PhytoFrontiers™
An Open Access Journal from The American Phytopathological Society

Research

Evaluation of Metabarcoding Methods for Plant Disease Surveillance

Eric A. Newberry[1,†] | Subodh Srivastava[1,2] | Schyler O. Nunziata[1] | Reny Mathew[1,2] | Nakhla Mark[1] | Yazmín Rívera[1] |

[1] Plant Pathogen Confirmatory Diagnostics Laboratory, USDA, Animal Plant Health Inspection Service, Plant Protection and Quarantine, Science and Technology, Laurel, MD 20708

[2] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695

[†] Corresponding author: E. A. Newberry; eric.newberry@usda.gov

Accepted for publication 28 April 2023.

Abstract

Metabarcoding holds great potential for general plant pathogen surveillance by providing an untargeted profile of the host microbiome. However, the standard marker utilized for microbiome analysis of prokaryotes, the 16 rRNA gene, offers limited diagnostic utility as it provides phylogenetic resolution primarily at the genus level, and universal primers often co-amplify plant DNA. Here, we evaluated two recently published universal primer sets targeting the DNA gyrase and RNA polymerase ß subunit genes (*gyrB* and *rpoB*, respectively), relative to a plant discriminating 16S primer set (799F/1115R), as disease surveillance tools. Comparative analysis of a mock bacterial community, as well as naturally infected citrus variegated chlorosis samples, indicated that the *gyrB* method displayed optimal performance in the amplification of DNA across a broad taxonomic spectrum of plant-pathogenic bacteria, providing resolution at the species and, in some cases, subspecies levels. It also generated high-quality datasets with minimal to no co-amplification of plant DNA, outperforming the *rpoB* and 16S assays in these regards. Further evaluation revealed that the *gyrB* method displayed an overall linear trend in the detection of several diverse bacterial pathogens; however, it was at least an order of magnitude less sensitive than a standard real-time PCR assay. Finally, we demonstrate the potential of this method to disentangle a mixed population of *Pantoea* spp. associated with a rice-bacterial blight outbreak and identify putative novel pathogens that would otherwise be overlooked with conventional PCR-based tests. This work represents steps toward establishing a robust, untargeted metabarcoding method for general plant disease surveillance.

Keywords: bacterial diagnostics, metabarcoding, pathogen surveillance

Data availability: The raw fastq reads generated from this project were submitted to NCBI GenBank under BioProject PRJNA917542.

Funding

This research was supported by the intramural research program of the U.S. Department of Agriculture, Animal and Plant Health Inspection Service.

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

Globalized agricultural trade provides a pathway for the movement of plant-pathogenic microorganisms across international borders and contributes to an increased risk of pathogen incursions into new geographic regions (Anderson et al. 2004; Sikes et al. 2018). Indeed, anthropogenic introductions associated with trade have been attributed to the emergence of numerous plant disease epidemics in the United States,

such as Asiatic citrus canker (Brown 2001), laurel wilt (Harrington et al. 2011), and tomato yellow leaf curl virus (Polston et al. 1999), to name a few. To minimize the potential for introduction of exotic pathogens, plant protection organizations and industry partners invest heavily in phytosanitary, quarantine, and plant disease surveillance measures. Traditional immunodiagnostic and/or polymerase chain reaction (PCR)-based techniques are routinely employed as part of these efforts to detect specific pathogens that pose a known risk to an agricultural commodity. High-throughput sequencing (HTS) technologies hold great potential to enhance plant disease surveillance by providing an untargeted profile of the microbiome associated with a sample, thus enabling the identification of novel and emerging plant disease threats that may go overlooked with traditional diagnostic techniques.

The utility of HTS and metagenomics has been recognized in plant-virus surveillance, enabling the discovery of numerous novel and potentially emerging viral lineages from both germplasm collections and regional surveys (Fowkes et al. 2021; Malapi-Wight et al. 2021; Saad et al. 2021). Although HTS for viral diagnostics has matured to the point at which protocols are being validated for routine testing of commodities such as grapevine and citrus (Bester et al. 2021; Soltani et al. 2021), the full potential of microbiome analysis has not been realized in plant-bacterial diagnostics. Several targeted, metagenomic sequencing methods have been developed for the bacterial pathogens '*Candidatus* Liberibacter asiaticus' (Cai et al. 2019), *Xylella fastidiosa* (Román-Reyna et al. 2021), and *Xanthomonas perforans* (Mechan Llontop et al. 2020; Newberry et al. 2020) from plant samples. Although these studies demonstrate the potential of whole genome shotgun sequencing to identify bacterial pathogens with fine-scale resolution, this approach may require specialized sample processing methods to minimize host contamination in the resulting sequence data, significant sequencing depth, and computational resources, as well as specialized bioinformatic tools developed for analysis of the bacterial pathogen of interest. Metabarcoding is an untargeted technique that has the potential to circumvent many of the challenges listed above and make HTS more accessible for general diagnostic and plant disease surveillance purposes.

Metabarcoding involves the PCR amplification of a partial gene fragment that is conserved among a defined group of organisms (i.e., DNA barcode) from a complex DNA sample. The resulting amplicon is subsequently subjected to HTS, and the composition of taxa present within the sample is resolved based on sequence divergence observed within the amplified region. A number of universal primers have been developed for metabarcoding analysis of prokaryotes, most of which target the 16S rRNA gene due to its strong conservation (Bukin et al. 2019). Unfortunately, this marker offers limited diagnostic utility as it provides phylogenetic resolution primarily at the genus level, and universal primers often co-amplify plant-chloroplast and mitochondrial genes, which may significantly reduce the sequencing depth of bacterial DNA. Several modifications have been made to adapt 16S metabarcoding for plant microbiome analysis, such as plant-discriminating primers, PCR clamps, and blocking oligos, the effectiveness of which may depend on the plant species of interest (Giangacomo et al. 2021). Similarly, metabarcoding methods that target so-called bacterial housekeeping genes, such as the DNA gyrase and RNA polymerase ß subunits (*gyrB* and *rpoB*, respectively), have been developed for the purposes of profiling seed (Barret et al. 2015) or nematode (Ogier et al. 2019) associated bacterial communities, respectively. These markers provide an advantage over the 16S rRNA in that they offer greater phylogenetic resolution in bacterial species (and subspecies) identification and have less potential to cross-react with plant DNA. An added benefit is that both are single-copy genes present in most bacteria, allowing for improved quantification of taxon abundances relative to the multicopy 16S rRNA gene (Poirier et al. 2018).

Although these advancements in bacterial metabarcoding have facilitated a growing number of plant and food microbiome studies (Bartoli et al. 2018; Poirier et al. 2018; Rochefort et al. 2021), an evaluation of their application in plant disease diagnostics is needed. Here, we assessed the performance of three metabarcoding primer sets as diagnostic tools, including universal primers for the housekeeping genes *gyrB* (Barret et al. 2015) and *rpoB* (Ogier et al. 2019), as well as a 16S primer set (799F/1115R) designed to discriminate plant organellar DNA via mismatches in the forward primer binding site (Giangacomo et al. 2021). We evaluated these methods for the detection of a range of plant-pathogenic bacterial species and previously characterized diagnostic samples. The results of this work will help to establish a robust, untargeted metabarcoding method for general plant disease surveillance and diagnostic applications.

MATERIALS AND METHODS

Sample information and description of sequencing experiments

A number of previously characterized DNA samples were used to evaluate the metabarcoding methods described below and were tested using several real-time quantitative PCR (qPCR) assays prior to sequencing, which were run on a QuantStudio 5 system (Thermo Fisher Scientific). For sequencing, the Nano Reagent Kit (Illumina) was chosen for this work due to its shorter sequencing time and reduced price per sample relative to the corresponding Standard Reagent Kit (~$360 per kit, 28 h sequencing time vs. ~$1,220 per kit, 39 h sequencing time), which makes it better suited for diagnostic purposes when performing sequencing in-house. To optimize sequencing depth, no more than nine multiplexed libraries were sequenced in a single MiSeq run (see description of experiments below). Except for the mock community, all sequencing experiments included a control library prepared from either a healthy plant or non-host DNA extract.

