



## OPEN ACCESS

EDITED BY  
Muhammad Waseem,  
Hainan University, China

REVIEWED BY  
Dixit Sharma,  
Central University of Himachal Pradesh,  
India  
Zheng Zheng,  
South China Agricultural University, China

\*CORRESPONDENCE  
Chun Ran  
✉ ranchun@cric.cn  
Xue-Feng Wang  
✉ wangxuefeng@cric.cn

RECEIVED 22 December 2022

ACCEPTED 02 May 2023

PUBLISHED 29 May 2023

## CITATION

Liu H-Q, Zhao Z-I, Li H-J, Yu S-J, Cong L,  
Ding L-L, Ran C and Wang X-F (2023)  
Accurate prediction of huanglongbing  
occurrence in citrus plants by machine  
learning-based analysis of  
symbiotic bacteria.  
*Front. Plant Sci.* 14:1129508.  
doi: 10.3389/fpls.2023.1129508

## COPYRIGHT

© 2023 Liu, Zhao, Li, Yu, Cong, Ding, Ran  
and Wang. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Accurate prediction of huanglongbing occurrence in citrus plants by machine learning-based analysis of symbiotic bacteria

Hao-Qiang Liu<sup>1</sup>, Ze-long Zhao<sup>2</sup>, Hong-Jun Li<sup>1</sup>, Shi-Jiang Yu<sup>1</sup>,  
Lin Cong<sup>1</sup>, Li-Li Ding<sup>1</sup>, Chun Ran<sup>1\*</sup> and Xue-Feng Wang<sup>1\*</sup>

<sup>1</sup>Citrus Research Institute, Southwest University/Chinese Academy of Agricultural Sciences, National Engineering Research Center for Citrus, Chongqing, China, <sup>2</sup>Shanghai BIOZERON Biotechnology Co., Ltd., Shanghai, China

Huanglongbing (HLB), the most prevalent citrus disease worldwide, is responsible for substantial yield and economic losses. Phytobiomes, which have critical effects on plant health, are associated with HLB outcomes. The development of a refined model for predicting HLB outbreaks based on phytobiome markers may facilitate early disease detection, thus enabling growers to minimize damages. Although some investigations have focused on differences in the phytobiomes of HLB-infected citrus plants and healthy ones, individual studies are inappropriate for generating common biomarkers useful for detecting HLB on a global scale. In this study, we therefore obtained bacterial information from several independent datasets representing hundreds of citrus samples from six continents and used these data to construct HLB prediction models based on 10 machine learning algorithms. We detected clear differences in the phyllosphere and rhizosphere microbiomes of HLB-infected and healthy citrus samples. Moreover, phytobiome alpha diversity indices were consistently higher for healthy samples. Furthermore, the contribution of stochastic processes to citrus rhizosphere and phyllosphere microbiome assemblies decreased in response to HLB. Comparison of all constructed models indicated that a random forest model based on 28 bacterial genera in the rhizosphere and a bagging model based on 17 bacterial species in the phyllosphere predicted the health status of citrus plants with almost 100% accuracy. Our results thus demonstrate that machine learning models and phytobiome biomarkers may be applied to evaluate the health status of citrus plants.

## KEYWORDS

citrus microbiome, Huanglongbing, machine learning, meta-analysis, community assembly

## Introduction

The plant phytobiome, consisting of the rhizosphere, phyllosphere, and endosphere, harbors diverse microbes that affect plant growth and health (Berendsen et al., 2012; Liu et al., 2019; Liu et al., 2020). Considerable research has been carried out on the diversity, composition, and function of the phytobiome of various plant species, including rice (Edwards et al., 2015), maize (Peiffer et al., 2013), corn (Jat et al., 2021), soybean (Zhang et al., 2018), cucumber (Zhou et al., 2022a), citrus (Xu et al., 2018), and tomato (Oyserman et al., 2022). Microbes inhabiting plant surfaces or internal tissues produce metabolites that support plant growth by regulating physiological processes (e.g., nutrient absorption and pathogen suppression) (Trivedi et al., 2020). The identification of microbes associated with specific phenotypes is currently a fundamental objective of phytobiome researchers. Moreover, changes to the plant microbiome are influenced by diverse biotic and abiotic factors of the host and surrounding environment (Dastogeer et al., 2020). Niche and neutral theory-based approaches have been used to explore the mechanisms modulating the microbiome assembly, with all factors classified into deterministic or stochastic processes (Chen et al., 2019). These studies have expanded our understanding of phytobiomes, with implications for managing plant-associated microbiomes to enhance crop production (French et al., 2021).

Citrus is an economically important fruit crop comprising several widely cultivated species initially domesticated more than 1,000 years ago (Gmitter and Hu, 1990). Because their fruits contain an abundance of diverse nutrients, vitamins, and dietary fiber, citrus species are extensively cultivated worldwide, with annual yields exceeding 120 million tons (Mahato et al., 2021). Nevertheless, citrus production has been restricted by global climate change and the prevalence of diseases (Wang, 2019; Wang, 2020). Huanglongbing (HLB), a fatal disease, is the biggest threat to citrus production. The high prevalence of HLB and associated, considerable yield losses have recently renewed interest in this disease (Das et al., 2019). HLB leads to a loss of citrus root carbonaceous compounds, malfunctioning phloem tissues, and decreased release of photosynthates, all of which impair the transport of photoassimilates (Trivedi et al., 2020). In addition, HLB upsets the nutrient balance by altering the ability of roots to absorb and transport nutrients and water (Wang, 2020). HLB is mainly caused by the bacterium *Candidatus Liberibacter asiaticus* (CLAs) (Li et al., 2017). Because CLAs cannot be monocultured (Killiny-Mansour, 2019), clarifying citrus-associated microbiome dynamics due to HLB pressure is of serious interest.

A comprehensive understanding of citrus-associated microbiomes may lead to the development of sustainable, environmentally friendly methods for increasing citrus plant health and productivity. Global patterns in citrus phytobiomes are already being revealed (Xu et al., 2018; Zhang et al., 2021; Penyalver et al., 2022). For instance, the International Citrus Microbiome Consortium, established in 2015, has performed sequencing analyses of citrus rhizosphere and soil microbiome samples from major citrus-producing regions on six continents (Wang et al., 2015). In addition, recent Illumina-based sequencing of 16S rRNA

genes has revealed that HLB alters the microbiome of citrus rhizospheres and phyllospheres (Srivastava et al., 2022). Most related research has only focused on healthy citrus microbiomes or HLB-infected ones, however, with relatively few comparative studies performed on both types of microbiomes (Li et al., 2017; Ginnan et al., 2018; Blacutt et al., 2020). Whether some citrus taxa in worldwide cultivation are robust and universally responsive to HLB remains unclear.

In this study, we systematically reviewed available data on the citrus phytobiome and compared the bacterial communities of healthy and HLB-infected citrus plants. In addition, machine learning approaches were applied to identify potential biomarkers for HLB occurrence after technical biases, geographic distribution, and tissue specificity were taken into account. These analyses allowed us to reveal the diversity, composition, and mechanisms underlying the bacterial community assembly of HLB-infected citrus plants. Finally, we developed a phytobiome-based model to predict HLB outbreaks under field conditions.

