



A Machine Learning Approach Reveals a Microbiota Signature for Infection with *Mycobacterium avium* subsp. *paratuberculosis* in Cattle

Sang-Mok Lee,^a Hong-Tae Park,^b Seojeong Park,^a Jun Ho Lee,^b Danil Kim,^c  Han Sang Yoo,^b Donghyuk Kim^a

^aSchool of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

^bDepartment of Infectious Diseases, College of Veterinary Medicine, Seoul National University, Seoul, Republic of Korea

^cDepartment of Farm Animal Medicine, College of Veterinary Medicine, Seoul National University, Seoul, Republic of Korea

Sang-Mok Lee and Hong-Tae Park contributed equally to this work. The author order was determined by their equal but graded contributions for this paper.

ABSTRACT Although *Mycobacterium avium* subsp. *paratuberculosis* (MAP) has threatened public health and the livestock industry, the current diagnostic tools (e.g., fecal PCR and enzyme-linked immunosorbent assay [ELISA]) for MAP infection have some limitations, such as inconsistent results due to intermittent bacterial shedding or low sensitivity during the early stage of infection. Therefore, this study aimed to develop a novel biomarker focusing on elucidating the gut microbial signature of MAP-positive ruminants, since the clinical signs of MAP infection are closely related to dysbiosis. 16S rRNA-based gut microbial community analysis revealed both a decrease in microbial diversity and the emergence of several distinct taxa following MAP infection. To determine the discriminant taxa diagnostic of MAP infection, machine learning-based feature selection and predictive model construction were applied to taxon abundance data or their transformed derivatives. The selected taxa, such as *Clostridioides* (formerly *Clostridium*) *difficile*, were used to build models using a support vector machine, linear support vector classification, *k*-nearest neighbor, and random forest with 10-fold cross-validation. The receiver operating characteristic-area under the curve (ROC-AUC) analysis of the models revealed their high accuracy, up to approximately 96%. Collectively, taxonomic signatures of cattle gut microbiotas according to MAP infection status could be identified by feature selection tools and applied to establish a predictive model for the infection state.

IMPORTANCE Due to the limitations, such as intermittent bacterial shedding or poor sensitivity, of the current diagnostic tools for Johne's disease, novel biomarkers are urgently needed to aid control of the disease. Here, we explored the fecal microbiota of Johne's disease-affected cattle and tried to discover distinct microbial characteristics which have the potential to be novel noninvasive biomarkers. Through 16S rRNA sequencing and machine learning approaches, a dozen taxa were selected as taxonomic signatures to discriminate the disease state. In addition, when constructing predictive models using relative abundance data of the corresponding taxa, the models showed high accuracy for classification, even including animals with subclinical infection. Thus, our study suggested novel noninvasive microbiological biomarkers that are robustly expressed regardless of subclinical infection and the applicability of machine learning for diagnosis of Johne's disease.

KEYWORDS *Mycobacterium avium* subsp. *paratuberculosis*, gut microbiota signature, machine learning-based predictive model, 16S rRNA sequencing, feature selection

M *ycobacterium avium* subspecies *paratuberculosis* (MAP) is an infectious pathogen causing Johne's disease (JD) or paratuberculosis (PTB) in ruminants (1). Animals infected with MAP suffer from chronic enteritis and diarrhea, which causes decreases

Editor Frederick S. B. Kibenge, University of Prince Edward Island

Copyright © 2023 Lee et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Han Sang Yoo, yoohs@snu.ac.kr, or Donghyuk Kim, dkim@unist.ac.kr.

The authors declare no conflict of interest.

Received 10 August 2022

Accepted 18 December 2022

Published 19 January 2023

in productivity, such as milk yield loss and infertility (1), and even leads to death (2). Although MAP infection can cause devastating effects, its long incubation period makes it difficult to eradicate (3). Clinically symptomatic animals are the “tip of the iceberg,” which indicates that many other individuals are in a silent or subclinical state, having acquired infection by herd transmission events (4). Additionally, MAP has been identified as a zoonotic pathogen (5) that plays a pivotal role in the pathogenesis of various diseases (6, 7). Therefore, diagnostic tools to detect MAP infection are crucial to prevent damage to the livestock industry and public health.

