

MBPD: A multiple bacterial pathogen detection pipeline for One Health practices

Xinrun Yang  | Gaofei Jiang  | Yaozhong Zhang  | Ningqi Wang  |
Yuling Zhang  | Xiaofang Wang  | Fang-Jie Zhao  | Yangchun Xu  |
Qirong Shen  | Zhong Wei 

Laboratory of Bio-Interactions and Crop Health, Jiangsu Provincial Key Laboratory for Organic Solid Waste Utilization, Joint International Research Laboratory of Soil Health, Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, National Engineering Research Center for Organic-Based Fertilizers, College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing, China

Correspondence

Gaofei Jiang and Zhong Wei, Laboratory of Bio-Interactions and Crop Health, Jiangsu Provincial Key Laboratory for Organic Solid Waste Utilization, Joint International Research Laboratory of Soil Health, Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, National Engineering Research Center for Organic-Based Fertilizers, College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China.
Email: gjiang@njau.edu.cn and weizhong@njau.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 42090062, 42090064, 31972504, 42007038, 3217018; National Key Research and Development Program of China, Grant/Award Number: 2022YFC3501501; Fundamental Research Funds for the Central Universities, Grant/Award Numbers: XUEKEN2022025, XUEKEN2022026; China National Tobacco Corporation, Grant/Award Number: 110202101047 (LS-07); Jiangxi Branch of China National Tobacco Corporation, Grant/Award Number: 2021.01.010

Abstract

Bacterial pathogens are one of the major threats to biosafety and environmental health, and advanced assessment is a prerequisite to combating bacterial pathogens. Currently, 16S rRNA gene sequencing is efficient in the open-view detection of bacterial pathogens. However, the taxonomic resolution and applicability of this method are limited by the domain-specific pathogen database, taxonomic profiling method, and sequencing target of 16S variable regions. Here, we present a pipeline of multiple bacterial pathogen detection (MBPD) to identify the animal, plant, and zoonotic pathogens. MBPD is based on a large, curated database of the full-length 16S genes of 1986 reported bacterial pathogen species covering 72,685 sequences. In silico comparison allowed MBPD to provide the appropriate similarity threshold for both full-length and variable-region sequencing platforms, while the subregion of V3–V4 (mean: 88.37%, accuracy rate compared to V1–V9) outperformed other variable regions in pathogen identification compared to full-length sequencing. Benchmarking on real data sets suggested the superiority of MBPD in a broader range of pathogen detections compared with other methods, including 16SPIP and MIP. Beyond detecting the known causal agent of animal, human, and plant diseases, MBPD is capable of identifying cocontaminating pathogens from biological and environmental samples. Overall, we provide a MBPD pipeline for agricultural, veterinary, medical, and environmental monitoring to achieve One Health.

KEY WORDS

bacterial pathogen detection, 16S rRNA gene sequencing, One Health

Xinrun Yang and Gaofei Jiang contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *iMeta* published by John Wiley & Sons Australia, Ltd on behalf of *iMeta* Science.

Highlights

- Multiple bacterial pathogen detection (MBPD) is a 16S-based pipeline for detecting animal, plant, and zoonotic bacterial pathogens.
- The curated reference database of MBPD contains 72,685 sequences from 1986 known pathogen species.
- MBPD provides appropriate thresholds for taxonomic profiling of common 16S variable regions.
- MBPD outperforms other methods in detecting both causal and coinfecting pathogens.

INTRODUCTION

Pathogenic infections are a formidable challenge in One Health, which is an approach that recognizes that the health of people is closely connected to the health of animals, plants, and our shared environment [1]. During the past two decades, approximately 15 million global deaths per year have been attributed to infectious diseases [2], and approximately 30% of global food production is lost to plant diseases every year [3]. Today, the serious damage to humans and the global economy caused by COVID-19, the risk of increased zoonotic host diversity, and the pandemic of plant pathogenic bacteria indicate the importance of whole-habitat (human-animal-plant-environment) detection of pathogens and surveillance for disease management and ecological health [4–6]. In addition, coinfection is common in both natural and agricultural environments [7]. The main problem is that coinfecting pathogens interact synergistically with each other, with the presence of one enhancing the abundance and/or virulence of the other [8]. However, there is no reliable means of detecting coinfection [9, 10]. Many national/international organizations, such as the National Center for Biotechnology Information (NCBI), World Health Organization, and International Society of Plant Pathology Committee on the Taxonomy of Plant Pathogenic Bacteria (ISPP-CTPPB), have been gathering information on pathogens, but the information is fragmentary and unable to satisfy the requirement of a holistic view of pathogen infections/contamination in biological and environmental samples for implementation of the One Health approach [11].

Pathogen detection is mainly categorized into culture-dependent and culture-independent approaches [12]. The former approaches relying on cultivation and biochemical testing are classic experiments for pathogen detection, which provide critical information, including the ability to satisfy Koch's postulates, drug susceptibility, and biochemical tests, but are highly time-consuming and

limited by the culturability of pathogens [13]. Culture-independent approaches include PCR- and high-throughput sequencing-based methods. PCR or probe microarrays reduce time in pathogen identification but are overreliant on the specificity and number of primers and prone to false-positive rates in pathogen detection [14, 15]. Common high-throughput sequencing-based methods involve 16S/ITS/18S/*rpoB/cpn60* amplicon sequencing and metagenomic sequencing [16]. Given the reported pathogen sequences, sequencing methods are capable of identifying various pathogens within almost all types of microorganisms in the samples [17]. Compared with culture-dependent and PCR methods, the main benefit of sequencing methods is being high throughput, but the matching analysis pipelines need to be improved to achieve high sensitivity [18].

16S rRNA gene sequencing is widely used in bacterial pathogen detection in animal, food, water, soil, and plant samples [16, 19–23]. However, the scope and sensitivity of pathogen detection using 16S sequencing are highly dependent on the sequencing methods, hypervariable regions sequenced, pathogen database, and analytic pipelines [9, 18]. Short-read sequencing is high-throughput and cost-effective but has limitations in read length [18]. Long-read sequencing is able to generate long reads for high-resolution taxonomy but has a limitation of sequencing error rates of approximately 13% [17]. Additionally, the composition and abundance of pathogens can be affected by the use of different bioinformatics tools [12]. For example, the common approaches to data processing include operational taxonomic unit (OTU) clustering or amplicon sequence variant (ASV) approaches [24]. Compared with the OTU-based approach, the ASV-based approach has the advantages of fewer sequencing errors and more accurate taxonomic resolution [25, 26]. Moreover, the outcome of pathogen detection fluctuates depending on the reference database, sequencing region, and similarity threshold [27, 28]. Several pipelines for pathogen identification have been developed based on 16S rRNA amplicon data, such as

16SPIP [29], MIP [30], and Krishna [31], but they only target pathogens in relation to human infection or water quality [29–31]. The widespread distribution and transmission of pathogens among multiple hosts and environments require a tool for systematic and rapid detection of pathogens suitable for all habitats to provide a basis for One Health diagnosis and therapy of multiple pathogens that coexist in humans, animals, and plants [1, 32].

In this study, we developed a bacterial pathogen detection pipeline, multiple bacterial pathogen detection (MBPD), to accurately detect pathogens based on 16S rRNA gene sequencing. To comprehensively identify pathogens in samples and meet the demand for accurate pathogen identification under different variable regions in 16S rRNA sequencing, we constructed a curated full-length 16S database matching the DADA2 pipeline [25] of the ASV-based approach for data processing and evaluated the optimal similarity based on in silico experiments. MBPD is able to detect a broad range of animal, plant, and zoonotic pathogens, which is essential for food safety, epidemic prevention, disease diagnosis, and environmental monitoring. MBPD is publicly available at <https://github.com/LorMeBioAI/MBPD>.