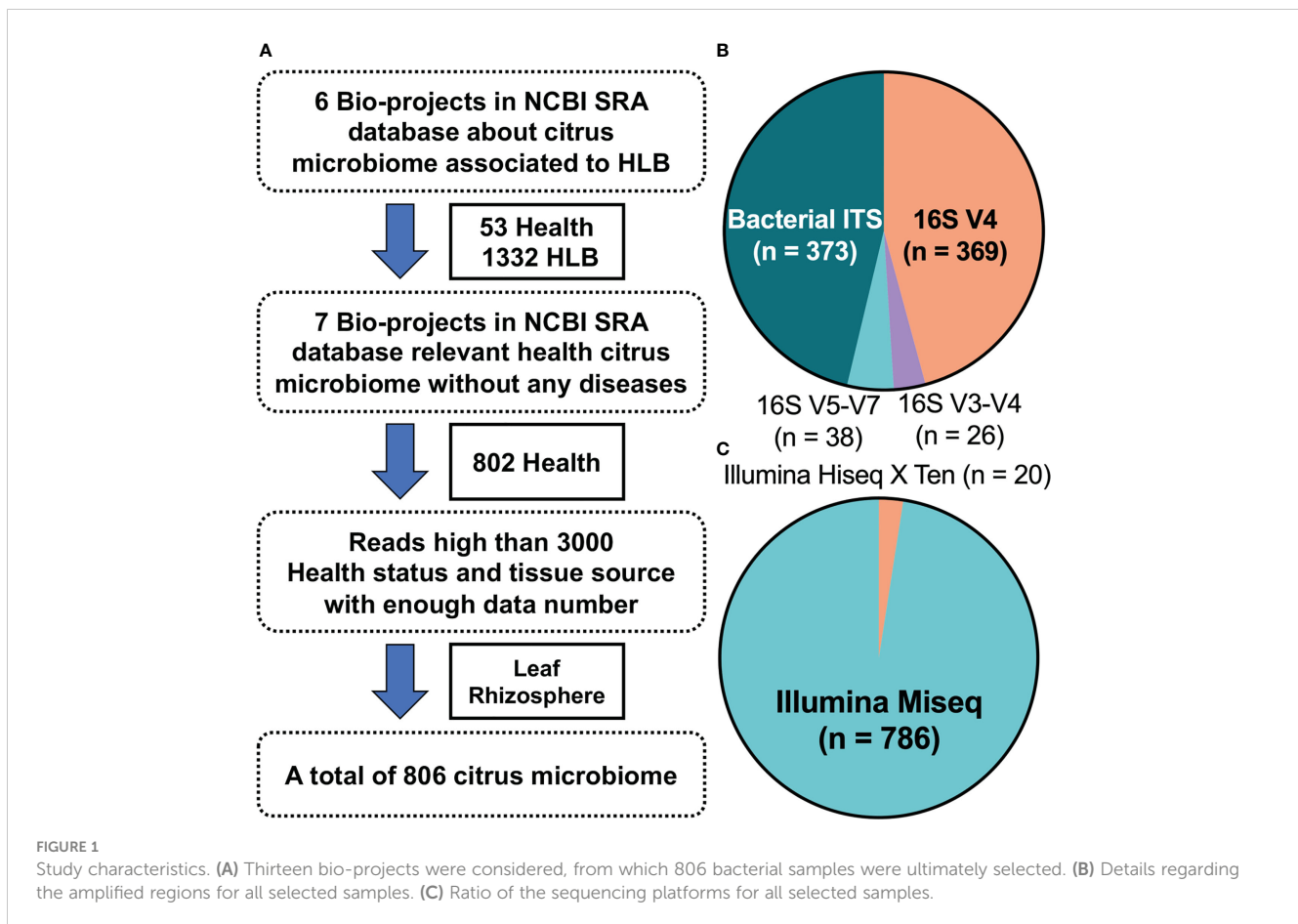
## Materials and methods

### Data collection and description

Screening of the National Center for Biotechnology Information Sequence Read Archive database using “citrus” and “huanglongbing” as keywords yielded six HLB-related citrus microbiome bio-projects that included 1,385 bacterial samples (53 healthy citrus samples and 1,332 HLB-infected ones). Another seven bio-projects with 802 bacterial samples from healthy citrus plants were also considered. The metadata for these bio-projects (Table S1) were classified according to source tissue/material into five groups (Table S2). Because samples from budwood, bulk soil, and attached insects were limited, only leaf and rhizosphere datasets were analyzed further (Table S2). Moreover, samples with fewer than 3,000 reads were removed to eliminate abnormal sequencing results. Finally, 806 citrus microbiome samples were retained (Figure 1A): 29 and 207 from healthy citrus leaves and rhizospheres, respectively, and 267 and 303 from HLB-infected citrus leaves and rhizospheres, respectively (Table S3). Four amplified regions, mainly bacterial ITS (46.28%) and 16S V4 (45.78%), were identified in the metadata (Figure 1B). The sequencing data were mostly produced on the Illumina MiSeq platform, with only 20 samples sequenced using the Illumina HiSeq X Ten system (Figure 1C).

### Data processing

The collected datasets, in FASTQ format, were analyzed using a standard Quantitative Insights Into Microbial Ecology 2 (QIIME2) pipeline (Bolyen et al., 2019). After removal of primer sequences and quality control, the resulting clean reads were clustered to obtain amplicon sequence variants (ASVs) using the DADA2 plugin unit (Bokulich et al., 2018). Each ASV was assigned to a taxon according to a closed-reference strategy using the SILVA database



(release 138) (Yilmaz et al., 2014). In this approach, a reference database comprising the full-length sequences of amplified targets was predefined and used to generate representative sequences and taxonomically classify sequences produced by different primers (Yu et al., 2018). Non-bacterial ASVs (i.e., chloroplasts and archaea) and singletons (ASVs with only one read) were discarded. Finally, the ASV abundance tables were rarefied to 3,000 reads per sample because of the unequal sequencing depth.

## Statistical analyses

All statistical analyses were conducted in R (v4.0.2), and the results were visualized using the R “ggplot2” package (R Core Team, 2020). The following three alpha diversity indices for bacterial communities were calculated using the “vegan” package (Oksanen et al., 2020): Chao1 (richness), Shannon’s (diversity), and Pielou’s J (evenness). Differences in the alpha diversity and relative abundance of bacterial phyla and genera between healthy and HLB-infected citrus leaf and rhizosphere samples were analyzed using the *t*-test function. Bray–Curtis distances between bacterial communities were calculated using the *vegdist* function in the “vegan” package. These distances were then used in a principal coordinate analysis (PCoA) with the *pcoa* function in the “ape” package and a permutational multivariate analysis of variance (PERMANOVA) with the *adonis* function in the “vegan” package

to evaluate differences in the bacterial communities of healthy and HLB-infected citrus leaf and rhizosphere samples. A Venn diagram-based analysis was performed using the “VennDiagram” package (Chen, 2022) to identify shared bacterial ASVs among samples. Differences in the total relative abundance of shared ASVs were assessed by ANOVA followed by Tukey’s HSD test (“multcomp” package). Finally, a neutral community model was used to determine the potential importance of deterministic and stochastic processes on the assembly of bacterial communities. In this model, two variables were defined: *m*, an estimate of the dispersal between communities, and *R*<sup>2</sup>, which represented the ratio of the contributions of stochastic processes (Sloan et al., 2006).

## Machine-learning modeling

To more precisely distinguish between bacterial communities of HLB-infected and healthy citrus plants, we applied the following 10 established machine learning algorithms (Gupta et al., 2021) to construct models according to relative abundances of bacteria (phylum to species levels): logistic regression, decision tree, *k*-nearest neighbor, bagging, gradient boosting, Bayes classification, artificial neural network, conditional inference tree, random forests, and support vector machines. Models were constructed for leaf and rhizosphere microbiomes separately. More specifically, 70% of healthy and HLB-infected citrus samples were randomly selected

as training data to construct models, and the remaining 30% were used as testing data for model validation. The predicted results were compared with actual health status using two metrics: receiver operating characteristic curve (ROC) and area under the curve (AUC) scores (Sing et al., 2005). Finally, the best-performing models (i.e., those with the highest AUC scores and accuracies) were identified, and the importance of various features in the classification was determined.