Mock community construction. Metabarcoding methods were first benchmarked using a mock community of plant-pathogenic bacteria listed in Table 1. The bacterial strains used to construct this sample were provided to the USDA-APHIS-PPQ Plant Pathogen Confirmatory Diagnostics Laboratory as reference material from various researchers or as diagnostic specimens that were previously identified through Sanger sequencing and/or other PCR-based tests. In total, 15 strains were selected to provide a diverse sample of plant-pathogenic bacteria (encompassing 13 species and 8 genera) for their relative importance as plant pathogens and to test the metabarcoding methods' discriminatory power at the species and subspecies levels. Genomic DNA was extracted from pure cultures grown on nutrient agar plates for 48 h at 28°C. Three to four single colonies from each plate were suspended in 500 µl of sterile deionized water using a sterile pipette tip and subjected to DNA extraction using a DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. The resulting genomic DNA was assessed for quality and quantity using a NanoDrop spectrophotometer and Qubit dsDNA BR fluorometric assay (Invitrogen), then pooled in equal concentrations. Approximately 10 ng of mock community DNA was used as a template for each metabarcoding assay (described below), which were run in triplicate and sequenced in a single MiSeq run for a total of nine multiplexed libraries.

Citrus variegated chlorosis (CVC) leaf samples. Archived DNA extracted from sweet orange leaf samples (*Citrus sinensis* L.) collected in Argentina during 2014 were evaluated for a side-by-side comparison of the metabarcoding methods. These samples were naturally infected with the bacterial pathogen *Xylella fastidiosa* and displayed typical CVC symptoms, except for the healthy plant control. DNA extraction and qPCR analysis were performed as described by Li et al. (2013) using the *Xylella fastidiosa*-specific primer/probe set, XF16S. The *gyrB* and 16S amplicons obtained from these samples were pooled and sequenced in the same MiSeq run (eight multiplexed libraries), and the *rpoB* amplicons were sequenced separately along with other *rpoB* samples for a total of seven multiplexed libraries.

Spiked pepper seed extracts. To evaluate the limit of detection and intermediate precision of the *gyrB* metabarcoding assay, total nucleic acid (~1 ng μl^{-1} DNA concentration as measured by Qubit) isolated from a 3.0-g sample of pepper seeds (*Capsicum annuum*) was used as a matrix to generate a five-point dilution series, spiked with purified genomic DNA from *Ralstonia solanacearum* strain 801-01 (isolated from a geranium plant in 2020). The pepper seed nucleic acid was extracted using an Sbeadex Maxi Plant Kit (Biosearch Technologies) according to the manufacturer's instructions. The final concentration of *R. solanacearum* DNA in each dilution point ranged from 1.5×10^4 to 1.5×10^0 fg μl^{-1}. Each dilution was assessed for the presence of *R. solanacearum* through qPCR using the RsSA3 assay described by Stulberg et al. (2016) prior to sequencing. In total, three MiSeq runs were performed by two operators. The first run included a negative control consisting of the original pepper seed extract (before spiking) and four dilution points, whereas the two subsequent runs included an additional dilution point for a total of six multiplexed libraries.

Assorted environmental DNA extracts. Previously characterized environmental DNA extracts collected from various sources were selected for sequencing to further evaluate the *gyrB* metabarcoding assay with a range of plant pathogens and testing backgrounds (Table 2). DNA extraction from these samples was performed using the DNeasy PlantMini Kit (Qiagen) according to the manufacturer's instructions, using 100 to 200 mg of tissue, depending on the sample type. These included three genomic DNA samples (*Echinacea*, alder, and potato leaves) that tested positive for '*Candidatus* phytoplasma' species, two rice (*Oryza sativa*) DNA samples collected from a bacterial blight outbreak in Arkansas associated with *Pantoea ananatis* (Luna et al. 2023), and a lime DNA sample (*Citrus* sp.) collected from a tree in California that was positive for '*Ca.* Liberibacter asiaticus'. Additionally, we evaluated three effluent water samples collected from the run-off of *Pelargonium* plants that were artificially inoculated with *R. solanacearum* strain UW551. Briefly, potted *Pelargonium* were subjected to drench inoculation as described by Stulberg and Huang (2015). Seven days after inoculation, 275 ml of effluent water was collected from individual plants after watering, and bacterial cells were concentrated using an analytical filter funnel with 47 mm diameter cellulose nitrate and 0.45 μm pore size filter membrane (Nalgene). Filters were collected, and DNA extraction was performed using a DNeasy Powersoil Pro Kit (Qiagen).

Library preparation and sequencing

Amplicon libraries were constructed using two rounds of PCR according to the workflow described in the Illumina 16S Sample Preparation Guide with minor modifications (Illumina 2013). The first-round PCRs were performed as previously described using the gyrB_aF64/gyrB_aR353 (2 μM; Barret et al. 2015), Univ_rpoB_F_deg/Univ_rpoB_R_deg (0.5 μM; Ogier et al. 2019), and 799F/1115R (0.2 μM; Giangacomo et al. 2021) primer sets. High-fidelity AccuPrime *Taq* DNA polymerase (Invitrogen) was used for the *gyrB* and *rpoB* PCRs, whereas the Kappa HiFi Hot Start Ready Mix (Roche) was used for the 16S with 2 μl of DNA template. Amplicons were purified with AMPure XP magnetic beads (Beckman Coulter), then used as a template (5 μl) for the second-round PCR to incorporate Nextera XT indexes (Illumina). Cycling conditions for the *gyrB* and *rpoB* amplicons were 94°C (2 min), followed by 12 cycles of 94°C (60 s), 55°C (60 s), and 68°C (60 s) and a final extension at 68°C (10 min). Cycling conditions for the 16S amplicons were 95°C (3 min), followed by eight cycles of 95°C (30 s), 55°C (30 s), and 72°C (30 s) and a final extension at 72°C (5 min). Amplicons were purified as described above, quantified using a Qubit BR dsDNA fluorometric assay (Invitrogen), and subsequently pooled in equimolar concentrations (4 nM). The pooled libraries were denatured and diluted to a final concentration of 12 pM, spiked with 30% PhiX, then sequenced on an Illumina MiSeq instrument using a v2 500 Nano Reagent Kit to generate 250-bp paired-end reads.

Sequence analysis

The fastq reads were processed using the DADA2 (v1.2) pipeline (Callahan et al. 2016). To minimize sequencing errors

TABLE 1

List of bacterial strains used to construct the mock bacterial community

Species	Strain number	Host	Source
Xanthomonas citri pv. *mangiferaeindicae*	DGS42	Mango	PPCDL[a]
Xanthomonas vasicola pv. *vasculorum*	NCPPB 1326	Sugarcane	J. Leach
Ralstonia pseudosolanacearum	GMI1000	Tomato	C. Allen
Ralstonia solanacearum	UW551	Geranium	C. Allen
Pseudomonas syringae (genomospecies 1)	ICMP 18960	Kiwi	J. Stack
Pseudomonas syringae (genomospecies 1)	ICMP 13107	Kiwi	J. Stack
Pseudomonas syringae pv. *actinidilfoliorum* (genomospecies 3)	ICMP 18803	Kiwi	J. Stack
Pseudomonas syringae pv. *actinidiae* (genomospecies 3)	ICMP 18801	Kiwi	J. Stack
Dickeya fangzhondai	12320	Orchid	PPCDL
Erwinia aphidicola	SA615	Melon	PPCDL
Pectobacterium carotovorum	Pcc16	Potato	A. Charkowski
Pectobacterium atrosepticum	Pa15	Potato	A. Charkowski
Musicola paradisiaca	Ech703	Banana	A. Charkowski
Dickeya solani	4321-01	Potato	A. Charkowski
Curtobacterium flaccumfaciens	2901-01	Watermelon	PPCDL