RESULTS

Overview of MBPD

We developed the MBPD approach for holistic pathogen detection based on 16S sequencing. The input of MBPD is 16S rRNA amplicon data without primers and barcodes, and the output is an ASV table with pathogen taxonomy. The pipeline consists of the following steps: (1) building the MBPD database based on the alignment between collected pathogen information and the Silva 16S database; (2) use of the MBPD database in in silico and sequence-based experiments to explore the optimal similarity and the accuracy of species/strain annotation using the UCLUST algorithm for the different variable regions of 16S; (3) use of the DADA2 tool for the conversion from 16S rRNA amplicon data to ASV sequences; (4) alignment of ASV sequences with the MBPD database using an appropriate similarity threshold based on various sequencing regions, where 90% is recommended for the V4 and V1–V2 regions, and 80% is recommended for other variable regions (Figure 1).

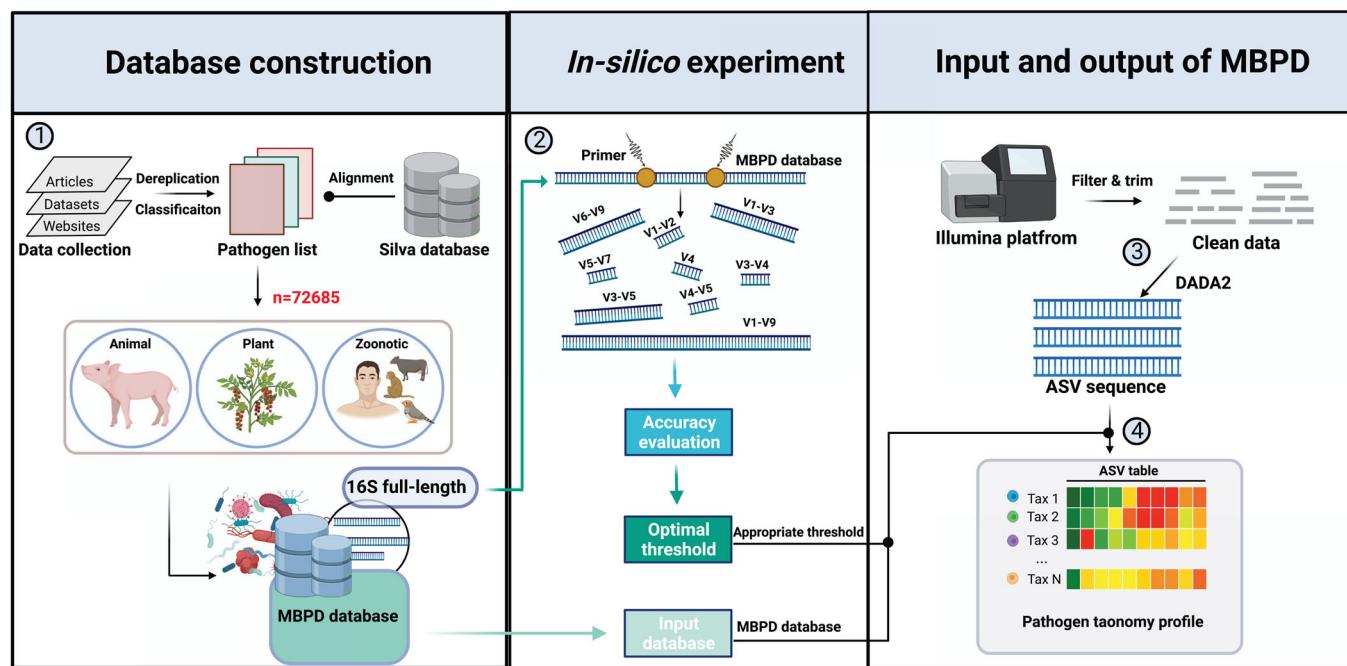


FIGURE 1 MBPD workflow for open-view bacterial pathogen detection based on 16S rRNA-encoding gene sequencing. Bacterial pathogens causing animal, plant, and zoonotic diseases were first collected and curated from publicly available literature, databases, and web resources to construct the multiple bacterial pathogen database (MBPD) with a total of 72,685 full-length sequences of 16S. Then, the accuracy and appropriate threshold of each variable region of 16S were evaluated through an in silico experiment. The clean data from 16S sequencing of bio/eco samples were subjected to the DADA2 pipeline using the MBPD database to obtain the amplicon sequence variant (ASV) sequences. The feature table of the ASV sequences was assigned using the UCLUST algorithm with an in silico experiment to optimize the cutoffs of similarity for the different variable regions of 16S. MBPD, multiple bacterial pathogen detection.

A large curated pathogen reference database of 72,685 16S sequences for MBPD

To construct a comprehensive database of bacterial pathogens causing animal, plant, and zoonotic diseases, we collected a total of 1986 pathogen species from public agencies, databases, and papers (Supporting Information: Tables S1 and S2). In general, they mainly belonged to four phyla, namely, Proteobacteria (758), Firmicutes (558), Actinobacteria (388), and Bacteroidetes (123) (Figure 2A). Then, we extracted sequences corresponding to those pathogens at the species or strain level from the Silva database, called the MBPD database, to improve data compatibility and better identify pathogen diversity. Briefly, the MBPD database is primarily composed of Proteobacteria (39,582), Firmicutes (24,284), Actinobacteria (4660), and Bacteroidetes (1449) at the phylum level (Figure 2B) and *Bacillus* (6945), *Escherichia-Shigella* (6187), *Salmonella* (5251), and *Pseudomonas* (4360) at

the genus level (Figure 2C). In addition, a high number of these pathogens originated from the “zoonotic” group (35,493 out of 72,685, 48.9%) and “animal” group (33,832 out of 72,685, 46.5%) compared to the “plant” group (3360 out of 72,685, 4.6%) (Figure 2B and Supporting Information: Figure S4). Details of the sequence number in each species are presented in Supporting Information: Table S5.

In silico experiments under different thresholds for sequencing platforms and amplicon targets

The 16S rRNA gene comprises nine variable regions interspersed throughout the highly conserved 16S sequence (Figure 3A). To provide better detection efficiency for various sequencing platforms and amplicon targets, we conducted an in silico experiment to