## Results

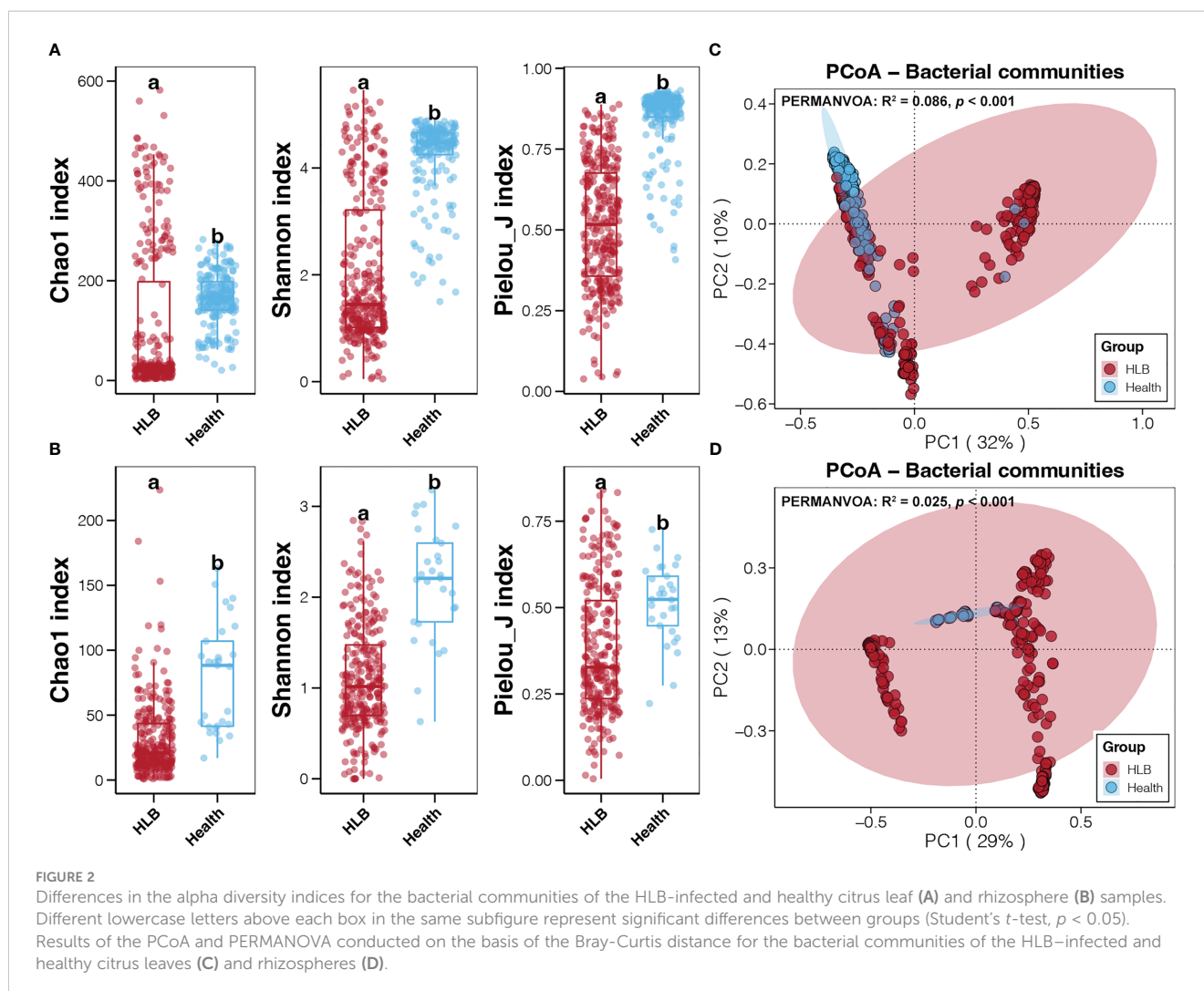
### Differences in the microbiome diversity of HLB-infected and healthy citrus samples

Meta-analysis of sequencing data for 806 bacterial samples from six continents generated a merged bacterial ASV table comprising more than 3,700 taxa. All of these bacterial ASVs were annotated at the phylum level, but only 67.45% and 25.5% were annotated at genus and species levels, respectively (Figure S1). We rarefied the sequencing data to 3,000 reads per sample before calculating alpha diversity indices, which were lowest and highest for HLB-infected

citrus leaf and healthy citrus rhizosphere microbiomes, respectively (Figure S2). More importantly, Chao1, Shannon's, and Pielou's J indices of leaf and rhizosphere bacterial communities were significantly lower for HLB-infected citrus samples than healthy ones (Student's *t*-test,  $p < 0.05$ ; Figures 2A, B). In addition, inter-individual differences in the alpha diversity indices of leaf and rhizosphere bacterial communities were more obvious for HLB-infected samples than for healthy samples (Figures 2A, B). According to PCoA and PERMANOVA, the bacterial community structures of HLB-infected and healthy citrus leaf and rhizosphere samples were significantly different ( $p < 0.05$ ; Figures 2C, D). Moreover, inter-individual differences in leaf and rhizosphere bacterial communities were greater for HLB-infected samples than for healthy ones (Figures 2C, D), consistent with the differences observed in alpha diversity indices.

### Taxonomic classification of bacteria in citrus leaf and rhizosphere microbiomes

The most dominant bacterial phylum in citrus leaf and rhizosphere samples was Proteobacteria, which was followed by



Cyanobacteria and Actinobacteriota in leaf and rhizosphere microbiomes, respectively (Figure 3A). Actinobacteriota and Firmicutes were more abundant in HLB-infected citrus rhizospheres than in healthy citrus rhizospheres, whereas the opposite pattern was observed for Proteobacteria and Bacteroidota (Figure 3B). HLB-infected leaves had a higher relative abundance of Proteobacteria, whereas healthy leaves had higher relative abundances of Cyanobacteria, Actinobacteriota, and Bacteroidota (Figure 3C). We also detected 186 shared taxa among bacterial communities (Figure 3D). The total relative abundances of shared bacteria in citrus leaf and rhizosphere microbiomes were significantly higher for HLB-infected samples than for healthy ones (Tukey's HSD test,  $p < 0.05$ ; Figure 3E). Most shared bacteria with increased abundances in HLB-infected samples belonged to Proteobacteria, Firmicutes, and Nitrospinota (leaf microbiome) or Actinobacteriota (rhizosphere microbiome) (Figure 3F).

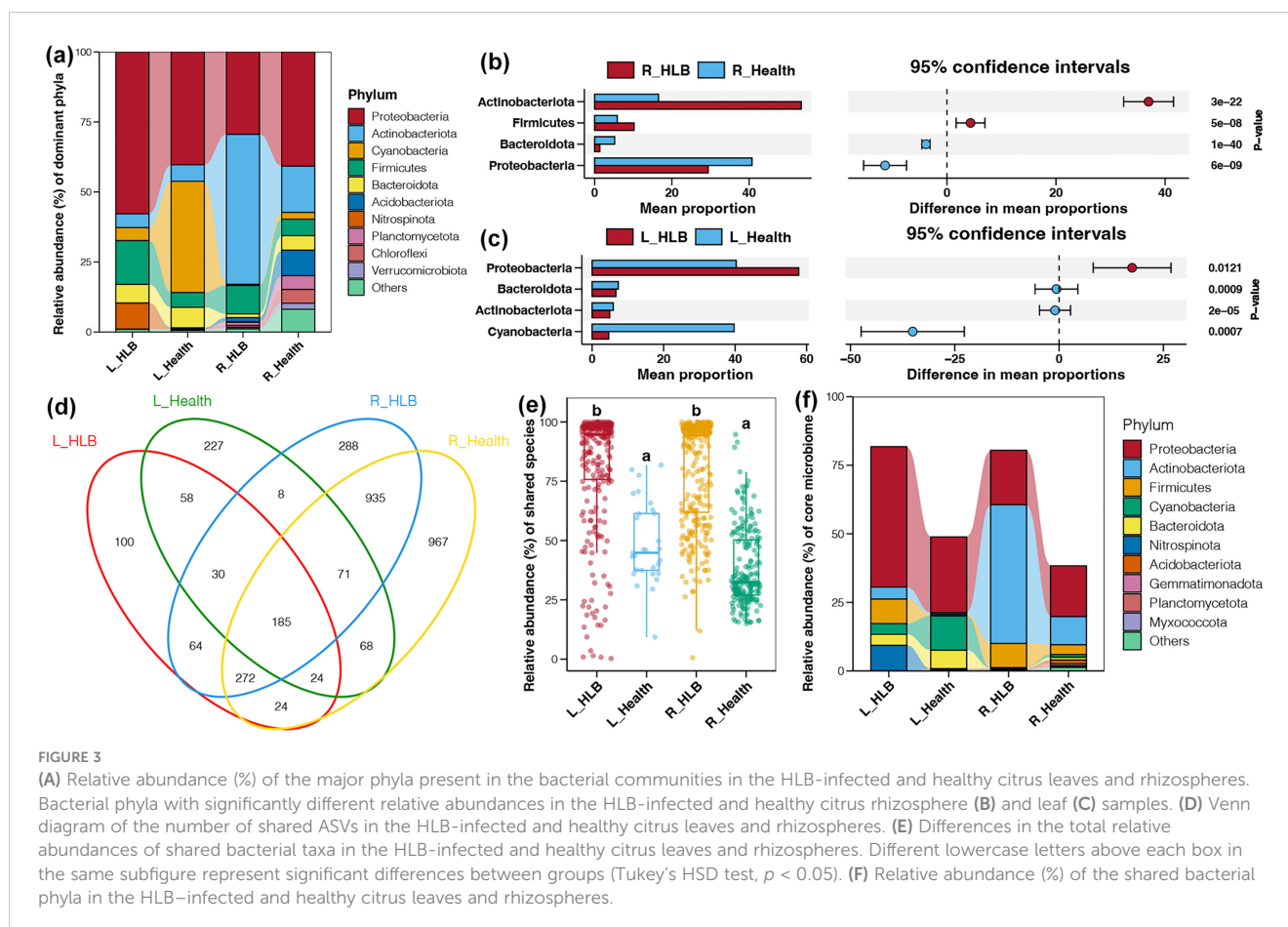
## Analysis of microbiome assembly mechanisms of HLB-infected and healthy citrus samples