To overcome the limitation of high costs and time-consuming diagnoses based on bacterial culture, several culture-independent methods have been developed, such as fecal PCR (8) and enzyme-linked immunosorbent assay (ELISA) (9). Fecal PCR can detect the infectious state with high sensitivity, but there are still challenges, such as PCR inhibitors in feces (10), insufficient primer specificity (11), and low-level intermittent shedding during the subclinical stage (12). ELISA kits for MAP detection using serum or milk samples are available, but their low sensitivity during the early stage of infection remains a limitation. Hence, the need for alternative diagnostic tools with novel approaches for detecting MAP infection is urgent.

The gut of a eukaryotic host harbors a complex and dynamic population of various microorganisms, referred to as the microbiota. As the composition of the gut microbiota is modulated by various factors, such as diet, antibiotics, and disease state (13), many researchers have suggested novel biomarkers to indicate specific conditions by forming indices with populations of differentially abundant taxa (e.g., the *Firmicutes/Bacteroidetes* [F/B] ratio [14, 15]) or identifying closely associated bacteria (e.g., *Faecalibacterium prausnitzii* [16]). As the main route of MAP infection is fecal-oral transmission (1) and its primary clinical sign is granulomatous diarrhea, which is closely related to dysbiosis, investigating the distinct features of the gut microbiota of MAP-infected individuals may provide insight into biomarker discovery. However, the structure of the microbial community and the microbial signatures of JD are still poorly understood.

Machine learning (ML) has been widely applied in biological studies since the quantity of data generation began to rapidly increase with advances in next-generation sequencing technologies (17). After the explosion of biological data, data mining to find patterns and extract useful biological insights from multiple types of data sets has become a bottleneck (18). Various ML algorithms have demonstrated their usefulness in integrating heterogeneous biological data with a noisy nature (19). For example, ML-based diagnostic models for general dysbiosis states (20) or diseases such as Crohn’s disease (CD) (21) have succeeded in classifying abnormal states with high accuracy. Data regarding bacterial abundance in microbiotas are also used as input for ML algorithms to train the predictive model (22).

This study aimed to capture the microbial signature of MAP infection, thereby developing a classification model to aid diagnosis and discover novel MAP-associated biomarkers using microbial composition data. First, the microbial diversity and taxonomic profile in the gut microbiota of MAP-infected cattle were explored based on 16S rRNA sequencing. Subsequently, five feature selection tools, including ridge regression, LASSO, ElasticNet, Feature Selector, and the filter method, were used to determine closely correlated features using high-dimensional microbial abundance data. Furthermore, ML-based predictive models for MAP infection using linear support vector classification (LinearSVC), k-nearest neighbor (KNN), random forest, and support vector machine (SVM) were constructed, and their performances were compared. Taken together, the results predicted that taxonomic signatures of MAP-positive cattle could be identified by feature selection tools, and the ML classification model could predict the infection state with high accuracy using these signatures.

RESULTS

MAP infection modulates the gut microbiota of cattle in the direction of decreasing microbial diversity. To determine the impact of MAP infection on the gut microbial community and to determine distinct microbiological features reflecting