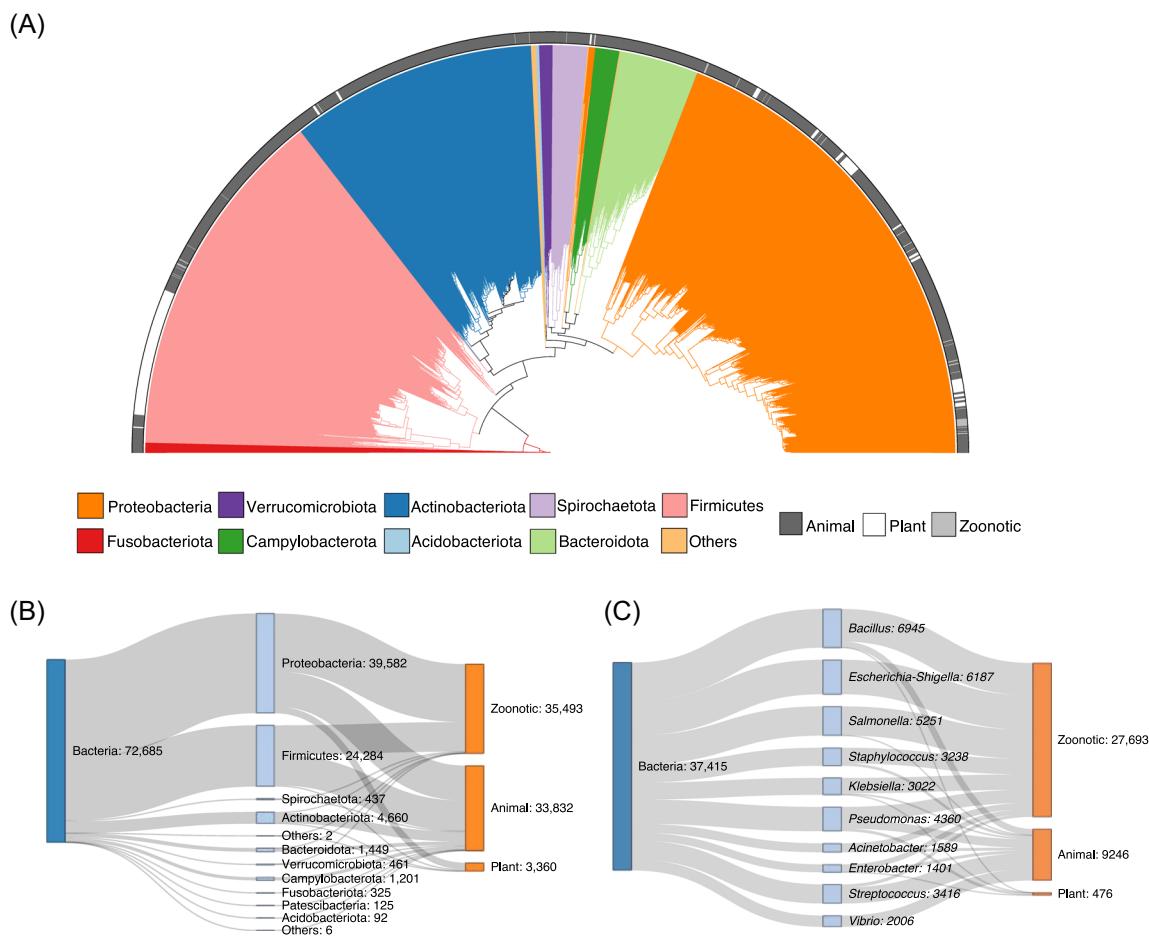


FIGURE 2 The curated bacterial pathogen database of MBPD. (A) Phylogenetic tree based on the 16S rRNA gene sequences of 1986 reference pathogens. Colors of branches represent the corresponding phyla, and the outer ring denotes animal, plant, and zoonotic types of pathogens. (B) Basic taxonomic composition of the MBPD database at the phylum level. Only the top 10 phyla are shown in the figure, and the remaining phyla are merged as others. (C) Basic taxonomic composition of the MBPD database at the genus level. Only the top 10 genera are shown in the figure.

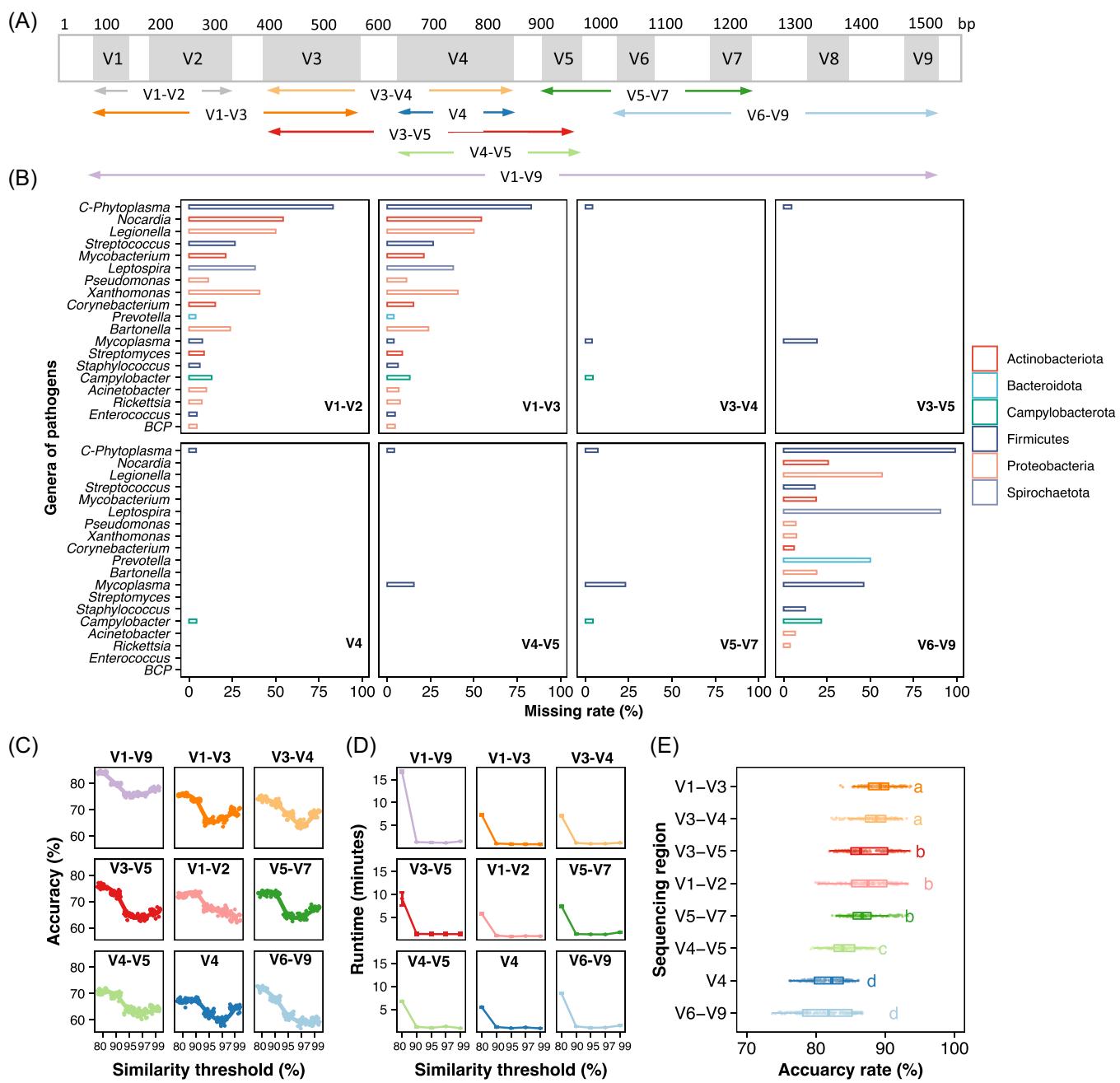


FIGURE 3 In silico comparison of 16S rRNA variable regions. (A) Common sequencing target of variable regions in Illumina and full-length sequencing. The 16S full-length database of MBPD was trimmed and generated in silico amplicons for different subregions based on the location of PCR primers commonly used in microbiome studies. Then, 10,000 sequences were extracted and repeated 30 times for pathogen alignment. (B) Missing rate of pathogens across bacterial genera under in silico experiments. Various facets denote the sequencing target of 16S variable regions (V). Colors denote the missing pathogen at the phylum level. *C-Phytoplasma* and *BCP* refer to *Candidatus Phytoplasma* and *Burkholderia-Caballeronia-Paraburkholderia*, respectively. (C) Accuracy rate of taxonomy assignment for in silico amplicons of the MBPD database with varying similarity thresholds (80%, 90%, 95%, 97%, and 99%) for different variable regions of 16S. (D) Runtime of taxonomy assignment for in silico amplicons with varying similarity thresholds for different variable regions of 16S. (E) Accuracy rate of the subregion in pathogen detection compared to the V1–V9 region. Different lowercase letters indicate significant differences in disease incidence across varieties (HSD post hoc test: $p < 0.05$). HSD, honest significant difference.

compare the missing rate, accuracy rate, and runtime under different thresholds (Figure 3, Supporting Information: Figures S1 and S2). We found that subregions differed substantially in the extent to which they could match the PCR primer. The V1–V2, V1–V3, and V6–V9 regions performed worst, with 29.3%–33.2% of in silico amplicons unmatched (Supporting Information: Figure S2A), mainly belonging to three phyla (Proteobacteria, Firmicutes, and Actinobacteria) (Figure 3B and Supporting Information: Figure S1). Meanwhile, certain pathogen genera were probably unstably detected by the sequencing of variable regions with a high missing rate by in silico amplicons (Figure 3B). For example, phytopathogenic *Candidatus Phytoplasma* spp. is highly missing in the V1–V2, V1–V3, and V6–V9 regions, while human pathogenic *Mycoplasma* spp. is not well assigned in V3–V5, V4–V5, and V5–V7 (Figure 3B). Next, various sequencing regions showed a distinct discrepancy in the accuracy of pathogen classification (Supporting Information: Figure S2B). The V1–V9 region (full-length) performed best (analysis of variance [ANOVA]: $F_{8,1341} = 183.9$, $p < 0.001$, Supporting Information: Figure S2B), and the subregions of V1–V3 and V3–V4 were second to V1–V9 after eliminating the difference in the similarity threshold at the sequence level (ANOVA: $F_{7,1192} = 174.7$, $p < 0.001$, Figure 3E). Above all, the choice of similarity threshold substantially affected the accuracy of pathogen identification and runtime. In terms of accuracy, most subregions performed well at the 80% similarity threshold, but the V4 and V1–V2 regions performed best at the 90% threshold (Figure 3C). Although the similarity threshold of 80% is recommended for both full-length and V3–V4 sequencing approaches to obtain the optimal accuracy of bacterial pathogen detection, it took more processing time at the 80% similarity threshold than at other thresholds (Figure 3D).