To explore the mechanisms underlying the differences in leaf and rhizosphere microbiome assemblies between HLB-infected and healthy citrus samples, we examined the relative effects of niche

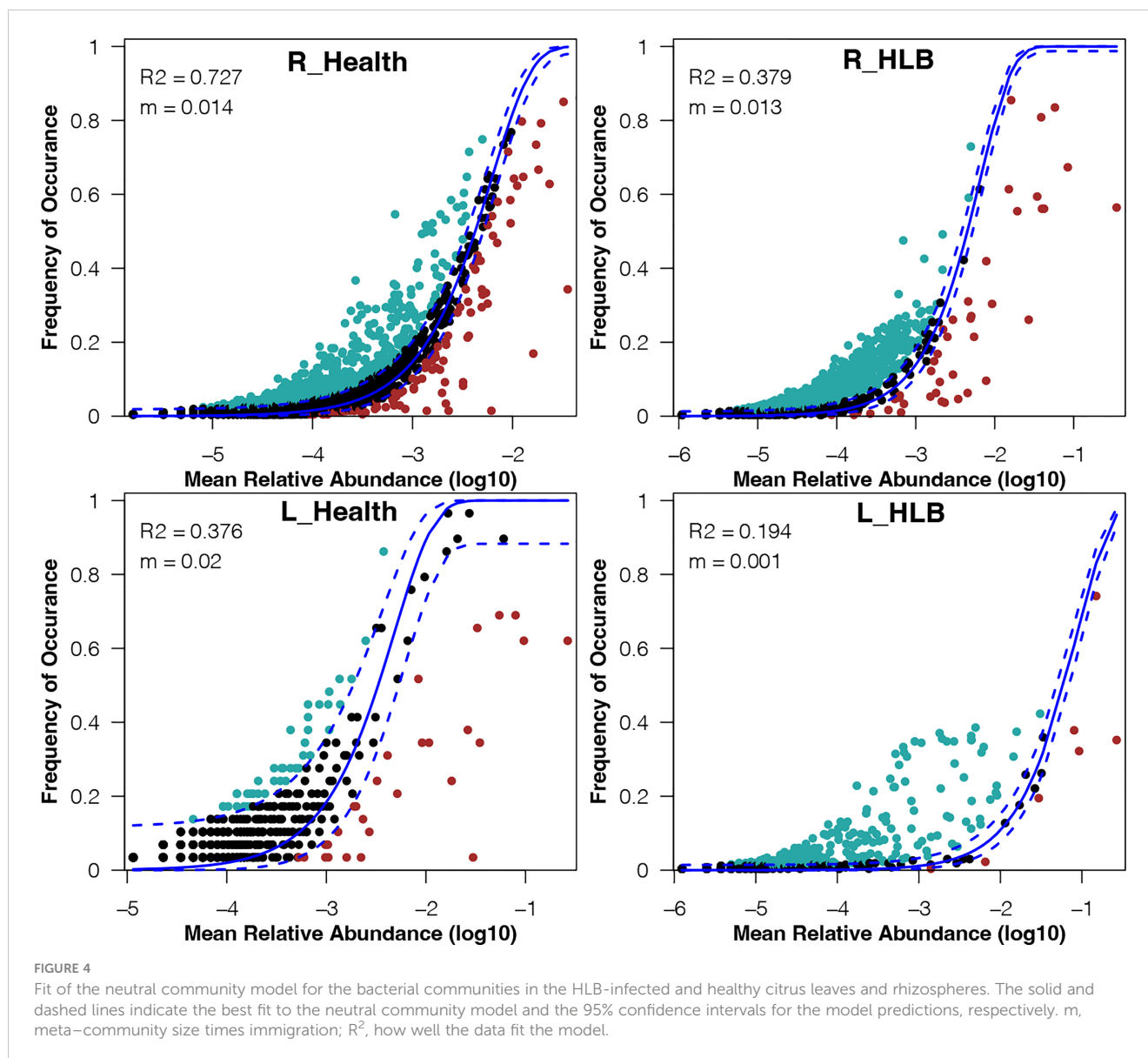
and neutral processes on the assembly of bacterial communities. The neutral community model explained a large proportion of the variance in the bacterial community of healthy citrus rhizospheres ( $R^2 = 0.727$ ), whereas only 37.9% of the corresponding variance in HLB-infected citrus rhizospheres was explained by this model (Figure 4). In contrast, the neutral community model explained only 19.4% and 37.6% of the variance in bacterial communities of HLB-infected and healthy citrus leaves, respectively (Figure 4). These results indicate that the bacterial community assembly in healthy citrus rhizospheres and leaves was respectively governed by stochastic and deterministic processes. More importantly, HLB obviously decreased the contribution of stochastic processes to the assembly of bacterial communities in citrus leaves and rhizospheres.

## Bacterial communities useful for distinguishing between HLB-infected and healthy citrus samples

To determine whether the properties of leaf and rhizosphere bacterial communities may be useful biomarkers for distinguishing between HLB-infected and healthy citrus plants, we constructed 10 machine learning models. The accuracy of model predictions based on the testing data as well as AUC and ROC data derived from the models were used to evaluate model performance (Figures S3, S4).







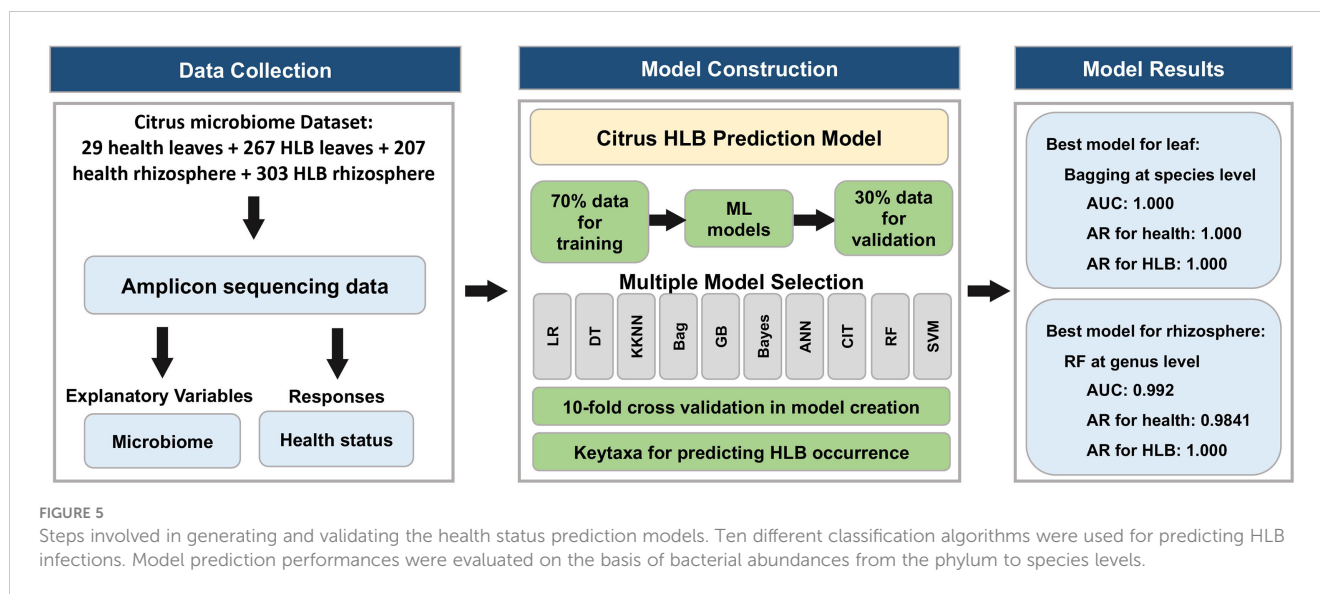
We also used the accuracy of predictions for healthy and HLB-infected samples to select the best models (Figure S5). The bagging model trained at the species level and the random forest model trained at the genus level were found to be the best models for classifying leaf and rhizosphere samples, respectively (Figure 5).