[a] PPCDL, archived diagnostic sample identified at the USDA-APHIS-PPQ, Plant Pathogen Confirmatory Diagnostic Laboratory (PPCDL).

observed in amplicon sequence variant (ASV) inference from the mock community dataset, more stringent quality filtering and trimming parameters were used, including minLen = 100; truncLen, c(240, 220); maxN = 0; maxEE = 1; truncQ = 5, and trimLeft set to c(20,40) for *gyrB* and 16S, whereas c(30,30) was used for *rpoB*. These parameters were laxened for datasets generated from environmental DNA extracts, with the maxEE option adjusted to 2 (indicating reads with > 2 expected errors will be discarded) and trimLeft set to c(20,25) for *gyrB* and c(20,20) for the 16S. Default parameters were used for the subsequent pipeline functions. Taxonomic annotations were assigned to the ASVs using the reference databases described by Barret et al. (2015) and Ogier et al. (2019) for the *gyrB* and *rpoB* datasets, respectively, whereas the Ribosomal Database Project training set 18 (Cole et al. 2014) was used for the 16S datasets. At least 70% bootstrap support was required to assign a taxonomic annotation. BLASTn analysis of NCBI GenBank was used for the initial identification of pathogen-specific ASVs in the datasets. For selected ASVs, multiple sequence alignments were constructed using the Muscle algorithm (v3.8; Edgar 2004) implemented in Geneious Prime software (v2020.0.5; Biomatters). Approximately maximum likelihood phylogenetic trees were generated from these alignments using FastTree (v2.1.11; Price et al. 2010) with the Jukes-Cantor model of nucleotide substitution. To verify the quantification of *X. fastidiosa* in the CVC datasets, the paired-end reads were used to construct a custom BLAST database in Geneious Prime and queried (E value $\leq 1e^{-100}$) with the following reference sequences obtained from GenBank: AF192343.1 (16S), CP002165.1:1124828-1127272 (*gyrB*), and NC_004556.1:c2364357-2360203 (*rpoB*). At least 95% pairwise identity and 90% query coverage were required for a read to be considered *X. fastidiosa*.

RESULTS AND DISCUSSION

Evaluation of metabarcoding methods

The *gyrB* (Barret et al. 2015), *rpoB* (Ogier et al. 2019), and 16S (799F-1115R, plant-discriminating) primer sets (Giangacomo et al. 2021) were first benchmarked by sequencing a mock community composed of 15 plant-pathogenic bacterial strains, encompassing 13 species and 8 genera (Table 1). The results of this experiment (Table 3) revealed that the *gyrB* assay successfully amplified DNA from all 13 species and generated four discrete ASVs for each of the strains within the *Pseudomonas* genus. Notably, two of these strains (*P. syringae* pv. *actinidiae* ICMP 18801 and *P. syringae* pv. *actinidifoliorum* ICMP 18803) differed by just a single-nucleotide polymorphism at the barcoding locus, highlighting the precision of the "de-noising" approach in analysis of amplicon sequence data implemented by DADA2. In contrast, the *rpoB* assay failed to detect *Curtobacterium flaccumfaciens* and generated an inflated number of ASVs within the *Pseudomonas* genus. A manual inspection suggested that these ASVs were likely chimeric sequences associated with merged reads derived from several closely related *Pseudomonas* strains/species and was likely an artifact of sequencing a larger insert (~430 bp) produced by the *rpoB* primers (relative to the 16S and *gyrB* assays). On average, this marker yielded fewer paired-end reads (32,142 ± 3,016) relative to the 16S (97,735 ± 46,067) and *gyrB* (79,770 ± 1,843) libraries, all sequenced in the same run. Therefore, it is unclear whether the lack of detection for *C. flaccumfaciens* was the result of shallower sequence depth and/or amplification bias of the primers. Finally, the 16S assay failed to produce corresponding ASVs for all *Xanthomonas* and *Ralstonia* species in the sample. It also generated a greater number of ASVs than

TABLE 2

Metadata for diagnostic and spiked effluent water samples evaluated with the *gyrB* metabarcoding assay

MiSeq run (pathogen)	Sample ID	Host (disease)	Collection state	Year	qPCR assay (reference)	Cq value[a]	Total number of reads[b]	Number of target reads
1 '*Ca*. Phytoplasma sp.'	J1	*Echinacea* (phyllody)	Maryland	NA[c]	23S-JH (Hodgetts et al. 2009)	17.5	128,984	35,919
	WA	Alder (alder yellows)	Washington	2021		21.3	94,302	4,660
	AK	Potato (purple top)	Alaska	2019		29.2	130,420	5 (2)[e]
	Negative control	Blank buffer extraction	NA	NA		Und[d]	106,931	0
2 *R. solanacearum*	4A	Effluent water	Maryland	2020	RsSA3 (Stulberg et al. 2016)	25.1	99,917	334
	2A	Effluent water	Maryland	2020		32.1	90,664	0
	0A	Effluent water	Maryland	2020		36.0	69,971	0
	Negative control	Effluent water	Maryland	2020		Und	119,392	0
3 (*P. ananatis* or '*Ca*. Liberibacter asiaticus')	R1A	Rice (bacterial blight)	Arkansas	2021	Multiplex cPCR (Kini et al. 2021)	NA	77,832	49
	R2A	Rice (bacterial blight)	Arkansas	2021		NA	90,202	66,352
	121420	Lime (HLB)	California	2019	RNR (Zheng et al. 2016)	19.5	26,032	12,247

[a] Real-time PCR cycle quantification (Cq) value.
[b] Total number of paired-end reads after quality filtering with DADA2.
[c] Not available (NA).
[d] Undetermined (Und).
[e] Five reads detected by DADA2 were suspected to be the result of index hopping from the *Echinacea* sample. Two additional, sample-specific reads were identified through BLASTn.

expected for *Pectobacterium* and *Erwinia*, indicating the lack of resolution provided by this marker.

To further evaluate these metabarcoding methods, we compared their performance in the detection of *Xylella fastidiosa* from naturally infected CVC samples. Similar to the mock community sequencing, the *rpoB* libraries exhibited shallower sequencing depth and greater sample-to-sample variation in overall yield relative to the 16S and *gyrB* libraries (Fig. 1). Due to the presence of a secondary, ~640-bp amplicon generated by the *rpoB* PCR (suggesting a ~570-bp insert; Supplementary Fig. S1), a greater proportion of reads did not pass quality filtering as they went unmerged, leaving between 48 and 59% of the original dataset for analysis. Surprisingly, investigation of these unmerged reads revealed that they did not originate from sweet orange DNA but were predominantly related to the bacterial genus *Terriglobus* (Supplementary Table S1), which was predicted to generate a 560-bp amplicon using the Univ_rpoB_F_deg/Univ_rpoB_R_deg primer pair based on in-silico analysis (data not shown). Thus, variation in the length of the *rpoB* barcoding locus targeted by these primers may present technical challenges for certain plant- and soil-associated bacterial communities with Illumina short-read technology. In contrast, a majority of the 16S (~88%) and *gyrB* (~76%) reads passed quality filtering and produced more consistent results in terms of sequencing depth. However, approximately 84% of the 16S reads could be attributed to host contamination, indicating that the 799F/1115R primer pair does not universally discriminate plant DNA and should be validated prior to use with a new plant species (Fig. 1).