MBPD outperformed other approaches in pathogen detection

We selected 16S data of 50 patient skin samples sourced from Lax et al. [33] to compare MBPD with two published similar pipelines (16SPIP [29] and MIP [30]) using the same data set. For MBPD, we used the similarity threshold of 90% in the process of pathogen alignment based on the results of the in silico experiment (Figure 3C) because they all amplified the V4 region of 16S. In this study, MBPD contained the largest number of identified pathogens among all the pipelines, with the number of detected pathogens reaching 310, and contained most of the other two, especially in MIP (Figure 4A). According to the evaluation of processing time, MBPD is as time-effective as the MIP pipeline in detecting pathogenic species with the same sample sizes (Figure 4B). Next, we compared the relative abundance of dominant genera (the six most abundant genera) with MIP because 16SPIP can be applied to pathogen identification but not abundance. We identified the dominant genus *Staphylococcus*, which was also detected in MIP, and detected genera not identified in MIP, such as *Corynebacterium* and *Finegoldia* (Supporting Information: Figure S5A).

MBPD accurately identified the dominant and coinfecting pathogens in real samples

We tested MBPD on paired samples between healthy and diseased individuals for human periodontitis, white shrimp disease, and plant bacterial wilt to further evaluate the performance of MBPD in real environmental samples, which are important sectors of One Health (human, animal, and environment). We used the threshold of 90% for pathogen alignment of all samples because they all amplified the V4 region of the 16S rRNA

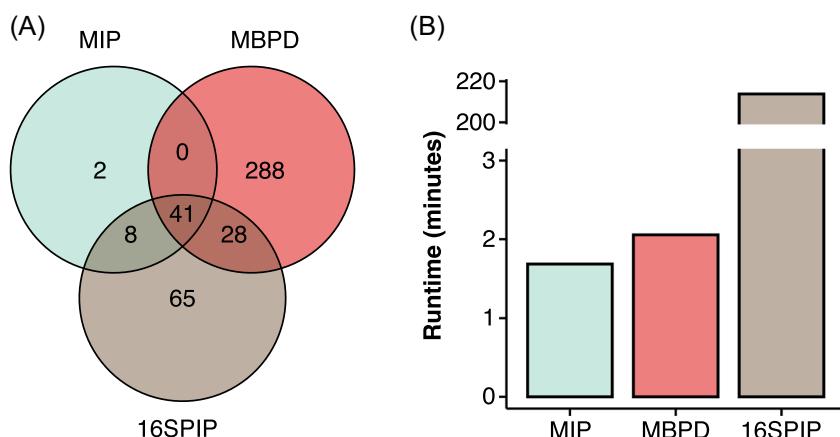


FIGURE 4 Performance of MBPD, 16SPIP, and MIP in bacterial pathogen detection.

(A) Venn diagram displaying the shared and specific species of pathogens detected by MBPD, 16SPIP, and MIP. (B) Comparison of runtime among MBPD, MIP, and 16SPIP. Runtime in min. The colors green, red, and brown denote MIP, MBPD, and 16SPIP, respectively.

gene. The abundance of the pathogen community and causal agent of diseases was significantly higher in diseased samples than in healthy samples (Student's *t*-test, $p < 0.001$, Figure 5A and Supporting Information: Figure S3A–C), but the evenness of the pathogen community in healthy samples was significantly higher than that in diseased samples (Supporting Information: Figure S3D–F). In addition, MBPD was able to identify other potential coinfecting pathogens (Figure 5B–D). For example, the highly enriched genera in diseased samples of humans and animals (*Treponema* and *Photobacterium*; $p < 0.001$ by Student's *t*-test) could be isolated from the same environment as the causal agent of diseases and seriously threaten the host.

DISCUSSION

We present MBPD, a method to comprehensively identify bacterial pathogens from bio/eco-samples using 16S rRNA gene sequencing, focusing on assessing pathogenic contamination in environmental and biological samples. To the best of our knowledge, MBPD represents a One Health-oriented pathogen detection approach fueled by a large, current reference pathogen database. Key advances underlying our pipeline include (i) a relatively complete database convenient for detection across a broad range of pathogenic bacteria infecting animals and plants and zoonoses; (ii) appropriate thresholds provided for current common 16S amplicon regions for sequence alignment;