The bagging model for detecting HLB based on bacterial species in citrus leaves was constructed using 17 species—the most important of which was *Solanum melongena* (eggplant) (Figure 6A). Two of these species, (CLas and *Paraburkholderia rhizoxinica* HKI 454), had higher relative abundances in HLB-infected leaves, whereas 15 species had higher relative abundances in healthy leaves (Table S4). In terms of the citrus rhizosphere, 28 bacterial genera were defined as biomarker taxa for HLB, of which *Nitrospira* was the most important (Figure 6B). Four of these genera had higher relative abundances in HLB-infected citrus rhizospheres, and the other 24 genera had higher relative abundances in healthy citrus rhizospheres (Table S5). *Streptomyces*, *Burkholderia-*

*Caballeronia-Paraburkholderia*, and *Bacillus* were all enriched in HLB-infected citrus rhizospheres to the same degree relative to healthy rhizospheres.

## Discussion

The exploration of biomarkers common to the phytobiome of citrus plants infected with HLB is critical for developing improved methods for diagnosing plant diseases caused by bacteria and for determining optimal treatments. Nevertheless, knowledge of whether certain microbial lineages consistently respond to HLB across global biogeographic regions is unclear. In this study, we performed a meta-analysis of HLB-infected and healthy citrus rhizosphere and phyllosphere microbiomes on a global scale to screen for biomarkers useful for detecting HLB. By combining global datasets, we determined that Proteobacteria,



Actinobacteria, Acidobacteria, and Bacteroidetes were the predominant bacterial phyla in healthy citrus rhizospheres (Figure 3A). This result is consistent with the findings of an earlier study on citrus rhizosphere samples from six continents (Xu et al., 2018). Furthermore, the dominant bacterial phyla detected in healthy citrus phyllospheres as well as the distribution of their relative abundances in this study are in accordance with published results (Blaustein et al., 2017; Bai et al., 2019; Wu et al., 2020). These observations revealing similarities in citrus phytobiomes from various geographical regions imply that host phylogeny may influence phytobiome assembly more than geographical factors. Recent reports have described adaptive matching between rhizosphere and phyllosphere microbiomes and plant hosts (Lajoie et al., 2020; Escudero-Martinez et al., 2022).

Symbiotic microbiome homeostasis is closely associated with host physiological features and health (Paasch and He, 2021). As a fundamental indicator of the stability and performance of microbial communities, diversity is a crucial index for phytobiomes (Sare et al., 2020). A decrease in phytobiome richness and diversity is often responsible for the increased susceptibility of plant hosts to potentially harmful factors (Agler et al., 2016). In addition, decreased phytobiome diversity may be due to insufficient competition between resident commensals and invading pathogens (Thoms et al., 2021). In the current study, alpha diversity decreased in the rhizosphere and phyllosphere microbiomes of citrus plants infected with HLB (Figures 2A, B). Alpha diversity indices varied considerably among HLB-infected samples, however, which may be related to the features of the sequenced regions and the methods used. Moreover, the PERMANOVA results revealed that data source and sequenced target region had larger effects on citrus phytobiome composition than did citrus health status, geographic location, and tissue source (Table S6). Alpha diversity and citrus phytobiome composition were thus strongly affected by the methods used in different bio-projects and may therefore not be robust indicators of citrus health status.

Most research conducted on a global scale has suggested that relatively few bacterial taxa represent a large proportion of highly diverse bacterial communities. For example, a study examining global soil samples found that 2% of bacterial taxa accounted for nearly half of bacterial communities at various sites (Delgado-Baquerizo et al., 2018). Notably, Xu et al. (2018) reported that a small number of bacterial taxa (< 10%) were the core taxa in rhizospheres of citrus samples collected on various continents. In the present study, we detected 138 bacterial taxa common to all samples; these taxa represented approximately half of the bacterial communities in healthy citrus samples (Figure 3E). In contrast, the median total abundances of shared bacterial taxa in HLB-infected citrus leaves and rhizospheres corresponded to approximately 95% of the entire communities (Figure 3F). Hence, a limited number of core taxa were associated with the occurrence of HLB in citrus plants. This result may help to explain the decrease in alpha diversity index values among HLB samples.

The mechanisms mediating the assembly of phytobiome communities must be considered when designing plant microbiome management strategies (Trivedi et al., 2020). From a meta-community perspective, bacterial community assembly is governed by both deterministic and stochastic processes (Tian et al., 2022). If communities are controlled by deterministic processes, species will occupy specific ecological niches in a predictable fashion (Vanwonderghem et al., 2014). In contrast, multiple species can exist in similar or overlapping habitats in communities affected by stochastic fluctuations (Sloan et al., 2006). Our findings suggest that stochastic and deterministic processes are critical for shaping healthy citrus rhizosphere and phyllosphere microbiomes (Figure 4). Inconsistencies between citrus rhizosphere and phyllosphere community assemblies may be attributed to their different lifestyles and functions. Plant leaves and roots are located above- and belowground, respectively, which allows phyllosphere and rhizosphere microbiomes to perform different functions (Trivedi et al., 2012). In addition, low nutrient levels and long-term illumination may lead to changes in the abundances of specific