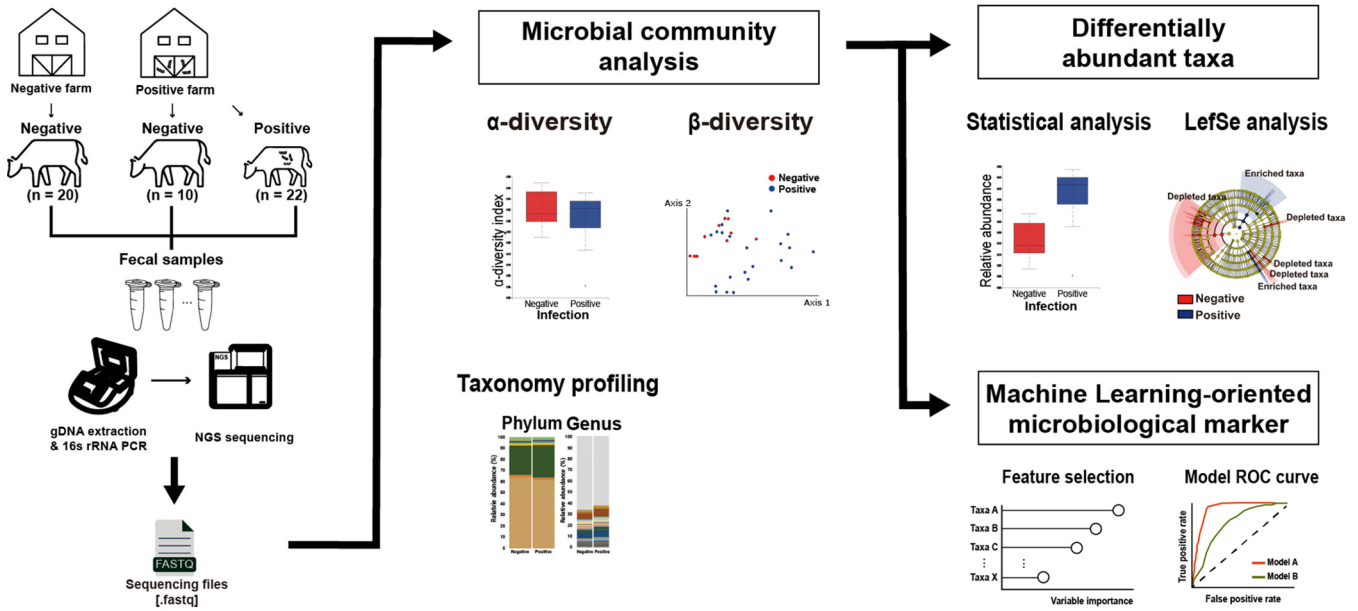


FIG 1 Overview of the analysis. The experiment was conducted using fecal samples from 52 cattle (negative, 30; positive, 22) divided into two groups by MAP infection. Microbial diversities (alpha and beta) and taxonomy profiles of the microbiotas in all samples were explored and compared by group. To determine the microbial features associated with MAP infection, significantly different taxa were identified by statistical analysis and LefSe analysis based on the relative abundances of each taxon. Several taxa that contributed to classifying MAP infection were selected using dimensionality reduction tools, and their discriminating potential for the MAP infection classifier was validated using ROC curve analysis.

infection, microbial community analysis using the fecal microbiotas of 22 MAP-positive and 30 MAP-negative cattle was conducted via 16S rRNA sequencing (Fig. 1). On average, $19,469 \pm 6,130$ paired-end reads were obtained for each sample from a total of 1,012,401 reads.

First, differences in various alpha and beta diversity indices related to MAP infection were investigated. The microbial richness (observed features; $P = 0.005$) (Fig. 2A) and diversity (Shannon’s index; $P < 0.001$) (Fig. 2B) were significantly decreased by MAP infection. Likewise, other alpha diversity indices for microbial evenness (Pielou’s evenness; $P < 0.001$) (see Fig. S2A in the supplemental material) and diversity (Faith’s phylogenetic diversity [PD] [$P < 0.001$] and Simpson’s index [$P < 0.001$]) (Fig. S2B and C) showed decreased patterns. Subsequently, principal coordinate analysis (PCoA) performed based on weighted UniFrac distances (Fig. 2C) and the distribution of the distances between samples in the MAP-negative and -positive groups (Fig. 2D) (permutational multivariate analysis of variance (PERMANOVA), 999 permutations; $P = 0.001$) revealed that the gut microbial communities were significantly altered in response to MAP infection. PCoA plots and distance distribution (PERMANOVA, 999 permutations; $P = 0.001$) based on unweighted UniFrac distances also showed apparent clustering of the microbiotas (Fig. S2). The analysis suggested that MAP infection diminished the microbial diversity of the gut microbiota in cattle.