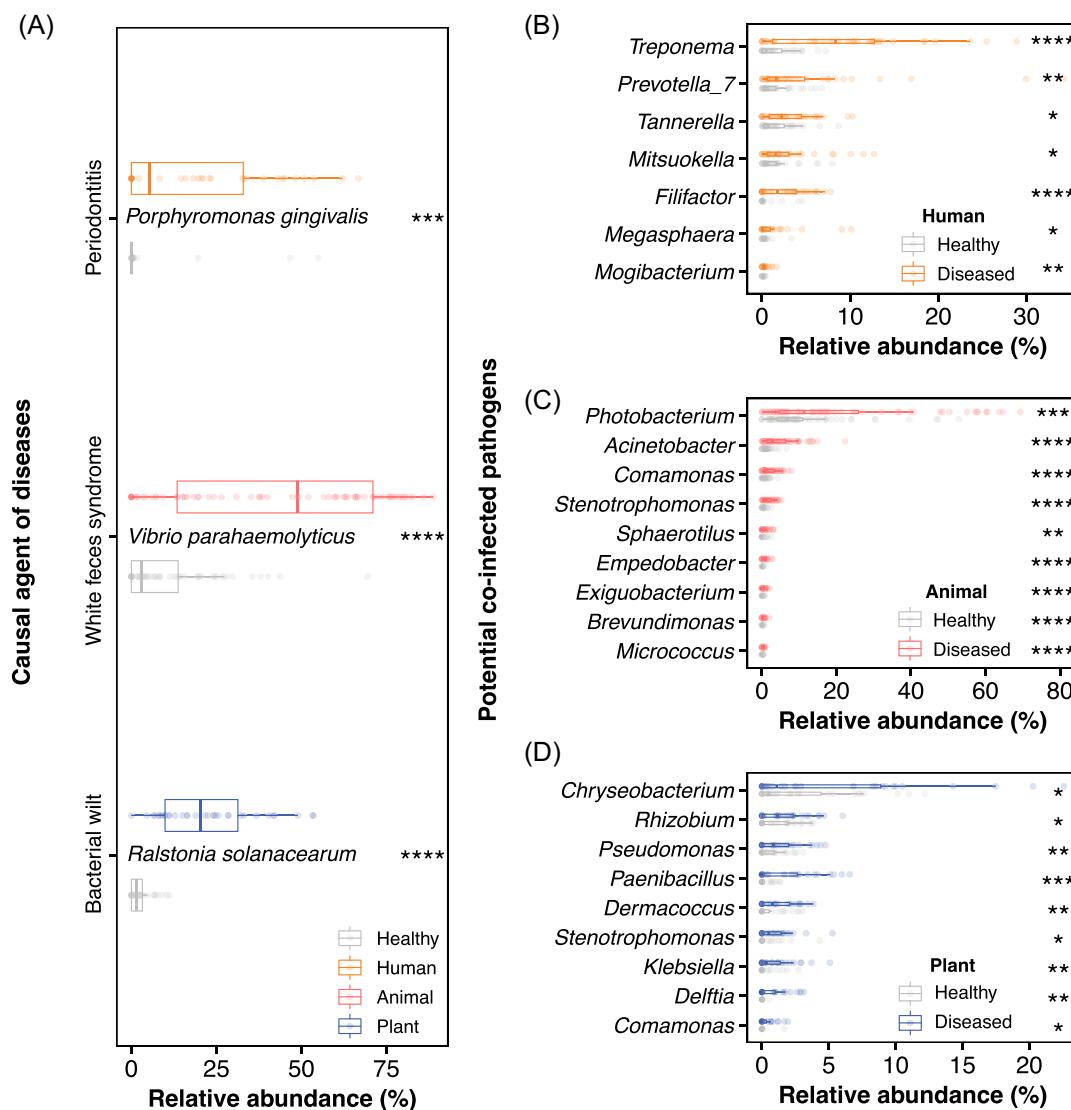


FIGURE 5 Pathogen detection in the healthy and diseased samples. Differences in the relative abundance of causal agents of diseases in human, animal, and plant rhizosphere samples (A). Relative abundances of other potential genera were enriched in the diseased group between human (B), animal (C), and plant rhizosphere (D) samples. Pairwise Student's *t*-test was used for statistical analyses (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).

(iii) high performance in the identification and quantification of pathogens; (iv) accurate identification for both dominant pathogens and their cocontaminating pathogens from biological and environmental samples; and (v) an open-source tool suitable for constant updating of pathogen databases and processes.

The performance of pathogen identification is affected by the reference database, sequencing region, and similarity threshold [27, 34]. Small databases limit detected pathogen richness [28], and diverse sequencing regions lead to differences in species [27]. Here, we introduced a reference database for pathogen identification based on the Silva database because Silva is comprehensive and updated in a timely manner [35]. However, the optimal similarity threshold under different sequencing regions is still unknown. We conducted an *in silico* comparison of 16S rRNA gene sequencing under different thresholds [27] and chose UCLUST [36] as the tool for pathogen alignment because UCLUST achieved significantly better performance than all other methods for 16S rRNA gene classifications at the species level [34]. Our assessment showed that 16S full-length sequencing at the 80% threshold performed best using MBPD, but variable regions will likely reduce the pathogen richness of a microbiome sample. In addition, considering the wider use of the Illumina sequencing platform (lower cost and production of short sequences), we also provided the optimal threshold under various common sequencing regions (V1–V2, V1–V3, V3–V4, V3–V5, V4, V4–V5, V5–V7, and V6–V9) for the best detection results. Next, MBPD enables the detection of cocontaminating pathogens from samples. For example, there are many cross-infections occurring in a hostile environment [37, 38], but specific pathogens are not clear, or the densities of other pathogens will increase after treatment with the current pathogen. MBPD can also be used for a deeper understanding of pathogens, for example, to elucidate the pathogen diversity in samples and its response to environmental changes.

Although several tools are available to identify pathogens using 16S sequences [29, 30], a benchmark study to assess the accuracy of these pipelines were needed. We downloaded 16S rRNA amplicon-based sequencing data of 50 patient skin samples from Lax et al. [33] and tested them with these tools. Here, we showed that MBPD might outperform other existing tools when processing the same data: MBPD contains the highest pathogen richness, and the dominant pathogens are consistent with those from other tools, which may result from the more comprehensive database providing a wider detection of pathogens [9]. In addition, we looked at the accuracy of MBPD using 16S rRNA amplicon data of humans, animals, and plants as well as diseased and healthy samples. Each type of sample

was adequate for detecting the causal agent of diseases and identifying the potential pathogens.

Although MBPD showed high performance in pathogen detection, there are still limitations and challenges in 16S sequencing diagnosis. First, short-read sequencing is limited in taxonomical resolution [12, 27, 39, 40]. The performance of the 16S subregion reached a maximum relative accuracy of 90% at the species/strain level compared to 16S full-length sequencing (Figure 3D). Different strains at the species level or closely related species are occasionally unstably detected using short-read sequencing [12, 27]. For example, *Staphylococcus* spp. was abundantly recognized as innocuous *S. epidermidis* by MIP but identified as pathogenic *S. aureus* by MBPD (Supporting Information: Figure S5B,C). To date, higher taxonomical resolution in differentiating bacterial strains at the subgenus level has been achieved by long-read sequencing (PacBio [27] and MinION™ [41]) and protein-coding taxonomic markers (*rpoB* [42] and *cpn60* [43]), but the target does not perform well in discerning microorganism below the genus level. Second, pathogen identification has primer bias. Studies have indicated that V4 is unstable in identifying enteric pathogens and phytopathogenic *Robbsia* spp. using OTU clustering [40] or the Greengenes database [39]. The current MBPD shows a decent resolution at the species/strain level and accurately detects these two pathogens using the ASV approach with our curated pathogen database (Figure 4A and Supporting Information: Figure S1). This suggested analytical method and database are critical for taxonomic profiling [25, 28]. However, some pathogen species, such as phytopathogenic *Candidatus Phytoplasma* spp., are still difficult to detect (missing rate >75%, Supporting Information: Figure S1) due to the limited strain resolution of 16S for identification [40]. Thus, the identification of *Phytoplasmas* spp. may rely on some protein-coding taxonomic markers (*rpoB* [42] and *cpn60* [43]) or PCR-based sequencing approaches applied to multiple genes other than 16S [12]. Finally, sequencing approaches lack direct evidence of lesions. Although the composition and risk of pathogen contamination have been assessed by amplicon and metagenome sequencing approaches [9, 18], the ground truth of causal pathogenic agents, lesions, and disease types require further experimental validations and antilytic method advances. For example, some pipelines, such as MIP, were developed to understand the relationship among pathogens, disease types, and human lesions [30, 44].

CONCLUSION

We developed a 16S-based pipeline, MBPD, for the holistic detection of animal, plant, and zoonotic pathogens. MBPD contains the most comprehensive reference

database for pathogen identification and quantification with feasible updating. This pipeline outperforms the existing tools for detecting bacterial pathogens and provides the optimal similarity threshold for pathogen alignment from varying biome samples. We anticipate a broad application of MBPD for assessment and diagnosis in the clinic, agriculture, fisheries, and veterinary medicine in the current One Health era.

METHODS

MBPD database construction

We first collected information on animal, plant, and zoonotic pathogens from various studies and databases, such as 16SPP [29], FAPROTAX [45], and Bull *et al.* [46], [47], [48] (for details, see Supporting Information: Table S1). Based on the species level, a pathogenic reference database containing 72,685 sequences was obtained by alignment with the Silva 138.1 SSU Ref NR 99 bacterial database, which was developed with a 99% identity to remove redundant sequences [35], and a label on pathogenic types was added to the basic taxonomy of the Silva database (Supporting Information: Table S4, type). For unclassified pathogens, we further searched for relatively accurate taxonomic information in the NCBI Taxonomy Database (<https://www.ncbi.nlm.nih.gov/taxonomy>). However, considering that the species and strain levels in the Silva database are not well distinguished [49], we added a label for the corrected species level (Supporting Information: Table S4, species_manual). To construct and explore the phylogenetic relationships of pathogens, the sequences of all species (1986) were aligned using MAFFT [50], and a de novo phylogenetic tree was constructed using FastTree2 based on the maximum-likelihood method [51]. The phylogenetic tree was further visualized using the iTOL web tool (<https://itol.embl.de>).

In silico comparison of 16S rRNA gene sequencing under different thresholds

The in silico analysis was carried out separately using the MBPD database. In silico amplicons demarcating different subregions of the 16S gene were generated by trimming regions defined by established primer sets (Supporting Information: Table S2) using the USEARCH *search_pcr2* command [36]. Sequences were discarded if one or more variable regions could not be matched to a primer pair [27]. To determine the taxonomic resolution afforded by different variable regions and thresholds, 10,000 sequences were randomly selected using SeqKit [52] from each in silico

amplicon, and classification was repeated 30 times against the MBPD database using the UCLUST algorithm [36, 53] with a variety of similarity thresholds (80%, 90%, 95%, 97%, and 99%). The accuracy rate, calculated as Accuracy rate = $100 \times (\text{accuracy}_{\text{full-length}} - \text{accuracy}_{\text{subregions}})/\text{accuracy}_{\text{full-length}}$, was employed to explore the best performance of various subregions in pathogen identification. The missing rate, calculated as Missing rate = $100 \times (\text{total number}_{\text{phyla}} - \text{miss number}_{\text{phyla}})/\text{total number}_{\text{phyla}}$, was used to reveal the proportion of sequences for each variable region that could not be matched in the in silico experiment.

