sequencing protocols and real-time analysis for phytopathogen diagnostics. *Plant Health Prog.* 22: 31-36.
- 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.
- Roistacher, C. N. 1991. *Graft-transmissible Diseases of Citrus: Handbook for Detection and Diagnosis*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Rong, W., Rollin, J., Hanafi, M., Roux, N., and Massart, S. 2023. Validation of high-throughput sequencing as virus indexing test for *Musa* germplasm: Performance criteria evaluation and contamination monitoring using an alien control. *PhytoFrontiers* 3:91-102.
- Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., Hayes, S., Lawrence, T., Birch, C., Bhagwat, B., and Rast, H. 2017. Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Dis.* 101:1489-1499.
- Shendure, J., and Ji, H. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26:1135-1145.
- Soltani, N., Stevens, K. A., Klaassen, V., Hwang, M.-S., Golino, D. A., and Al Rwahnih, M. 2021. Quality assessment and validation of high-throughput sequencing for grapevine virus diagnostics. *Viruses* 13: 1130.
- 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.
- Stobbe, A. H., Schneider, W. L., Hoyt, P. R., and Melcher, U. 2014. Screening metagenomic data for viruses using the e-probe diagnostic nucleic acid assay. *Phytopathology* 104:1125-1129.
- Vidalakis, G., Garnsey, S. M., Bash, J. A., Greer, G. D., and Gumpf, D. J. 2004. Efficacy of bioindexing for graft-transmissible Citrus pathogens in mixed infections. *Plant Dis.* 88:1328-1334.
- Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. 2019. High throughput sequencing for plant virus detection and discovery. *Phytopathology* 109:716-725.
- Visser, M., Bester, R., Burger, J. T., and Maree, H. J. 2016. Next-generation sequencing for virus detection: Covering all the bases. *Virol. J.* 13:85.
- Wang, X., Doddapaneni, H., Chen, J., and Yokomi, R. K. 2015. Improved real-time PCR diagnosis of Citrus stubborn disease by targeting prophage genes of *Spiroplasma citri*. *Plant Dis.* 99:149-154.