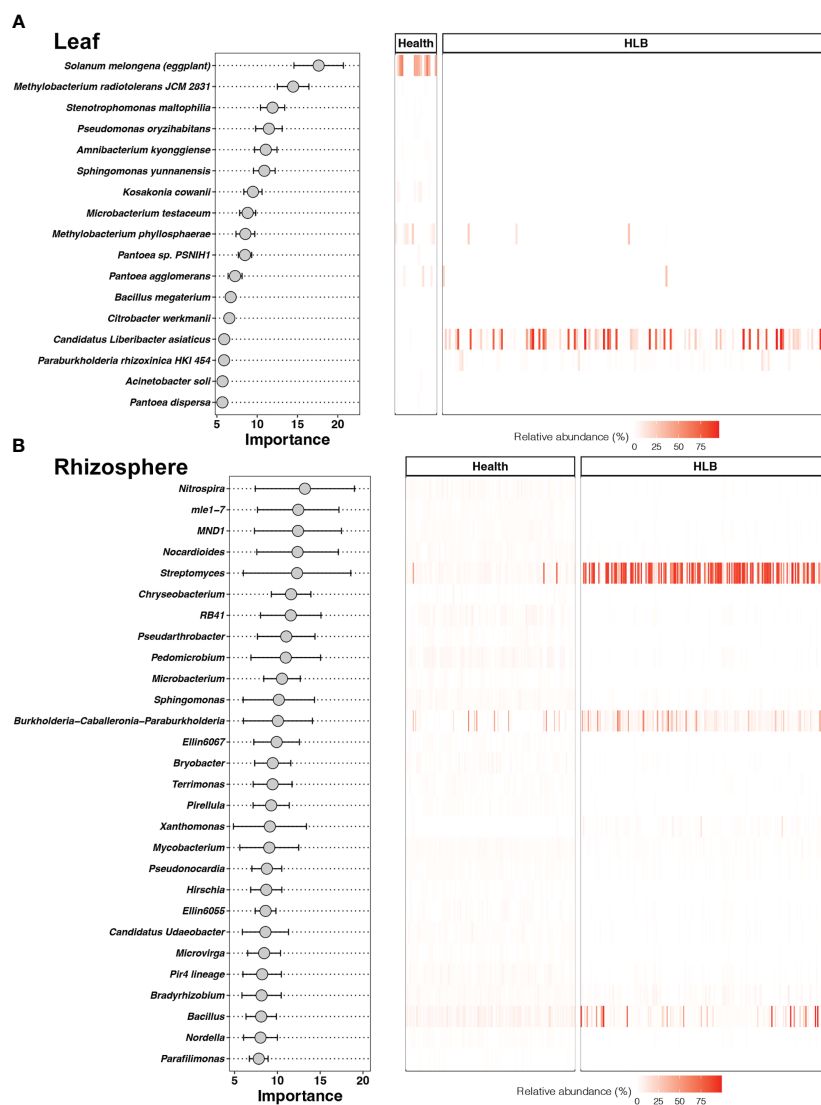


FIGURE 6

Biomarker taxa ranked in descending order of importance for the accuracy of predictions and their relative abundances in the leaf (A) and rhizosphere (B) models.

bacteria in the phyllosphere microbiome (Carvalho and Castillo, 2018), thereby increasing the importance of deterministic processes. Moreover, we observed that deterministic processes mediated phytobiome assembly more substantially in HLB-infected citrus samples than in healthy ones (Figure 4). Similar observations have been reported for phytobiomes of other diseased plants and the gut microbiota of diseased animals (Yao et al., 2018; Liu et al., 2022; Zhang et al., 2022). HLB is a disease caused by pathogenic bacteria, and the enriched pathogens will decrease the abundance of species with an overlapping niche and select for species without niche conflicts through competition (Chase, 2011). This phenomenon may explain why deterministic processes shaped the HLB-infected citrus microbiome.

Among the many statistical methods for elucidating the complex relationships between microbial communities and specific phenotypes, machine learning-based methods are considered the most promising (Torija and Ruiz, 2015). Machine

learning approaches take various forms according to their algorithms (e.g., unsupervised, semi-supervised, or supervised learning) (Ghannam and Techtmann, 2021). In the present study, we assessed the relationship between the health status of citrus samples collected worldwide and the relative abundance of bacteria at different taxonomic levels in the rhizosphere and phyllosphere. We used 10 machine learning approaches and found that the most appropriate models for predicting HLB infections were supervised learning methods (random forest and bagging) that were based on rhizosphere and phyllosphere bacterial genera and species (Figure 5). A previous meta-analysis demonstrated that supervised learning models for soil microbiomes may be used to predict the potential occurrence of *Fusarium* wilt disease in plants (Yuan et al., 2020). In addition, supervised learning methods have identified robust and reproducible features relevant for diagnosing shrimp diseases according to meta-analyses of gut microbiota (Sha et al., 2022). The origin and quality of sea cucumber cultured in



diverse geographic regions has also been accurately predicted using random forest models for gut microbiota (Zhao et al., 2022). Furthermore, supervised machine learning approaches have accurately predicted environmental health variables following analyses of microbiome data (Zhou et al., 2022b). Random forest models for microbial communities have predicted the soil health parameters of agroecosystems, with accuracies exceeding 80% (Wilhelm et al., 2022). These results provide convincing evidence of the utility of supervised learning methods for establishing models that accurately predict the health status of plants.

The vector of HLB, a bacterial infection of citrus trees, is believed to be the Asian citrus psyllid *Diaphorina citri* (Galdeano et al., 2020). At present, the dominant control strategies for HLB are removal of HLB-symptomatic citrus trees and the spraying of insecticides to restrict the psyllid (Coletta-Filho et al., 2014). Because HLB-infected trees may remain asymptomatic for several months, however, the efficacy of current disease control measures is limited (Lee et al., 2015). Using the bagging model, we identified crucial bacterial taxa related to citrus HLB disease incidence, including CLas and *Paraburkholderia rhizoxinica* in the phyllosphere and *Streptomyces*, *Burkholderia-Caballeronia-Paraburkholderia*, and *Bacillus* in the rhizosphere. A recent study indicated that CLas is the main pathogen responsible for HLB outbreaks (Ginnan et al., 2020). *Paraburkholderia rhizoxinica* is an endofungal bacterium that has a symbiotic relationship with phytopathogenic fungi (Braga et al., 2019). In contrast, bacteria enriched in HLB-infected citrus rhizospheres in our study were not directly related to the disease phenotype. These bacteria included antibiotic producers and species with detrimental effects on community stability (Nicholson, 2002; de Lima et al., 2012). Our findings imply that the risk of HLB can be assessed by screening for a few specific known pathogens in citrus leaves. Data for only 29 healthy citrus phyllospheres were included in our analyses, however, and a limited sample size and bias between two classifications can cause machine learning models to overestimate the risks of HLB outbreaks (Gupta et al., 2021). We thus recommend the use of a random forest model based on bacterial genera in the rhizosphere to predict the likelihood of HLB in citrus plants.

## Conclusions

In this study, we analyzed the utility of phytobiome examinations for detecting HLB-infected citrus plants on a global scale. Meta-analyses involving the phytobiome data of hundreds of citrus samples revealed significant decreases in rhizosphere and phyllosphere microbiome diversities of HLB-infected samples relative to healthy ones. Furthermore, the onset of HLB increased the contribution of deterministic processes to citrus rhizosphere and phyllosphere microbiome assemblies. We also identified 17 and 28 HLB-related taxa in the phyllosphere and rhizosphere, respectively. These taxa may be exploited to accurately predict citrus HLB outbreaks on the basis of selected machine learning models. The findings of this study are relevant for evaluating the risks of HLB in citrus plants according to phytobiome compositions

derived from 16S rRNA gene sequencing data. Advances in high-throughput sequencing technology and decreases in associated costs should enable researchers to further improve models for predicting the health status of agriculturally important plant species.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

CR and X-FW: funding and project administration. H-QL: methodology, ideas, data curation, and statistical analysis. Z-LZ, H-JL, S-JY, LC and L-LD participated in this work. All authors contributed to the article and approved the submitted version.

## Funding

This research was financially supported by the National Key R & D Program of China (grant nos. 2021YFD1400800, 2020YFD1000102, 2019YFD1002100, and 2018YFD0201500) and the Chongqing Scientific Research Project (grant no. cstc2021jcyj-bsh0082). We thank Liwen Bianji (Edanz) ([www.liwenbianji.cn](http://www.liwenbianji.cn)) for editing the English text of a draft of this manuscript.