When investigating the microbial community composition, we identified a discrete *X. fastidiosa* ASV in two samples (CVC4 and CVC1) for both the *gyrB* and 16S datasets; however, the pathogen went undetected with *rpoB* (Table 4). We further verified these results through BLASTn analysis of the paired-end reads, which corroborated the accuracy of DADA2 and revealed a single *X. fastidiosa* read in the *rpoB* sample CVC4, which likely did not pass quality filtering. As this sample carried the highest *X. fastidiosa* titer (as indicated by qPCR) and exhibited comparable sequenc-

TABLE 3

Evaluation of the 16S, *gyrB*, and *rpoB* primer sets in the detection of a mock community composed of 13 plant-pathogenic bacterial species (data are the average of three technical replicates sequenced in a single MiSeq run)

Genus	Number of species (number of strains)	Number of observed ASVs[a]		
		16S	*gyrB*	*rpoB*
Curtobacterium	1.0	1.0	1.0	0.0
Dickeya	2.0	2.0	2.0	2.0
Erwinia	1.0	2.0	1.0	1.0
Musicola	1.0	1.0	1.0	1.0
Pectobacterium	2.0	3.3	2.0	2.0
Pseudomonas	2.0 (4.0)	2.3	4.7	12.0
Ralstonia	2.0	1.0	2.0	2.0
Xanthomonas	2.0	1.0	2.0	2.0
Summary statistics				
Total	13.0 (15.0)	13.7	15.7	22.0
Undetected species[b]	–	2.0	0.0	1.0
Redundant ASVs[c]	–	2.7	0.7	8.0

[a] Average number of observed amplicon sequence variants (ASVs) with a sequencing depth of ≥10 reads.
[b] Average number of species/strains present but not detected in the dataset.
[c] Average number of observed ASVs that exceeded the number of species/strains.

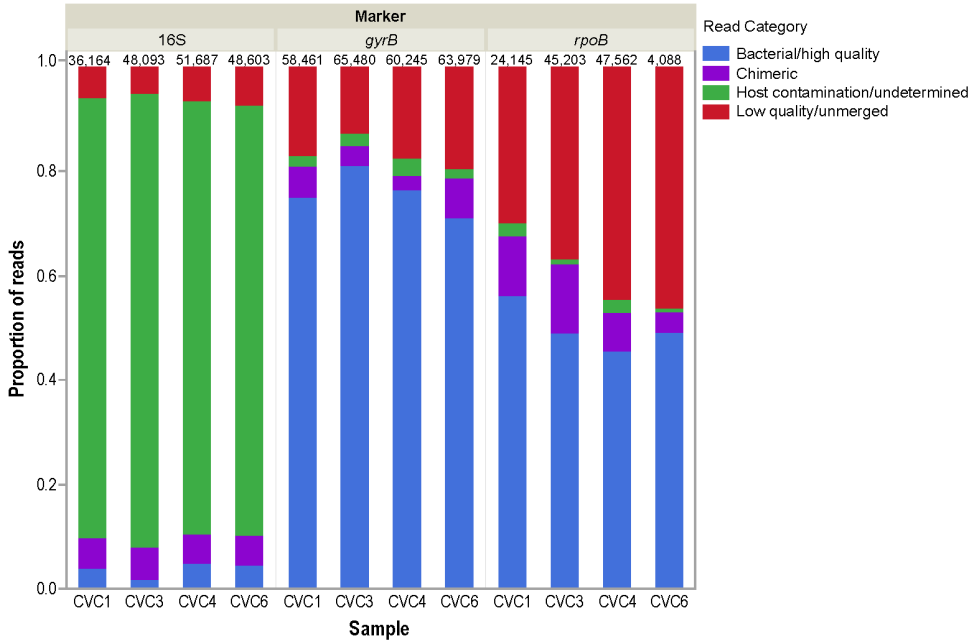

FIGURE 1
Stacked bar chart summarizing the proportion of reads from each citrus variegated chlorosis sample discarded due to low quality (red), being tagged as chimeric (purple), host contamination, or undetermined origin (green) and high-quality bacterial reads obtained for downstream analysis (blue). The *gyrB* and 16S amplicons were sequenced in a single run using a MiSeq V2 500-cycle Nano Reagent Kit, and the *rpoB* amplicons were sequenced in a separate run to optimize sequencing depth. Numbers above the bars indicate the total number of paired-end reads obtained per sample.

ing depth (47,562 reads) relative to the 16S (51,687 reads) and *gyrB* (60,245 reads) libraries, lack of detection for *X. fastidiosa* with the *rpoB* primers here was likely the result of amplification bias. Finally, it was surprising to note that both the *gyrB* and 16S assays generated a similar number of *X. fastidiosa* reads per sample, despite significant differences in the sequencing depth of bacterial DNA. This result may be an indication of the overall low abundance of background microflora associated with the orange leaf petioles, which has been similarly reported for metagenomic studies of *X. fastidiosa* in grapevine (Johnson et al. 2022). Finally, a phylogenetic analysis of *X. fastidiosa* ASVs revealed that the *gyrB* haplotypes correlated with the major subspecies of this pathogen, suggesting the presence of *X. fastidiosa* subsp. *pauca* in the sweet orange samples (Fig. 2), which is the subgroup responsible for CVC disease (Cordeiro et al. 2014). Similar intraspecific resolution was not obtained with 16S *Xylella* ASVs (data not shown).

Overall, it was unclear why the 799F/1115R primer pair failed to discriminate the sweet orange DNA, as these primers have proven effective with numerous other plant species (Anguita-Maeso et al. 2022; Laforest-Lapointe et al. 2016). Nonetheless, the lack of phylogenetic resolution offered by this marker limits its diagnostic utility. Although the *rpoB* method overcame these limitations, amplification bias and technical complications associated with sequencing a larger amplicon with Illumina technology were concerns. Although it may be possible to improve the performance of the *rpoB* assay through optimization of our sequencing and bioinformatic methods, *gyrB* did not suffer from these limitations. As this method successfully amplified DNA from a broad range of plant-pathogenic bacteria, providing insights into intraspecific pathogen diversity with little to no evidence of host contamination, we chose to further characterize its performance as an untargeted plant disease diagnostic tool.

TABLE 4

Identification of *Xylella fastidiosa* (*Xf*) in archived citrus variegated chlorosis samples (sweet orange) among three metabarcoding primer sets using BLASTn or DADA2

Sample	Xf16S Cq[a]	Marker	Number of *Xf* reads detected with BLASTn	Number of *Xf* reads detected with DADA2 (%)[b]
CVC4	25.34	*gyrB*	65	65 (0.14)
		rpoB	1	0
		16S	73	73 (2.95)
CVC1	28.29	*gyrB*	6	6 (0.01)
		rpoB	0	0
		16S	10	10 (0.71)
CVC6	32.29	*gyrB*	0	0
		rpoB	0	0
		16S	0	0
CVC3	Und.	*gyrB*	0	0
		rpoB	0	0
		16S	0	0

[a] Cycle quotient (Cq) value for the *X. fastidiosa* Xf16S qPCR (Li et al. 2013).
[b] Percentage of total bacterial reads.

Validation of the *gyrB* metabarcoding assay

As the *gyrB* metabarcoding method was developed for the purposes of profiling the seed microbiome, we assessed its perfor-