Data processing

Sequences with primers and barcodes removed were processed using the DADA2 pipeline [25], which is designed to obtain sequences with a single-nucleotide difference, known as ASVs. Trimming and filtering were performed on paired reads with a maximum of one expected error per read (maxEE = 1). After merging paired reads and chimera filtering, the phylogenetic affiliation of each 16S rRNA gene sequence was analyzed by the UCLUST algorithm [36, 53] against the MBPD database using the appropriate threshold based on the results of in silico experiments.

Amplicon sequencing data for testing the performance of MBPD

To compare the performance of MBPD with that of 16SPP and MIP, we randomly picked 50 patient skin samples from projects [33] targeting the V4 region of 16S for taxonomic profiling, which had been previously tested on the MIP pipeline [30]. In addition, three published V4 regions of 16S data from human, animal, and plant samples, comprising paired healthy versus diseased samples, were downloaded from the NCBI according to accession number: periodontitis (pathogen: *Porphyromonas gingivalis*) [20], white shrimp disease (*Vibrio parahaemolyticus*) [19], and bacterial wilt (*Ralstonia solanacearum*) [54]. The raw sequences were first processed using fastp [55] to truncate low-quality reads, potential primers, and barcodes. Then, paired-end sequences were merged using the QIIME *join_paired_ends.py* command [53]. Finally, cleaned fastq files were processed using the MBPD pipeline.

Statistical analysis

Student's *t*-tests (two-sided) were used to test for statistical significance between pairs of samples, where

p values below 0.05 were considered statistically significant. ANOVA and Tukey's honest significant difference tests was used to determine the statistical significance of multiple comparisons. All statistical analyses were carried out using the R 4.1.2 program (www.r-project.org).

AUTHOR CONTRIBUTIONS

Conceptualization: Xinrun Yang, Gaofei Jiang, Zhong Wei. *Resources:* Gaofei Jiang, Zhong Wei. *Methodology:* Xinrun Yang, Gaofei Jiang, Zhong Wei. *Data curation:* Xinrun Yang, Gaofei Jiang, Yuling Zhang, Ningqi Wang. *Formal analysis:* Xinrun Yang, Gaofei Jiang, Yaozhong Zhang, Ningqi Wang, Yuling Zhang. *Funding acquisition:* Gaofei Jiang, Zhong Wei, Xiaofang Wang, Fang-Jie Zhao, Yangchun Xu, Qirong Shen. *Investigation:* Xinrun Yang, Gaofei Jiang, Yaozhong Zhang. *Project administration:* Xinrun Yang, Gaofei Jiang, Zhong Wei, Xiaofang Wang, Fang-Jie Zhao, Yangchun Xu, Qirong Shen. *Supervision:* Xinrun Yang, Gaofei Jiang, Zhong Wei, Yaozhong Zhang. *Software:* Xinrun Yang, Gaofei Jiang, Yaozhong Zhang, Zhong Wei, Yangchun Xu. *Visualization:* Xinrun Yang, Gaofei Jiang, Yaozhong Zhang. *Writing—original draft:* Xinrun Yang. *Writing—review & editing:* Gaofei Jiang, Yaozhong Zhang, Zhong Wei, Fang-Jie Zhao.

ACKNOWLEDGMENTS

We thank Chen Liu and Ye Tao for their suggestions. This research was financially supported by the National Natural Science Foundation of China (42090062, 42090064, 31972504, and 42007038), the National Key Research and Development Program of China (2022YFC3501501), and the Fundamental Research Funds for the Central Universities (XUEKEN2022025-XUEKEN2022026), and China National Tobacco Corporation (110202101047, LS-07), Jiangxi Branch of China National Tobacco Corporation (2021.01.010).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

MBPD is publicly available on GitHub (<https://github.com/LorMeBioAI/MBPD>). Supplementary materials (figures, tables, scripts, graphical abstract, slides, videos, Chinese translated version, and updated materials) may be found in the online DOI or iMeta Science <http://www.imeta.science/>.

ORCID

- Xinrun Yang  <http://orcid.org/0000-0002-6482-8734>
 Gaofei Jiang  <http://orcid.org/0000-0002-5331-739X>
 Yaozhong Zhang  <https://orcid.org/0000-0002-8024-6523>
 Ningqi Wang  <https://orcid.org/0000-0002-7788-4364>
 Yuling Zhang  <https://orcid.org/0000-0001-6352-570X>
 Xiaofang Wang  <https://orcid.org/0000-0001-5030-1686>
 Fang-Jie Zhao  <https://orcid.org/0000-0002-0164-169X>
 Yangchun Xu  <https://orcid.org/0000-0003-2740-5561>
 Qirong Shen  <https://orcid.org/0000-0002-5662-9620>
 Zhong Wei  <https://orcid.org/0000-0002-7967-4897>

REFERENCES

- Ghai, Ria R., Ryan M. Wallace, James C. Kile, Trevor R. Shoemaker, Antonio R. Vieira, Maria E. Negron, and Sean V. Shadomy, et al. 2022. "A Generalizable One Health Framework for the Control of Zoonotic Diseases." *Scientific Reports* 12: 8588. <https://doi.org/10.1038/s41598-022-12619-1>
- Fauci, Anthony S., and David M. Morens. 2012. "The Perpetual Challenge of Infectious Diseases." *New England Journal of Medicine* 366: 454–61. <https://doi.org/10.1056/NEJMra1108296>
- Savary, Serge, Laetitia Willocquet, Sarah Jane Pethybridge, Paul Esker, Neil McRoberts, and Andy Nelson. 2019. "The Global Burden of Pathogens and Pests on Major Food Crops." *Nature Ecology & Evolution* 3: 430–39. <https://doi.org/10.1038/s41559-018-0793-y>
- Gruetzmacher, Kim, William B. Karesh, John H. Amuasi, Adnan Arshad, Andrew Farlow, Sabine Gabrysch, Jens Jetzkowitz, et al. 2021. "The Berlin Principles on One Health—Bridging Global Health and Conservation." *Science of The Total Environment* 764: 142919. <https://doi.org/10.1016/j.scitotenv.2020.142919>
- Gibb, Rory, David W. Redding, Kai Qing Chin, Christl A. Donnelly, Tim M. Blackburn, Tim Newbold, and Kate E. Jones. 2020. "Zoonotic Host Diversity Increases in Human-Dominated Ecosystems." *Nature* 584: 398–402. <https://doi.org/10.1038/s41586-020-2562-8>
- Delgado-Baquerizo, Manuel, Carlos A. Guerra, Concha Cano-Díaz, Eleonora Egidi, Jun-Tao Wang, Nico Eisenhauer, Brajesh K. Singh, and Fernando T. Maestre. 2020. "The Proportion of Soil-Borne Pathogens Increases with Warming at the Global Scale." *Nature Climate Change* 10: 550–54. <https://doi.org/10.1038/s41558-020-0759-3>
- Tollenaere, Charlotte, Hanna Susi, and Anna-Liisa Laine. 2016. "Evolutionary and Epidemiological Implications of Multiple Infection in Plants." *Trends in Plant Science* 21: 80–90. <https://doi.org/10.1016/j.tplants.2015.10.014>
- Griffiths, Emily C., Amy B. Pedersen, Andy Fenton, and Owen L. Petchey. 2011. "The Nature and Consequences of Coinfection in Humans." *Journal of Infection* 63: 200–206. <https://doi.org/10.1016/j.jinf.2011.06.005>
- Han, Dongsheng, Ziyang Li, Rui Li, Ping Tan, Rui Zhang, and Jinming Li. 2019. "mNGS in Clinical Microbiology Laboratories: On the Road to Maturity." *Critical Reviews in*