## Conflict of interest

Author Z-LZ was employed by Shanghai BIOZERON Biotechnology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1129508/full#supplementary-material>

## References

- Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S. T., Weigel, D., et al. (2016). Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol.* 14 (1), e1002352. doi: 10.1371/journal.pbio.1002352
- Bai, Y. N., Wang, J. C., Jin, L., Zhan, Z. G., Guan, L. M., Zheng, G. H., et al. (2019). Deciphering bacterial community variation during soil and leaf treatments with biologicals and biofertilizers to control Huanglongbing in citrus trees. *J. Phytopathol.* 167, 686–694. doi: 10.1111/jph.12860
- Berendsen, R. L., Pieterse, C. M., and Bakker, P. A. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17 (8), 478–486. doi: 10.1016/j.tplants.2012.04.001
- Blacutt, A., Ginnan, N., Dang, T., Bodaghi, S., Vidalakis, G., Ruegger, P., et al. (2020). An in vitro pipeline for screening and selection of marker-gene amplicon sequences with potential anti-*Candidatus liberibacter asiaticus* properties. *Appl. Environ. Microbiol.* 86 (8), e02883–e02819.
- Blaustein, R. A., Lorca, G. L., Meyer, J. L., Gonzalez, C. F., and Teplitski, M. (2017). Defining the core citrus leaf- and root-associated microbiota: factors associated with community structure and implications for managing Huanglongbing (citrus greening) disease. *Appl. Environ. Microbiol.* 83, e00210–e00217. doi: 10.1128/AEM.00210-17
- Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with qiime 2's q2-feature-classifier plugin. *Microbiome* 6 (1), 90. doi: 10.1186/s40168-018-0470-z
- Bolyen, E., Rideout, J., Dillon, M., Bokulich, N., Abnet, C., Al-Ghalith, G., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9
- Braga, D., Last, D., Hasan, M., Guo, H., Leichnitz, D., Uzum, Z., et al. (2019). Metabolic pathway rerouting in *Paraburkholderia rhizoxinica* evolved long-overlooked derivatives of coenzyme F420. *ACS Chem. Biol.* 14 (9), 2088–2094. doi: 10.1021/acscchembio.9b00605
- Carvalho, S. D., and Castillo, J. A. (2018). Influence of light on plant-phylosphere interaction. *Front. Plant Sci.* 9, 1482. doi: 10.3389/fpls.2018.01482
- Chase, J. M. (2011). Ecological niche theory. *Theory Ecol.*, 93–107.
- Chen, H. (2022) *VennDiagram: generate high-resolution Venn and Euler plots. r package version 1.7.3*. Available at: <https://CRAN.R-project.org/package=VennDiagram>.
- Chen, W., Ren, K., Isabue, A., Chen, H., Liu, M., and Yang, J. (2019). Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* 7 (1), 1–16.
- Coletta-Filho, H. D., Daugherty, M. P., Ferreira, C., and Lopes, J. R. S. (2014). Temporal progression of *Candidatus liberibacter asiaticus* infection in citrus and acquisition efficiency by *Diaphorina citri*. *Phytopathology* 104 (4), 416–421. doi: 10.1094/PHYTO-06-13-0157-R
- Das, A. K., Sharma, S. K., and Thakre, N. (2019). “Diagnostics for citrus greening disease (HLB): current and emerging technologies,” in *Plant biotechnology: progress genomic era*. Eds. S. Khurana and R. Gaur (Singapore: Springer), 597–630.
- Dastogeer, K. M., Tumpa, F. H., Sultana, A., Akter, M. A., and Chakraborty, A. (2020). Plant microbiome—an account of the factors that shape community composition and diversity. *Curr. Plant Biol.* 23, 100161. doi: 10.1016/j.cpb.2020.100161
- Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-Gonzalez, A., Eldridge, D. J., Bardgett, R. D., et al. (2018). A global atlas of the dominant bacteria found in soil. *Science* 359, 320–325. doi: 10.1126/science.aap9516
- de Lima, P. R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L., and de Araújo, J. M. (2012). Antibiotics produced by streptomycetes. *Braz. J. Infect. Dis.* 16 (5), 466–471.
- Edwards, J., Johnson, C., Santos-Medellin, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., et al. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U.S.A.* 112 (8), E911–E920. doi: 10.1073/pnas.1414592112
- Escudero-Martinez, C., Coulter, M., Alegria Terrazas, R., Foito, A., Kapadia, R., Pietrangeli, L., et al. (2022). Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nat. Commun.* 13 (1), 1–14. doi: 10.1038/s41467-022-31022-y
- French, E., Kaplan, I., Iyer-Pascuzzi, A., Nakatsu, C. H., and Enders, L. (2021). Emerging strategies for precision microbiome management in diverse agroecosystems. *Nat. Plants* 7 (3), 256–267. doi: 10.1038/s41477-020-00830-9
- Galdeano, D. M., de Souza Pacheco, I., Alves, G. R., Granato, L. M., Rashidi, M., Turner, D., et al. (2020). Friend or foe? relationship between *Candidatus liberibacter asiaticus* and *Diaphorina citri*. *Trop. Plant Pathol.* 45 (6), 559–571. doi: 10.1007/s40858-020-00375-4
- Ghannam, R. B., and Techtmann, S. M. (2021). Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Comput. Struct. Biotech.* 19, 1092–1107. doi: 10.1016/j.csbj.2021.01.028
- Ginnan, N. A., Dang, T., Bodaghi, S., Ruegger, P. M., Peacock, B. B., McCollum, G., et al. (2018). Bacterial and fungal next generation sequencing datasets and metadata from citrus infected with *Candidatus liberibacter asiaticus*. *Phytobiomes* 2 (2), 64–70. doi: 10.1094/PBIOMES-08-17-0032-A
- Ginnan, N. A., Dang, T., Bodaghi, S., Ruegger, P. M., McCollum, G., England, G., et al. (2020). Disease-induced microbial shifts in citrus indicate microbiome-derived responses to HLB across the disease severity spectrum. *Phytobiomes* 4, 375–387. doi: 10.1094/PBIOMES-04-20-0027-R
- Gmitter, F. G., and Hu, X. (1990). The possible role of yunnan, China, in the origin of contemporary citrus species (rutaceae). *Econ. Bot.* 44, 267–277. doi: 10.1007/BF02860491
- Gupta, S., Aga, D., Pruden, A., Zhang, L., and Vikesland, P. (2021). Data analytics for environmental science and engineering research. *Environ. Sci. Technol.* 55 (16), 10895–10907. doi: 10.1021/acs.est.1c01026
- Jat, S. L., Suby, S. B., Parihar, C. M., Gambhir, G., Kumar, N., and Rakshit, S. (2021). Microbiome for sustainable agriculture: a review with special reference to the corn production system. *Arch. Microbiol.* 203 (6), 2771–2793. doi: 10.1007/s00203-021-02320-8
- Killiny-Mansour, N. (2019). *Greening bacterium is now available in culture-so what's next?* (Florida, FL: EDIS).
- Lajoie, G., Maglione, R., and Kembel, S. W. (2020). Adaptive matching between phyllosphere bacteria and their tree hosts in a neotropical forest. *Microbiome* 8 (1), 1–10. doi: 10.1186/s40168-020-00844-7
- Lee, J. A., Halbert, S. E., Dawson, W. O., Robertson, C. J., Keesling, J. E., and Singer, B. H. (2015). Asymptomatic spread of Huanglongbing and implications for disease control. *Proc. Natl. Acad. Sci. U.S.A.* 112 (24), 7605–7610. doi: 10.1073/pnas.1508253112
- Li, J., Pang, Z., Trivedi, P., Zhou, X., Ying, X., Jia, H., et al. (2017). *Candidatus liberibacter asiaticus* encodes a functional salicylic acid (SA) hydroxylase that degrades SA to suppress plant defenses. *Mol. Plant Microbe In.* 30 (8), 620–630. doi: 10.1094/MPMI-12-16-0257-R
- Liu, H., Brettell, L. E., and Singh, B. (2020). Linking the phyllosphere microbiome to plant health. *Trends Plant Sci.* 25 (9), 841–844. doi: 10.1016/j.tplants.2020.06.003
- Liu, H., Sun, F., Peng, J., Shen, M., Li, J., and Dong, Y. (2022). Deterministic process dominated belowground community assembly when suffering tomato bacterial wilt disease. *Agronomy* 12 (5), 1024. doi: 10.3390/agronomy12051024
- Liu, Y., Zhu, A., Tan, H., Cao, L., and Zhang, R. (2019). Engineering banana endosphere microbiome to improve fusarium wilt resistance in banana. *Microbiome* 7 (1), 1–15. doi: 10.1186/s40168-019-0690-x
- Mahato, N., Sharma, K., Sinha, M., Dhyani, A., Pathak, B., Jang, H., et al. (2021). Biotransformation of citrus waste—I: production of biofuel and valuable compounds by fermentation. *Processes* 9 (2), 220. doi: 10.3390/pr9020220
- Nicholson, W. L. (2002). Roles of bacillus endospores in the environment. *Cell Mol. Life Sci.* 59 (3), 410–416. doi: 10.1007/s00018-002-8433-7
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., et al. (2020) *Vegan: community ecology package. r package version 2.6–2*. Available at: <https://CRAN.R-project.org/package=vegan>.
- Oyserman, B. O., Flores, S. S., Griffioen, T., Pan, X., van der Wijk, E., Pronk, L., et al. (2022). Disentangling the genetic basis of rhizosphere microbiome assembly in tomato. *Nat. Commun.* 13 (1), 1–16. doi: 10.1038/s41467-022-30849-9
- Paasch, B. C., and He, S. Y. (2021). Toward understanding microbiota homeostasis in the plant kingdom. *PLoS Pathog.* 17 (4), e1009472. doi: 10.1371/journal.ppat.1009472
- Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangel, J. L., et al. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. U.S.A.* 110 (16), 6548–6553. doi: 10.1073/pnas.1302837110
- Penyalver, R., Roesch, L. F., Piquer-Salcedo, J. E., Forner-Giner, M. A., and Alguacil, M. D. M. (2022). From the bacterial citrus microbiome to the selection of potentially host-beneficial microbes. *New Biotechnol.* 70, 116–128. doi: 10.1016/j.nbt.2022.06.002
- R Core Team (2020). *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: <https://www.R-project.org/>.
- Sare, A. R., Stouvenakers, G., Eck, M., Lampens, A., Goormachtig, S., Jijakli, M. H., et al. (2020). Standardization of plant microbiome studies: which proportion of the microbiota is really harvested? *Microorganisms* 8 (3), 342. doi: 10.3390/microorganisms8030342
- Sha, H., Lu, J., Chen, J., and Xiong, J. (2022). A meta-analysis study of the robustness and universality of gut microbiota-shrimp diseases relationship. *Environ. Microbiol.* 24 (9), 3924–3938. doi: 10.1111/1462-2920.16024
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCr: visualizing classifier performance in r. *Bioinformatics* 21, 3940–3941. doi: 10.1093/bioinformatics/bti623
- Sloan, W. T., Lunn, M., Woodcock, S., Head, I. M., Nee, S., and Curtis, T. P. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* 8, 732–740. doi: 10.1111/j.1462-2920.2005.00956.x
- Srivastava, A. K., Das, A. K., Jagannadham, P. T. K., Bora, P., Ansari, F. A., and Bhate, R. (2022). Bioprospecting microbiome for soil and plant health management amidst Huanglongbing threat in citrus: a review. *Front. Plant Sci.* 13, 858842. doi: 10.3389/fpls.2022.858842
- Thoms, D., Liang, Y., and Haney, C. H. (2021). Maintaining symbiotic homeostasis: how do plants engage with beneficial microorganisms while at the