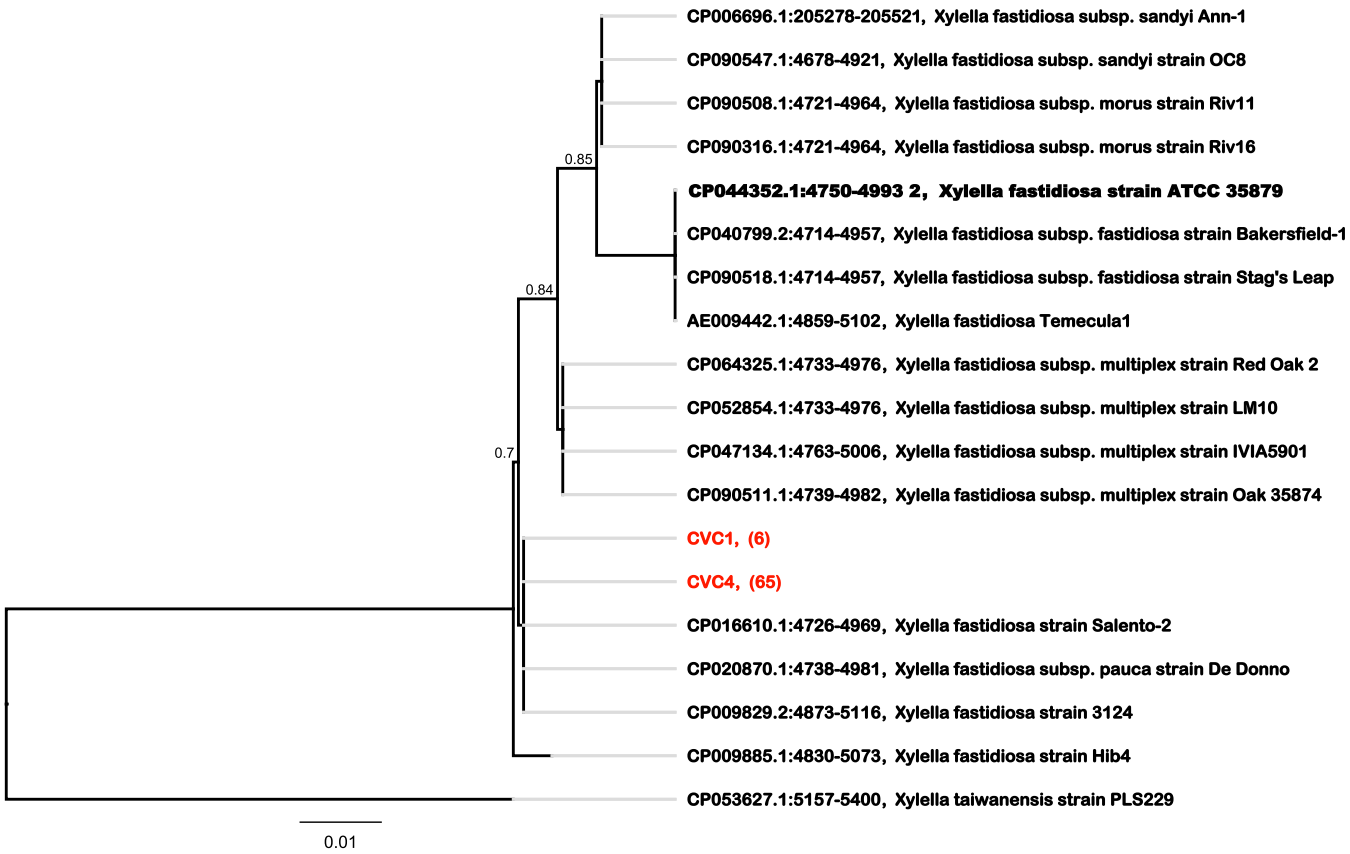

FIGURE 2
Phylogenetic tree of 19 *Xylella fastidiosa* strains based on a 244-bp alignment of the partial *gyrB* gene. The amplicon sequence variants (ASVs) obtained from citrus variegated chlorosis samples CVC1 and CVC4 are shown in red, with the number of reads sequenced for each ASV in parentheses. The *X. fastidiosa* type strain is shown in bold.

mance in the detection of *R. solanacearum* (*Rs*) DNA spiked in a matrix of pepper seed extract. The results of these experiments (Table 5) indicated a strong linear trend ($R^2 = 0.964$, $P < 0.0001$) between the *Rs* titer and $\log_{10}$ number of reads sequenced for the target across four dilution points, indicating the semi-quantitative nature of this method. A limit of detection (LoD) of approximately 300 fg of *Rs* DNA per PCR reaction (Cq value = 31.6) in the background of the larger pepper seed microbiome was determined, which was just above the LoD observed for the RsA3 qPCR. Among two operators and three independent library preps/MiSeq runs, the coefficient of variability (CV) for overall yield in quality filtered reads ranged from 2.0 to 19.0%. Some of this variation can be attributed to the number of multiplexed libraries sequenced per run, where the first included five and the two subsequent runs included six. Despite this moderate variability in overall yield, the $\log_{10}$ transformed number of *Rs* reads sequenced was highly consistent for the first two dilution points (CV $\leq$ 3.0%), whereas variability spiked at the LoD (14.0%), likely due to stochastic error.

Having established the sensitivity and intermediate precision, we sought to further characterize the performance of the *gyrB* assay with a range of previously characterized environmental DNA extracts (Table 2). Among the plant species tested, alder and lime generated secondary, nonspecific PCR products due to the co-amplification of plant DNA. However, this only significantly affected the proportion of high-quality bacterial reads sequenced for lime (Supplementary Fig. S2). Nonetheless, ASVs for the corresponding target pathogens, including '*Ca*. Phytoplasma sp.', *R. solanacearum* in the background of effluent water, and '*Ca*. Liberibacter asiaticus', were detected in samples up to a Cq value of 29.2 (Table 2). Despite significant differences in dataset quality, target pathogen, and sample background, we still observed a strong correlation between the $\log_{10}$ transformed number of target reads and Cq value (Fig. 3). The linear regression model generated from these data suggested that to obtain 10 reads for a given pathogen, one should expect a Cq value of approximately 27.8. This result was well above the estimated sensitivity of the spiked pepper seed extracts described above and is likely associated with the use of artificially spiked DNA for those samples. Despite these discrepancies, our results indicate that this method is at least a magnitude of order less sensitive than a standard qPCR assay, indicating that metabarcoding analysis may be effective for plant samples when a pathogen is expected to be present in moderate to high titer. However, it is important to note that the sensitivity of this method is not absolute and may be influenced by the target pathogen, complexity of the background microbial community, and sequencing depth, among other factors.

Considerations for index hopping and sample determination

When investigating the identity of the sequences obtained from the phytoplasma dataset, we noted that the potato and *Echinacea* samples contained identical ASVs that clustered with other members of the 16SrI group ('*Ca*. P. asteris' related; Fig. 4). Although the specific identity of the phytoplasmas in these samples was unknown prior to sequencing, we suspected that this result might be an artifact of index hopping, which is an issue that has been previously documented (Barret et al. 2015). This was especially of concern because the *Echinacea* was a high-titer sample (Cq = 17.5), the potato was low titer (Cq = 29.2), and no phytoplasma sequences were detected in the negative control to indicate contamination or user error. We subsequently performed BLASTn analysis of the potato reads and identified two additional phytoplasma sequences that were distinct from the ASV inferred by DADA2. One clustered with other members of the 16SrIII group ('*Ca*. P. pruni' related), which have been previously been associated with potato purple top disease in Alaska (Lee et al. 2009),

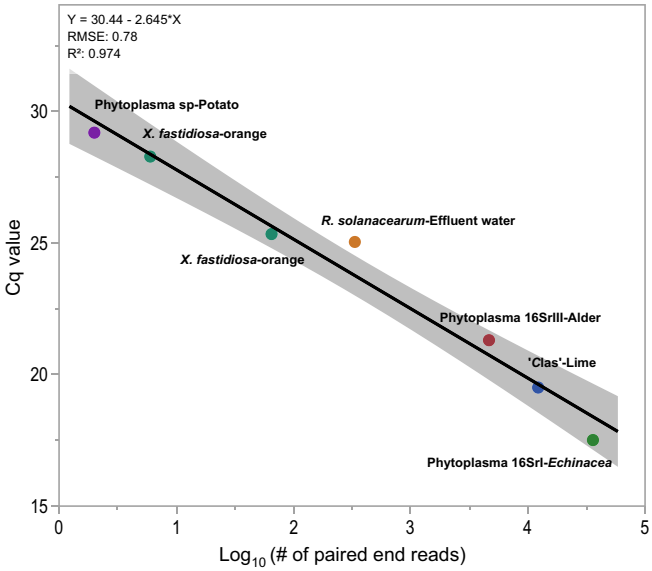

FIGURE 3

Linearity of the *gyrB* metabarcoding method in the detection of several plant-pathogenic bacteria from naturally infected plant samples and effluent water collected from artificially infected *Pelargonium* plants with *Ralstonia solanacearum*. The 16S rRNA group of the '*Candidatus* Phytoplasma' species is shown where available.