- Microbiology* 45: 668–85. <https://doi.org/10.1080/1040841X.2019.1618933>
10. Sundin, George W., and Nian Wang. 2018. “Antibiotic Resistance in Plant-Pathogenic Bacteria.” *Annual Review of Phytopathology* 56: 161–80. <https://doi.org/10.1146/annurev-phyto-080417-045946>
11. Timme, Ruth E., William J. Wolfgang, Maria Balkey, Sai Laxmi Gubbala Venkata, Robyn Randolph, Marc Allard, and Errol Strain. 2020. “Optimizing Open Data to Support One Health: Best Practices to Ensure Interoperability of Genomic Data from Bacterial Pathogens.” *One Health Outlook* 2: 20. <https://doi.org/10.1186/s42522-020-00026-3>
12. Muhamad Rizal, Nurnabila Syafiqah, Hui-min Neoh, Ramliza Ramli, Petrick Ramesh A/L K Periyasamy, Alfizah Hanafiah, Muttaqillah Najihan Abdul Samat, Toh Leong Tan, et al. 2020. “Advantages and Limitations of 16S rRNA Next-Generation Sequencing for Pathogen Identification in the Diagnostic Microbiology Laboratory: Perspectives from a Middle-Income Country.” *Diagnostics* 10: 816. <https://doi.org/10.3390/diagnostics10100816>
13. Lagier, Jean-Christophe, Sophie Edouard, Isabelle Pagnier, Oleg Mediannikov, Michel Drancourt, and Didier Raoult. 2015. “Current and Past Strategies for Bacterial Culture in Clinical Microbiology.” *Clinical Microbiology Reviews* 28: 208–36. <https://doi.org/10.1128/CMR.00110-14>
14. Palacio-Bielsa, A., M. A. Cambra, and M. M. López. 2009. “PCR Detection and Identification of Plant-Pathogenic Bacteria: Updated Review and Identification of Plant-Pathogenic Bacteria: Updated Review of Protocols (1989–2007).” *Journal of Plant Pathology* 91: 249–97. <https://doi.org/10.4454/jpp.v91i2.957>
15. Allicock, Orchid M., Cheng Guo, Anne-Catrin Uhlemann, Susan Whittier, Lokendra V. Chauhan, Joel Garcia, Adam Price, et al. 2018. “BacCapSeq: A Platform for Diagnosis and Characterization of Bacterial Infections.” *mBio* 9: e02007–02018. <https://doi.org/10.1128/mBio.02007-18>
16. Garner, Emily, Benjamin C. Davis, Erin Milligan, Matthew Forrest Blair, Ishi Keenum, Ayella Maile-Moskowitz, Jin Pan, et al. 2021. “Next Generation Sequencing Approaches to Evaluate Water and Wastewater Quality.” *Water Research* 194: 116907. <https://doi.org/10.1016/j.watres.2021.116907>
17. He, Xuejun, Ningzhi Zhang, Wenye Cao, Yiqiao Xing, and Ning Yang. 2022. “Application Progress of High-Throughput Sequencing in Ocular Diseases.” *Journal of Clinical Medicine* 11: 3485. <https://doi.org/10.3390/jcm11123485>
18. Chen, Peixin, Wenwen Sun, and Yayi He. 2020. “Comparison of the Next-Generation Sequencing (NGS) Technology with Culture Methods in the Diagnosis of Bacterial and Fungal Infections.” *Journal of Thoracic Disease* 12: 4924–29. <https://doi.org/10.21037/jtd-20-930>
19. Huang, Zhijian, Shenzheng Zeng, Jinbo Xiong, Dongwei Hou, Renjun Zhou, Chengguang Xing, Dongdong Wei, et al. 2020. “Microecological Koch’s Postulates Reveal that Intestinal Microbiota Dysbiosis Contributes to Shrimp White Feces Syndrome.” *Microbiome* 8: 32. <https://doi.org/10.1186/s40168-020-00802-3>
20. Chen, Casey, Chris Hemme, Joan Beleno, Zhou Jason Shi, Daliang Ning, Yujia Qin, and Qichao Tu, et al. 2018. “Oral Microbiota of Periodontal Health and Disease and Their Changes after Nonsurgical Periodontal Therapy.” *The ISME Journal* 12: 1210–24. <https://doi.org/10.1038/s41396-017-0037-1>
21. Lewis, Elizabeth, John A. Hudson, Nigel Cook, Jeremy D. Barnes, and Edward Haynes. 2020. “Next-Generation Sequencing as a Screening Tool for Foodborne Pathogens in Fresh Produce.” *Journal of Microbiological Methods* 171: 105840. <https://doi.org/10.1016/j.mimet.2020.105840>
22. Wei, Zhong, Jie Hu, Yi'an Gu, Shixue Yin, Yangchun Xu, Alexandre Jousset, Qirong Shen, and Ville-Petri Friman. 2018. “*Ralstonia solanacearum* Pathogen Disrupts Bacterial Rhizosphere Microbiome during an Invasion.” *Soil Biology & Biochemistry* 118: 8–17. <https://doi.org/10.1016/j.soilbio.2017.11.012>
23. Chen, Tao, Kinya Nomura, Xiaolin Wang, Reza Sohrabi, Jin Xu, Lingya Yao, Bradley C. Paasch, et al. 2020. “A Plant Genetic Network for Preventing Dysbiosis in the Phyllosphere.” *Nature* 580: 653–57. <https://doi.org/10.1038/s41586-020-2185-0>
24. Edgar, Robert C. 2013. “UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads.” *Nature Methods* 10: 996–98. <https://doi.org/10.1038/nmeth.2604>
25. Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. “DADA2: High-Resolution Sample Inference from Illumina Amplicon Data.” *Nature Methods* 13: 581–83. <https://doi.org/10.1038/nmeth.3869>
26. Prodan, Andrei, Valentina Tremaroli, Harald Brolin, Aeilko H. Zwinderman, Max Nieuwdorp, and Evgeni Levin. 2020. “Comparing Bioinformatic Pipelines for Microbial 16S rRNA Amplicon Sequencing.” *Plos One* 15: e0227434. <https://doi.org/10.1371/journal.pone.0227434>
27. Johnson, Jethro S., Daniel J. Spakowicz, Bo-Young Hong, Lauren M. Petersen, Patrick Demkowicz, Lei Chen, Shana R. Leopold, et al. 2019. “Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis.” *Nature Communications* 10: 5029. <https://doi.org/10.1038/s41467-019-13036-1>
28. Liao, Chun-Chieh, Po-Ying Fu, Chih-Wei Huang, Chia-Hsien Chuang, Yun Yen, Chung-Yen Lin, and Shu-Hwa Chen. 2022. “MetaSquare: An Integrated Metadatabase of 16S rRNA Gene Amplicon for Microbiome Taxonomic Classification.” *Bioinformatics* 38: 2930–31. <https://doi.org/10.1093/bioinformatics/btac184>
29. Miao, Jiaojiao, Na Han, Yujun Qiang, Tingting Zhang, Xiwen Li, and Wen Zhang. 2017. “16SPIP: A Comprehensive Analysis Pipeline for Rapid Pathogen Detection in Clinical Samples Based on 16S Metagenomic Sequencing.” *BMC Bioinformatics* 18: 568. <https://doi.org/10.1186/s12859-017-1975-3>
30. Sun, Zheng, Xudong Liu, Gongchao Jing, Yuzhu Chen, Shuaiming Jiang, Meng Zhang, Jiquan Liu, Jian Xu, and Xiaoquan Su. 2022. “Comprehensive Understanding to the Public Health Risk of Environmental Microbes via a Microbiome-Based Index.” *Journal of Genetics and Genomics* 49: 685–88. <https://doi.org/10.1016/j.jgg.2021.12.011>
31. Bal Krishna, K. C., Arumugam Sathasivan, and Maneesha P. Ginige. 2021. “An Assessment of the Persistence