- same time restricting pathogens? *Mol. Plant Microbe In.* 34 (5), 462–469. doi: 10.1094/MPMI-11-20-0318-FI
- Tian, G., Qiu, H., Li, D., Wang, Y., Zhen, B., Li, H., et al. (2022). Little environmental adaptation and high stability of bacterial communities in rhizosphere rather than bulk soils in rice fields. *Appl. Soil Ecol.* 169, 104183. doi: 10.1016/j.apsoil.2021.104183
- Torija, A. J., and Ruiz, D. P. (2015). A general procedure to generate models for urban environmental-noise pollution using feature selection and machine learning methods. *Sci. Total Environ.* 505, 680–693. doi: 10.1016/j.scitotenv.2014.08.060
- Trivedi, P., He, Z. L., Van Nostrand, J. D., Albrigo, G., Zhou, J. Z., and Wang, N. (2012). HLB alters the structure and functional diversity of microbial communities associated with citrus rhizosphere. *ISME J.* 6, 363–383. doi: 10.1038/ismej.2011.100
- Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., and Singh, B. K. (2020). Plant-microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* 18 (11), 607–621. doi: 10.1038/s41579-020-0412-1
- Vanwongerghem, I., Jensen, P. D., Dennis, P. G., Hugenholtz, P., Rabaey, K., and Tyson, G. W. (2014). Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J.* 8, 2015–2028. doi: 10.1038/ismej.2014.50
- Wang, N. (2019). The citrus huanglongbing crisis and potential solutions. *Mol. Plant* 12 (5), 607–609. doi: 10.1016/j.molp.2019.03.008
- Wang, N. (2020). A perspective of citrus HLB in the context of the Mediterranean basin. *J. Plant Pathol.* 102, 635–640. doi: 10.1007/s42161-020-00555-w
- Wang, N., Jin, T., Trivedi, P., Setubal, J. C., and Tang, J. (2015). Announcement of the international citrus microbiome (phytobiome) consortium. *J. Citrus Pathol.* 2, 1–2. doi: 10.5070/C421027940
- Wilhelm, R. C., van Es, H. M., and Buckley, D. H. (2022). Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biol. Biochem.* 164, 108472. doi: 10.1016/j.soilbio.2021.108472
- Wu, Y., Qu, M., Pu, X., Lin, J., and Shu, B. (2020). Distinct microbial communities among different tissues of citrus tree *Citrus reticulata* cv. *Chachiensis*. *Sci. Rep.* 10, 6068. doi: 10.1038/s41598-020-62991-z
- Xu, J., Zhang, Y., Zhang, P., Trivedi, P., Riera, N., Wang, Y., et al. (2018). The structure and function of the global citrus rhizosphere microbiome. *Nat. Commun.* 9 (1), 1–10. doi: 10.1038/s41467-018-07343-2
- Yao, Z., Yang, K., Huang, L., Huang, X., Qiuqian, L., Wang, K., et al. (2018). Disease outbreak accompanies the dispersive structure of shrimp gut bacterial community with a simple core microbiota. *AMB Express.* 8 (1), 1–10. doi: 10.1186/s13568-018-0644-x
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "All-species living tree project (LTOP)". taxonomic frameworks. *Nucleic Acids Res.* 42, D643–D648.
- Yu, W., Wu, J. H., Zhang, J., Yang, W., Chen, J., and Xiong, J. (2018). A meta-analysis reveals universal gut bacterial signatures for diagnosing the incidence of shrimp disease. *FEMS Microbiol. Ecol.* 94, fyy147. doi: 10.1093/femsec/fyy147
- Yuan, J., Wen, T., Zhang, H., Zhao, M., Penton, C. R., Thomashow, L. S., et al. (2020). Predicting disease occurrence with high accuracy based on soil macroecological patterns of *Fusarium* wilt. *ISME J.* 14 (12), 2936–2950. doi: 10.1038/s41396-020-0720-5
- Zhang, B., Zhang, J., Liu, Y., Shi, P., and Wei, G. (2018). Co-Occurrence patterns of soybean rhizosphere microbiome at a continental scale. *Soil Biol. Biochem.* 118, 178–186. doi: 10.1016/j.soilbio.2017.12.011
- Zhang, H., Wu, C., Wang, F., Wang, H., Chen, G., Cheng, Y., et al. (2022). Wheat yellow mosaic enhances bacterial deterministic processes in a plant-soil system. *Sci. Total Environ.* 812, 151430. doi: 10.1016/j.scitotenv.2021.151430
- Zhang, Y., Trivedi, P., Xu, J., Roper, M. C., and Wang, N. (2021). The citrus microbiome: from structure and function to microbiome engineering and beyond. *Phytobiomes J.* 5 (3), 249–262. doi: 10.1094/PBIOMES-11-20-0084-RVW
- Zhao, Z., Jiang, J., Zheng, J., Pan, Y., Dong, Y., Chen, Z., et al. (2022). Exploiting the gut microbiota to predict the origins and quality traits of cultured sea cucumbers. *Environ. Microbiol.* 24 (9), 3882–3897. doi: 10.1111/1462-2920.15972
- Zhou, F., Wu, X., Gao, Y., Fan, S., Zhou, H., and Zhang, X. (2022a). Diversity shifts in the root microbiome of cucumber under different plant cultivation substrates. *Front. Microbiol.* 13, 878409. doi: 10.3389/fmicb.2022.878409
- Zhou, L., Zhao, Z., Shao, L., Fang, S., Li, T., Gan, L., et al. (2022b). Predicting the abundance of metal resistance genes in subtropical estuaries using amplicon sequencing and machine learning. *Ecotox. Environ. Safe* 241, 113844. doi: 10.1016/j.ecoenv.2022.113844