TABLE 5

Sensitivity and intermediate precision of the *gyrB* metabarcoding method in the detection of *R. solanacearum* (*Rs*) spiked in a matrix of pepper-seed DNA

Dilution	*Rs* fg/reaction	RsSA3 Cq[a]	Total reads (N)[b]		Mean number of *Rs* reads	$\log_{10}$ (number of *Rs* reads)		Fraction positive
			Mean	CV[c]		Mean	CV	
1	30,000	24.5	79,203	19.0%	15,474	4.2	3.0%	3/3
2	3,000	28.1	83,502	11.8%	1,906	3.3	2.1%	3/3
3	300	31.6	77,477	7.0%	130	2.1	14.9%	3/3
4	30	35.5	77,613	5.0%	24	1.4	–	1/3
5	3	Und[d]	71,744	7.9%	0	0	–	0/2
Control	0	Und	77,557	2.0%	0	0	–	0/3
Mean CV				8.79 ± 5.96%			5.48 ± 7.13%	

[a] Cycle quotient (Cq) value for the RsSA3 qPCR described by Stulberg et al. (2016).
[b] Total number of quality-filtered reads.
[c] Coefficient of variation (CV).
[d] Undetermined (Und).

whereas the other could not be identified due to lack of reference sequences and branched intermediate of a '*Ca*. P. pini'-related strain (16SrXXI) and '*Ca.* P. ziziphi' (16SrV), suggesting a potential mixed infection (Fig. 5).

Although it is routine practice to control for index hopping by setting a threshold filter for low abundant ASVs in the dataset (Barret et al. 2015; Ogier et al. 2019), this practice may discard true biological sequences. Therefore, caution should be exercised when determining the presence of a pathogen with low sequencing depth using this method for diagnostic purposes, especially when multiplexed along with other high-titer samples. Nonetheless, this is not a limiting factor as further confirmation with more specific diagnostic tests is routinely performed following a positive result with a screening assay such as this. Although preliminary data obtained from the mock community analysis described above indicated that ASVs with a sequencing depth of <10 could primarily be attributed to index hopping or were redundant (data not shown), further work is needed to establish a more precise cutoff for sequencing with a Nano Reagent Kit.

Identification of a putative mixed infection associated with rice bacterial blight

Finally, we performed metabarcoding analysis of two plant DNA samples associated with a rice-bacterial blight outbreak in Arkansas during the summer of 2021 (Luna et al. 2023), suspect for *Pantoea* sp. (Table 2). Using a conventional multiplex PCR assay described by Kini et al. (2021), one of these samples (R2A) tested positive for *P. ananatis*, whereas the other (R1A) was negative. Surprisingly, our results revealed the presence of multiple *Pantoea* ASVs in both samples, corresponding to a least three different species, *P. ananatis*, *P. eucrina*, and *P. deleyi*, (Fig. 5). Although *P. ananatis* was clearly the dominant species in the PCR-positive sample R2A (comprising ~74% of the reads), *P. eucrina*, *P. deleyi*, and a third *P. eucrina*-related ASV (R1A_ASV3) were codominant (each comprising ~3.4% of the reads) in the PCR-negative sample, with *P. ananatis* collecting just 0.06% of the reads. Although we cannot be certain as to the role *P. eucrina* and *P. deleyi* may play in this outbreak without isolation of the bacterial species and pathogenicity analysis, *P. eucrina* has been identified as part of a species complex responsible for bacterial blight of cotton in Pakistan (Tufail et al. 2020). Similarly, *P. deleyi* was originally isolated from eucalyptus leaves showing symptoms of bacterial blight and die-back in Uganda (Brady et al. 2009). This example highlights the application of *gyrB* metabarcoding to disentangle a potential mixed infection that would otherwise be overlooked using conventional diagnostic techniques. Bio-surveillance using methods such as those evaluated here could be used to inform research priorities and evaluate the risk of emerging pathogens as part of a broader early detection and disease surveillance program.

CONCLUSIONS

The combined results of these experiments highlight the challenges of marker selection and the need for method validation prior to implementing metabarcoding in plant disease diagnostics. The *gyrB* method proved the most versatile in detecting a range of plant-pathogenic bacteria across the taxonomic spectrum and in producing high-quality metagenetic datasets from a variety of plant and environmental DNA extracts. Furthermore, the reproducible and semi-quantitative nature of this technique was

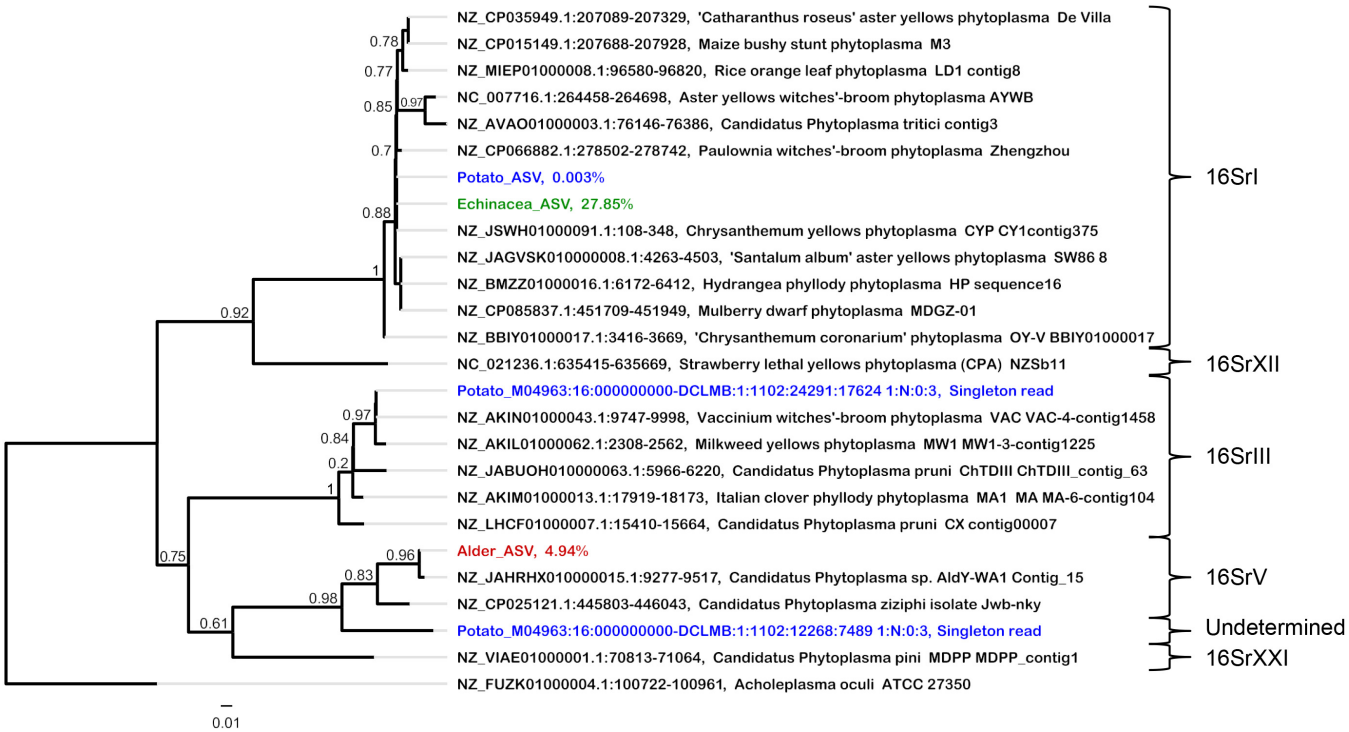

FIGURE 4
Identification of '*Candidatus* Phytoplasma sp.' in the metabarcoding datasets. The phylogenetic tree was based on a 241-bp alignment of 26 partial *gyrB* gene sequences. Amplicon sequence variants inferred by DADA2 or singleton reads identified through BLASTn analysis from the *Echinacea*, alder, and potato samples are color coded, with their relative abundance in the dataset to the right of the taxon label. Reference strains and their corresponding GenBank accession numbers are in black. Brackets denote the 16S rRNA group classification of the reference strains. The tree was rooted on *Acholeplasma oculi* strain ATCC 27350, and branch support was assessed using 1,000 bootstrap replicates.