- of Putative Pathogenic Bacteria in Chloraminated Water Distribution Systems." *Water Research* 190: 116677. <https://doi.org/10.1016/j.watres.2020.116677>
32. Bird, Brian H., and Jonna A. K. Mazet. 2018. "Detection of Emerging Zoonotic Pathogens: An Integrated One Health Approach." *Annual Review of Animal Biosciences* 6: 121–39. <https://doi.org/10.1146/annurev-animal-030117-014628>
33. Lax, Simon, Naseer Sangwan, Daniel Smith, Peter Larsen, Kim M. Handley, Miles Richardson, Kristina Guyton, et al. 2017. "Bacterial Colonization and Succession in a Newly Opened Hospital." *Science Translational Medicine* 9: eaah6500. <https://doi.org/10.1126/scitranslmed.aah6500>
34. Bokulich, Nicholas A., Benjamin D. Kaehler, Jai Ram Rideout, Matthew Dillon, Evan Bolyen, Rob Knight, Gavin A. Huttley, and J. Gregory Caporaso. 2018. "Optimizing Taxonomic Classification of Marker-Gene Amplicon Sequences with QIIME 2's q2-Feature-Classifier Plugin." *Microbiome* 6: 90. <https://doi.org/10.1186/s40168-018-0470-z>
35. Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2012. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41: D590–596. <https://doi.org/10.1093/nar/gks1219>
36. Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics (Oxford, England)* 26: 2460–61. <https://doi.org/10.1093/bioinformatics/btq461>
37. Wang, Ji-Xiang, Xiang Cao, and Yong-Ping Chen. 2021. "An Air Distribution Optimization of Hospital Wards for Minimizing Cross-Infection." *Journal of Cleaner Production* 279: 123431. <https://doi.org/10.1016/j.jclepro.2020.123431>
38. Son, Sungwan, and Choon-Man Jang. 2022. "Effects of Internal Airflow on IAQ and Cross-Infection of Infectious Diseases between Students in Classrooms." *Atmospheric Environment* 279: 119112. <https://doi.org/10.1016/j.atmosenv.2022.119112>
39. Greay, Telleasha L., Alexander W. Gofton, Alireza Zahedi, Andrea Paparini, Kathryn L. Linge, Cynthia A. Joll, and Una M. Ryan. 2019. "Evaluation of 16S Next-Generation Sequencing of Hypervariable Region 4 in Wastewater Samples: An Unsuitable Approach for Bacterial Enteric Pathogen Identification." *Science of The Total Environment* 670: 1111–24. <https://doi.org/10.1016/j.scitotenv.2019.03.278>
40. Mechan-Llontop, Marco Enrique, Long Tian, Vivian Bernal-Galeano, Ella Reeves, Mary Ann Hansen, Elizabeth Ann Bush, and Boris Alexander Vinatzer. 2019. "Assessing the Potential of Culture-Independent 16S rRNA Microbiome Analysis in Disease Diagnostics: The Example of *Dianthus gratianopolitanus* and *Robbsia andropogonis*." *European Journal of Plant Pathology* 155: 1211–23. <https://doi.org/10.1007/s10658-019-01850-8>
41. Matsuo, Yoshiyuki, Shinnosuke Komiya, Yoshiaki Yasumizu, Yuki Yasuoka, Katsura Mizushima, Tomohisa Takagi, Kirill Kryukov, et al. 2021. "Full-Length 16S rRNA Gene Amplicon Analysis of Human Gut Microbiota Using MinION™ Nanopore Sequencing Confers Species-Level Resolution." *BMC Microbiology* 21: 35. <https://doi.org/10.1186/s12866-021-02094-5>
42. Ogier, Jean-Claude, Sylvie Pagès, Maxime Galan, Matthieu Barret, and Sophie Gaudriault. 2019. "rpoB, A Promising Marker for Analyzing the Diversity of Bacterial Communities by Amplicon Sequencing." *BMC Microbiology* 19: 171. <https://doi.org/10.1186/s12866-019-1546-z>
43. Vancuren, Sarah J., and Janet E. Hill. 2019. "Update on cpnDB: A Reference Database of Chaperonin Sequences." *Database* 2019: baz033. <https://doi.org/10.1093/database/baz033>
44. Chen, Yuzhu, Jian Li, Yufeng Zhang, Mingqian Zhang, Zheng Sun, Gongchao Jing, Shi Huang, and Xiaoquan Su. 2022. "Parallel-Meta Suite: Interactive and Rapid Microbiome Data Analysis on Multiple Platforms." *iMeta* 1: e1. <https://doi.org/10.1002/imt2.1>
45. Louca, S., L. W. Parfrey, and M. Doebeli. 2016. "Decoupling Function and Taxonomy in the Global Ocean Microbiome." *Science* 353: 1272–77. <https://doi.org/10.1126/science.aaf4507>
46. Bull, Carolee T., Solke H. De Boer, Timothy P. Denny, Giuseppe Firrao, Marion Fischer-Le Saux, Gerry S. Saddler, Marco Scortichini, David E. Stead, and Yuichi Takikawa. 2010. "Comprehensive List of Names of Plant Pathogenic Bacteria, 1980–2007." *Journal of Plant Pathology* 92: 551–92. <https://doi.org/10.4454/jpp.v92i3.302>
47. Bull, Carolee T., Solke H. De Boer, Timothy P. Denny, Giuseppe Firrao, Marion Fischer-Le Saux, Gerry S. Saddler, y Marco Scortichini, David E. Stead, and Yuichi Takikawa. 2012. "List of New Names of Plant Pathogenic Bacteria (2008–2010)." *Journal of Plant Pathology* 94: 21–27. <https://doi.org/10.4454/jpp.fa.2011.003>
48. Bull, Carolee T., Teresa A. Coutinho, Timothy P. Denny, Giuseppe. Firrao, Marco Fischer-Le Saux, Xuguang Li, Gerry S. Saddler, et al. 2014. "List of New Names of Plant Pathogenic Bacteria (2011–2012)." *Journal of Plant Pathology* 96: 223–26. <https://doi.org/10.4454/jpp.v96i2.026>
49. Yilmaz, Pelin, Laura Wegener Parfrey, Pablo Yarza, Jan Gerken, Elmar Pruesse, Christian Quast, Timmy Schweer, et al. 2014. "The SILVA and "All-Species Living Tree Project (LTP)" Taxonomic Frameworks." *Nucleic Acids Research* 42: D643–648. <https://doi.org/10.1093/nar/gkt1209>
50. Katoh, Kazutaka. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30: 3059–66. <https://doi.org/10.1093/nar/gkf436>
51. Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. "FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments." *Plos One* 5: e9490. <https://doi.org/10.1371/journal.pone.0009490>
52. Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation." *Plos One* 11: e0163962. <https://doi.org/10.1371/journal.pone.0163962>
53. Caporaso, J. Gregory, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, et al. 2010. "QIIME Allows Analysis of High-Throughput Community Sequencing Data." *Nature Methods* 7: 335–36. <https://doi.org/10.1038/nmeth.f.303>
54. Jiang, Gaofei, Ningqi Wang, Yaoyu Zhang, Zhen Wang, Yuling Zhang, Jiabao Yu, Yong Zhang, et al. 2021. "The Relative Importance of Soil Moisture in Predicting Bacterial Wilt Disease Occurrence." *Soil Ecology Letters* 3: 356–66. <https://doi.org/10.1007/s42832-021-0086-2>

55. Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. "fastp: An Ultra-Fast All-in-One FASTQ Preprocessor." *Bioinformatics* 34: i884–890. <https://doi.org/10.1093/bioinformatics/bty560>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yang, Xinrun, Gaofei Jiang, Yaohong Zhang, Ningqi Wang, Yuling Zhang, Xiaofang Wang, Fang-Jie Zhao, Yangchun Xu, Qirong Shen, and Zhong Wei. 2023. "MBPD: A Multiple Bacterial Pathogen Detection Pipeline for One Health practices." *iMeta* e82.

<https://doi.org/10.1002/imt2.82>