Several microbial taxa showed distinct population changes in response to MAP infection. The relative abundance values of microbial taxa in each group were investigated. To identify taxa specific for MAP infection, linear discriminant analysis (LDA) effect size (LefSe) analysis was conducted by the MAP-negative/-positive group (Fig. 3A). Several differentially abundant taxa were identified at various taxonomic levels. The circular cladogram for LefSe (threshold, LDA score > 3.0) indicated that the classes *Clostridia* and *Bacteroidia* were the discriminant taxa of the MAP-negative and MAP-positive groups, respectively. Additionally, the amplicon sequence variants (ASVs) assigned to the class *Clostridia* and the order *Clostridiales* were significantly enriched in the MAP-negative groups, whereas the MAP-positive group was characterized by a significantly high abundance of ASVs assigned to the orders *Bacteroidales* and *Enterobacteriales*, the families *Bacteroidaceae* and *Enterobacteriaceae*, and the species *Clostridioides* (formerly *Clostridium*) *difficile*. The effects of the differential composition of the taxonomic profile caused by MAP

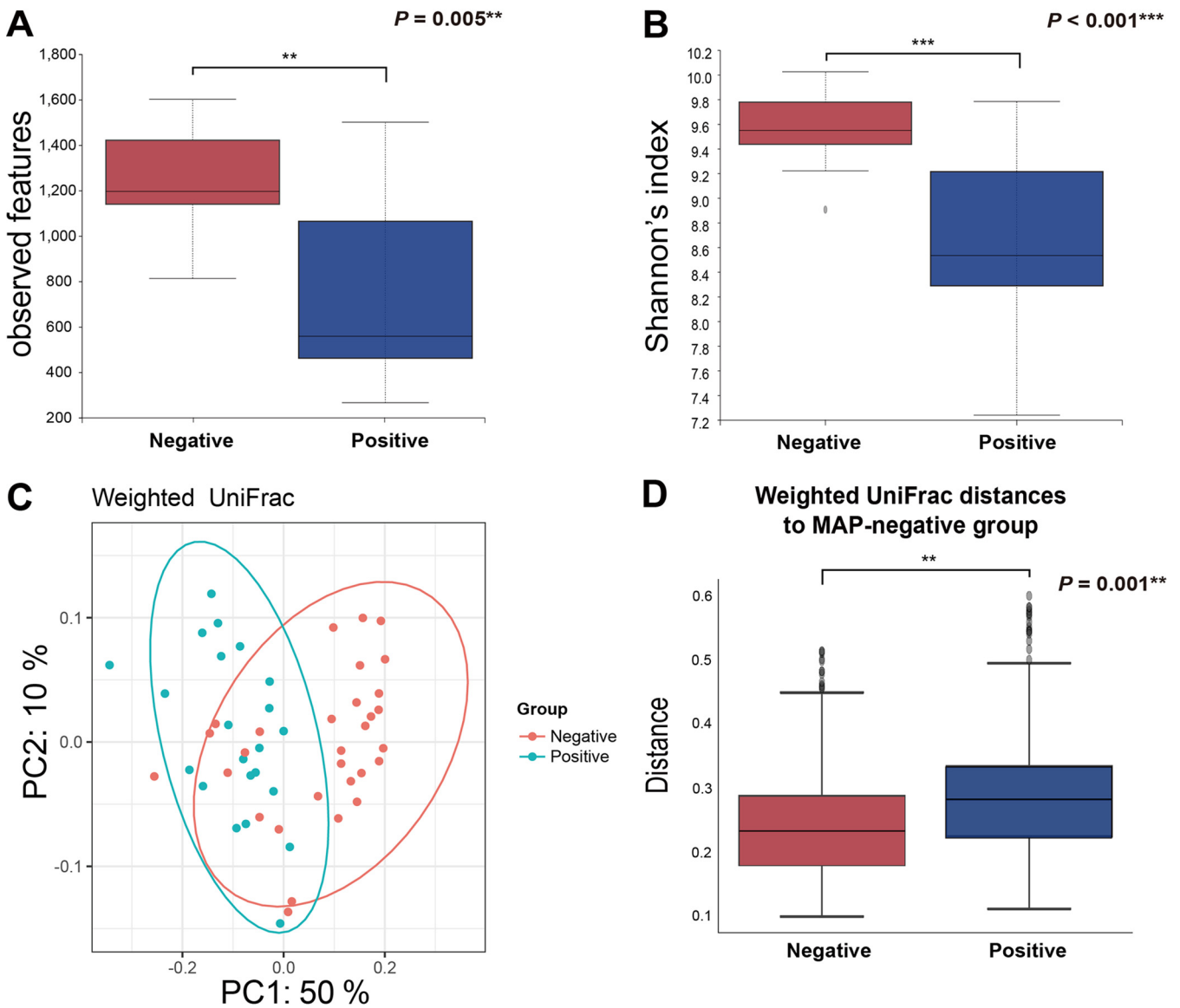


FIG 2 Changes in microbial diversity indices of fecal microbiotas by MAP infection. (A and B) Alpha diversity indices for richness (A) (observed features) and diversity (B) (Shannon's index). (C and D) Beta diversity indices based on weighted UniFrac distances visualized in the form of PCoA plots that demonstrate significant differences by their distances from each sample between groups.

infection on the metabolic function of the gut metagenome were also investigated. The abundances of Kyoto Encyclopedia of Genes and Genome (KEGG) pathways predicted by PICRUSt2 software were used to perform LefSe analysis to determine differentially abundant pathways for each group (Fig. S3A). Among the 25,961 ASVs, 7 were above the maximum NSTI (nearest-sequenced taxon index) cutoff of 2.0 and were therefore removed from the downstream analysis. Moreover, the bovine gut microbiome was also predicted using CowPI (23), a functional inference tool specific to rumen microbiomes (Fig. S3B). Notably, pathways such as "metabolism" and "amino acid metabolism" were commonly predicted at significantly higher levels in the MAP-negative group, whereas "environmental information processing," "membrane transport," and "transporters" were commonly predicted at significantly higher levels in the MAP-positive group (threshold, LDA score > 3.0).

Feature selection identified microbial taxa closely correlated with MAP infection. To investigate major informative taxa that were closely correlated with MAP infection and to develop a predictive model, five feature selection methods (ridge regression, LASSO, ElasticNet, Feature Selector, and the filter method) were applied to