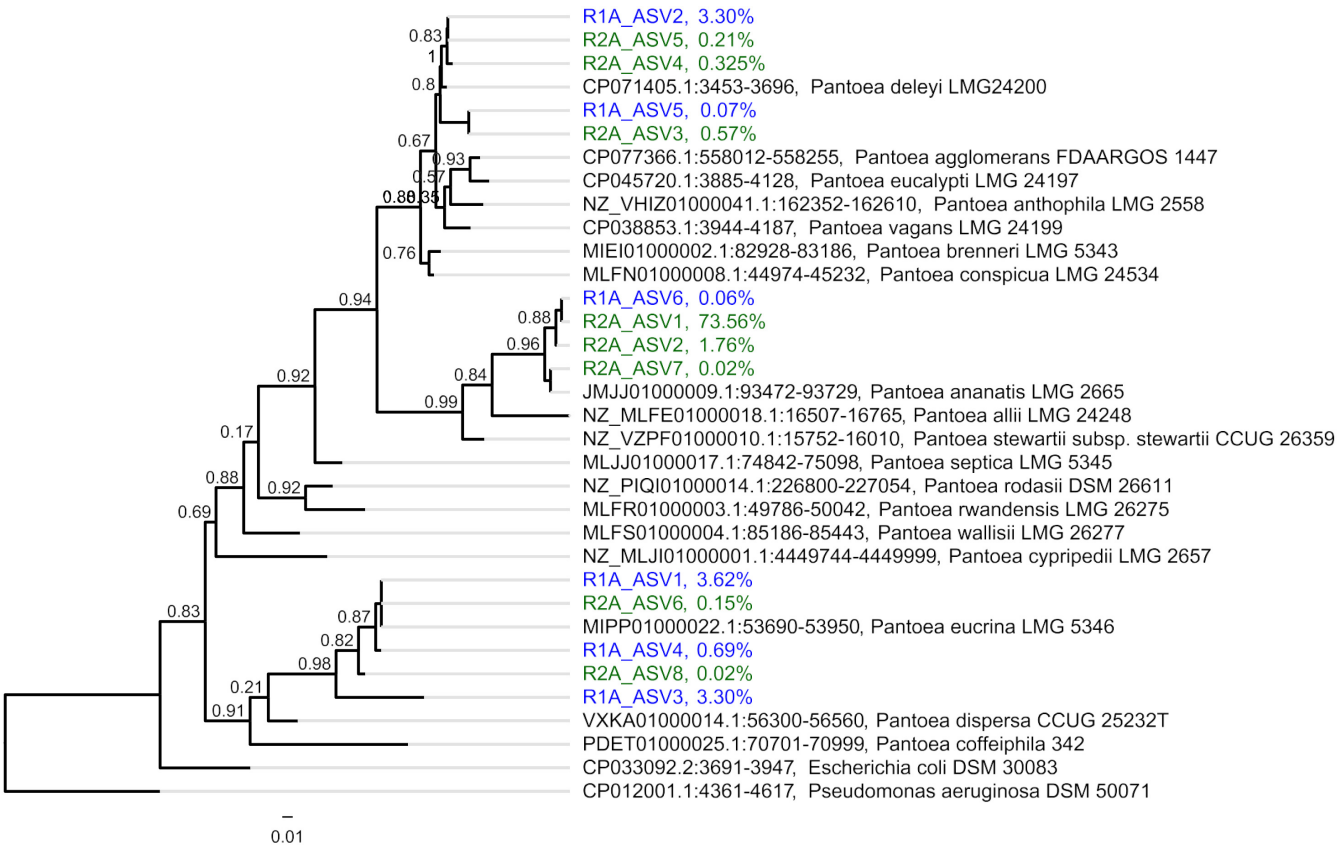

FIGURE 5

Identification of *Pantoea* species in rice-bacterial blight samples R1A (blue) and R2A (green). The phylogenetic tree was based on a 244-bp alignment of 34 partial *gyrB* gene sequences. The relative abundances of amplicon sequence variants (ASVs) in the datasets are shown to the right of the taxon label. Reference strains and corresponding GenBank accession numbers, representative of 18 *Pantoea* type species, are shown in black. The tree was rooted on *Pseudomonas aeruginosa* strain DSM 50071, and branch support was assessed using 1,000 bootstrap replicates.

verified, indicating its potential as a general plant disease diagnostic and surveillance tool. An inherent limitation of metabarcoding is amplification bias associated with the primer pair in use (Giangacomo et al. 2021). Despite issues with host contamination and the limited phylogenetic resolution provided by 16S metabarcoding, it may be best practice to pair *gyrB* amplicon sequencing with the 16S rRNA gene to ensure sufficient representation of the bacterial community and that taxa of interest are not missed. Nonetheless, the potential of *gyrB* metabarcoding to disentangle mixed infections and identify putative novel pathogens with intraspecific resolution makes it an attractive complement to the 16S rRNA gene and traditional plant disease diagnostic techniques.

For 16S metabarcoding, it is generally accepted that 100k raw read pairs is sufficient to survey bacterial community composition (Gołębiewski and Tretyn 2020). We found that this level of sequencing depth provided moderate sensitivity in pathogen detection (Cq value $< \sim 30$) from diseased plant samples and seeds with *gyrB* and was at least a magnitude of order less sensitive than a standard qPCR assay. Therefore, we recommend multiplexing no more than six samples with the Nano Reagent Kit to obtain sufficient sequencing depth for reproducible results. Of course, improved sensitivity may be obtained by increasing the sequencing depth using a standard MiSeq reagent kit; however, the price per sample may be prohibitive when sequencing a small number of diagnostic samples in-house. It should be noted that the sensitivity of this method is also influenced by the complexity of the background microbial population and may need to be verified for nonroutine diagnostic samples, for example, DNA extracted from soil or roots. Index hopping is also a concern with this method. Therefore, caution should be exercised when determining the presence of a pathogen with low sequencing depth, especially when multiplexed along with other high-titer samples. Finally, further benchmarking with a variety of plant species and bacterial pathogens is needed to bolster the adoption of this technique for plant disease surveillance and diagnostics.

ACKNOWLEDGMENTS

We thank G. Santillana and S. Costanzo for providing the DNA samples extracted from effluent water and the '*Ca*. Phytoplasma sp.' samples, respectively.

LITERATURE CITED

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., and Daszak, P. 2004. Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 19:535-544.

Anguita-Maeso, M., Haro, C., Navas-Cortés, J. A., and Landa, B. B. 2022. Primer choice and xylem-microbiome-extraction method are important determinants in assessing xylem bacterial community in olive trees. Plants 11:1320.

Barret, M., Briand, M., Bonneau, S., Préveaux, A., Valière, S., Bouchez, O., Hunault, G., Simoneau, P., and Jacquesa, M.-A. 2015. Emergence shapes the structure of the seed microbiota. Appl. Environ. Microbiol. 81: 1257-1266.

Bartoli, C., Frachon, L., Barret, M., Rigal, M., Huard-Chauveau, C., Mayjonade, B., Zanchetta, C., Bouchez, O., Roby, D., Carrère, S., and Roux, F.

2018. *In situ* relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. ISME J. 12:2024-2038.
Bester, R., Cook, G., Breytenbach, J. H. J., Steyn, C., De Bruyn, R., and Maree, H. J. 2021. Towards the validation of high-throughput sequencing (HTS) for routine plant virus diagnostics: Measurement of variation linked to HTS detection of citrus viruses and viroids. Virol. J. 18:61.
Brady, C. L., Venter, S. N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J., and Coutinho, T. A. 2009. *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. Int. J. Syst. Evol. Microbiol. 59:2339-2345.
Brown, K. 2001. Florida fights to stop citrus canker. Science 292:2275-2276.
Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., and Zemskaya, T. I. 2019. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci Data. 6:190007.
Cai, W., Nunziata, S., Rascoe, J., and Stulberg, M. J. 2019. SureSelect targeted enrichment, a new cost effective method for the whole genome sequencing of *Candidatus* Liberibacter asiaticus. Sci Rep. 9:18962.
Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13:581-583.
Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., and Tiedje, J. M. 2014. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42:D633-D642.
Cordeiro, A. B., Sugahara, V. H., Stein, B., and Leite, R. P., Jr. 2014. Evaluation by PCR of *Xylella fastidiosa* subsp. *pauca* transmission through citrus seeds with special emphasis on lemons (*Citrus limon* (L.) Burm. f). Crop Prot. 62:86-92.
Edgar, R. C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5:113.
Fowkes, A. R., McGreig, S., Pufal, H., Duffy, S., Howard, B., Adams, I. P., Macarthur, R., Weekes, R., and Fox, A. 2021. Integrating high throughput sequencing into survey design reveals turnip yellows virus and soybean dwarf virus in pea (*Pisum sativum*) in the United Kingdom. Viruses 13:2530.
Giangacomo, C., Mohseni, M., Kovar, L., and Wallace, J. G. 2021. Comparing DNA extraction and 16S rRNA gene amplification methods for plant-associated bacterial communities. Phytobiomes J. 5:190-201.
Gołębiewski, M., and Tretyn, A. 2020. Generating amplicon reads for microbial community assessment with next-generation sequencing. J. Appl. Microbiol. 128:330-354.
Harrington, T. C., Yun, H. Y., Lu, S.-S., Goto, H., Aghayeva, D. N., and Fraedrich, S. W. 2011. Isolations from the redbay ambrosia beetle, *Xyleborus glabratus*, confirm that the laurel wilt pathogen, *Raffaelea lauricola*, originated in Asia. Mycologia 103:1028-1036.
Hodgetts, J., Boonham, N., Mumford, R., and Dickinson, M. 2009. Panel of 23S rRNA gene-based real-time PCR assays for improved universal and group-specific detection of phytoplasmas. Appl. Environ. Microbiol. 75:2945-2950.
Illumina. 2013. 16S metagenomic sequencing library preparation. Part #15044223. Rev. B.
Johnson, M. A., Liu, H., Bush, E., Sharma, P., Yang, S., Mazloom, R., Heath, L. S., Nita, M., Li, S., and Vinatzer, B. A. 2022. Investigating plant disease outbreaks with long-read metagenomics: Sensitive detection and highly resolved phylogenetic reconstruction applied to *Xylella fastidiosa*. Microb. Genom. 8:mgen000822.
Kini, K., Agnimonhan, R., Dossa, R., Silué, D., and Koebnik, R. 2021. Genomics-informed multiplex PCR scheme for rapid identification of rice-associated bacteria of the genus *Pantoea*. Plant Dis. 105: 2389-2394.
Laforest-Lapointe, I., Messier, C., and Kembel, S. W. 2016. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. Microbiome 4:27.
Lee, I.-M., Bottner, K. D., and Sun, M. 2009. An emerging potato purple top disease associated with a new 16SrIII group phytoplasma in Montana. Plant Dis. 93:970.
Li, W., Teixeira, D. C., Hartung, J. S., Huang, Q., Duan, Y., Zhou, L., Chen, J., Lin, H., Lopes, S., Ayres, A. J., and Levy, L. 2013. Development and systematic validation of qPCR assays for rapid and reliable differentiation of *Xylella fastidiosa* strains causing citrus variegated chlorosis. J. Microbiol. Methods 92:79-89.
Luna, E., Lang, J. M., McClung, A., Wamishe, Y., Jia, Y., and Leach, J. E. 2023. First report of rice bacterial leaf blight disease caused by *Pantoea ananatis* in the United States. Plant Dis. https://doi.org/10.1094/PDIS-08-22-2014-PDN
Malapi-Wight, M., Adhikari, B., Zhou, J., Hendrickson, L., Maroon-Lango, C. J., McFarland, C., Foster, J. A., and Hurtado-Gonzales, O. P. 2021. HTS-based diagnostics of sugarcane viruses: Seasonal variation and its implications for accurate detection. Viruses 13:1627.
Mechan Llontop, M. E., Sharma, P., Aguilera Flores, M., Yang, S., Pollok, J., Tian, L., Huang, C., Rideout, S., Heath, L. S., Li, S., and Vinatzer, B. A. 2020. Strain-level identification of bacterial tomato pathogens directly from metagenomic sequences. Phytopathology 110:768-779.
Newberry, E., Bhandari, R., Kemble, J., Sikora, E., and Potnis, N. 2020. Genome-resolved metagenomics to study co-occurrence patterns and intraspecific heterogeneity among plant pathogen metapopulations. Environ. Microbiol. 22:2693-2708.
Ogier, J.-C., Pagès, S., Galan, M., Barret, M., and Gaudriault, S. 2019. *rpoB*, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. BMC Microbiol. 19:171.
Poirier, S., Rué, O., Peguilhan, R., Coeuret, G., Zagorec, M., Champomier-Vergès, M.-C., Loux, V., and Chaillou, S. 2018. Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using *gyrB* amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing. PLoS One 13:e0204629.
Polston, J. E., McGovern, R. J., and Brown, L. G. 1999. Introduction of tomato yellow leaf curl virus in Florida and implications for the spread of this and other geminiviruses of tomato. Plant Dis. 83:984-988.
Price, M. N., Dehal, P. S., and Arkin, A. P. 2010. FastTree 2 – Approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.
Rochefort, A., Simonin, M., Marais, C., Guillerm-Erckelboudt, A.-Y., Barret, M., and Sarniguet, A. 2021. Transmission of seed and soil microbiota to seedling. mSystems 6:e00446-21.
Román-Reyna, V., Dupas, E., Cesbron, S., Marchi, G., Campigli, S., Hansen, M. A., Bush, E., Prarat, M., Shiplett, K., Ivey, M. L. L., Pierzynski, J., Miller, S. A., Peduto Hand, F., Jacques, M.-A., and Jacobs, J. M. 2021. Metagenomic sequencing for identification of *Xylella fastidiosa* from leaf samples. mSystems 6:e00591-21.
Saad, N., Olmstead, J. W., Varsani, A., Polston, J. E., Jones, J. B., Folimonova, S. Y., and Harmon, P. F. 2021. Discovery of known and novel viruses in wild and cultivated blueberry in Florida through viral metagenomic approaches. Viruses 13:1165.
Sikes, B. A., Bufford, J. L., Hulme, P. E., Cooper, J. A., Johnston, P. R., and Duncan, R. P. 2018. Import volumes and biosecurity interventions shape the arrival rate of fungal pathogens. PLoS Biol. 16:e2006025.
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.
Stulberg, M. J., and Huang, Q. 2015. A TaqMan-based multiplex qPCR assay and DNA extraction method for phylotype IIB sequevars 1&2 (Select Agent) strains of *Ralstonia solanacearum*. PLoS One 10:e0139637.
Stulberg, M. J., Rascoe, J., Li, W., Yan, Z., Nakhla, M. K., and Huang, Q. 2016. Development and comparison of TaqMan-based real-time PCR assays for detection and differentiation of *Ralstonia solanacearum* strains. Curr. Microbiol. 73:542-549.
Tufail, M. R., Fayyaz, A., Sarfraz, S., Waqas, M., Ahmad, M. Z., Hameed, A., Masood, A., Abdin, Z. U., Amrao, L., Sahi, S. T., and Abbas, A. 2020. First report of Pantoea leaf blight of cotton caused by *Pantoea* spp. in Punjab, Pakistan. Plant Dis. 104:1534.
Zheng, Z., Xu, M., Bao, M., Wu, F., Chen, J., and Deng, X. 2016. Unusual five copies and dual forms of *nrdB* in "*Candidatus* Liberibacter asiaticus": Biological implications and PCR detection application. Sci Rep. 6:39020.