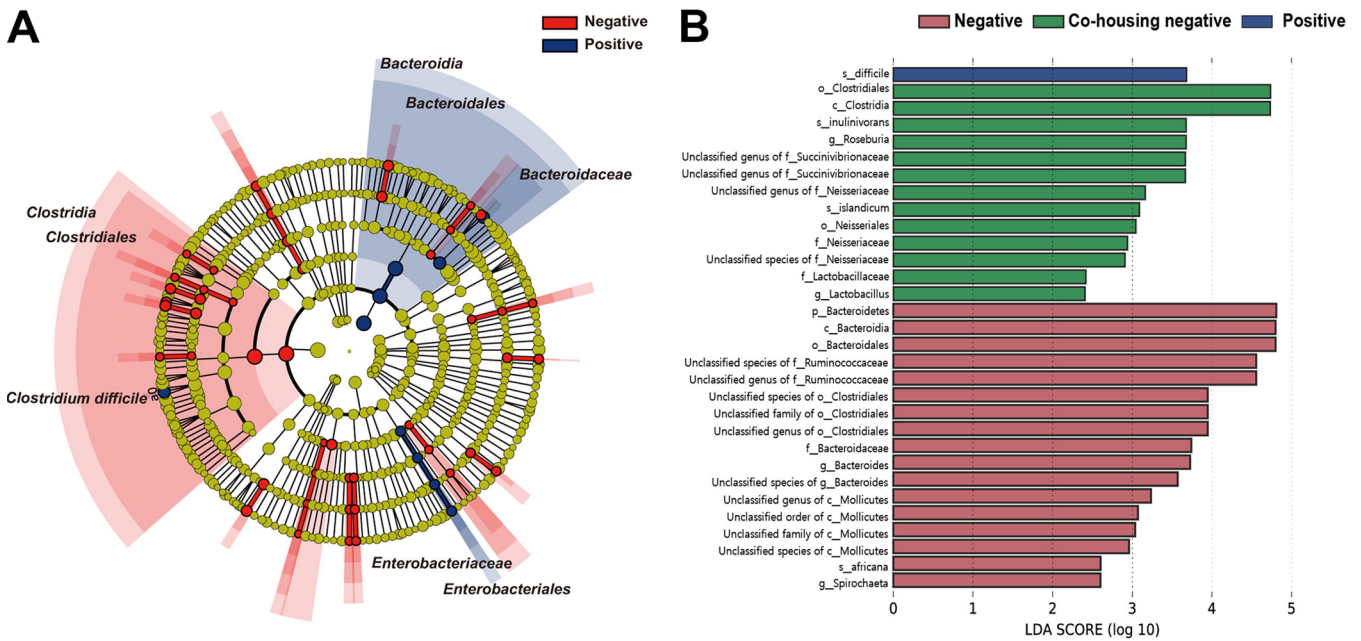
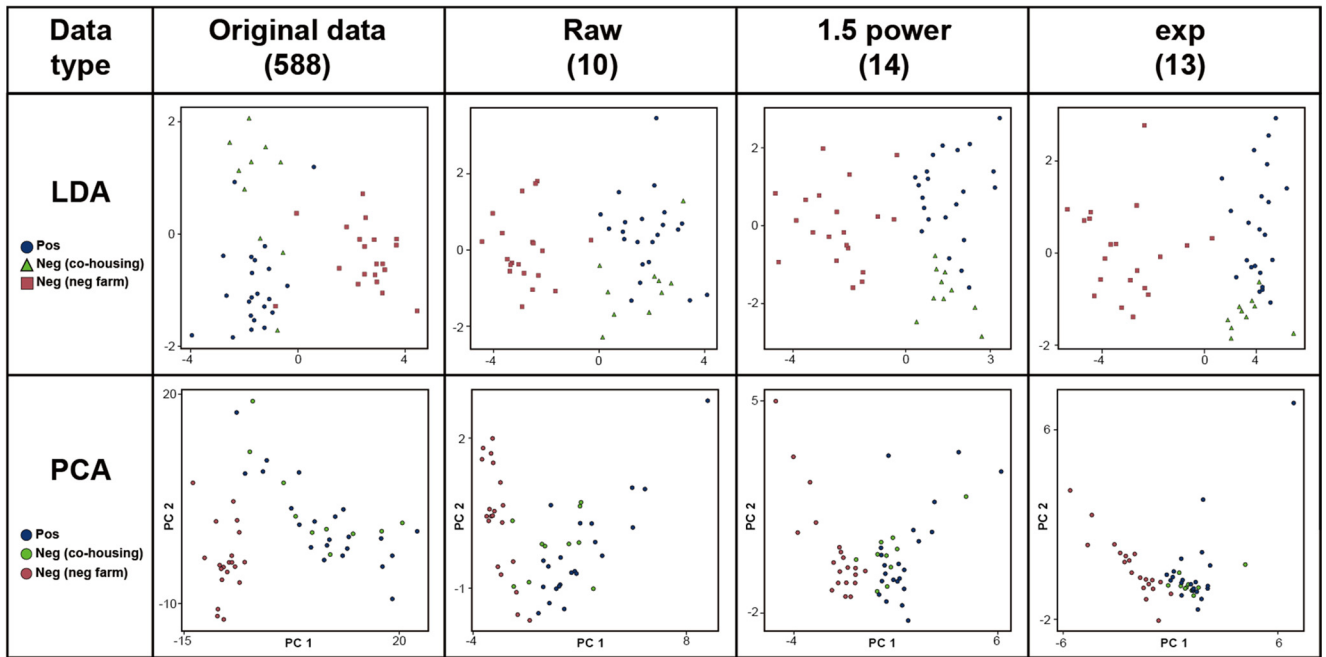


FIG 3 LefSe analysis for determining differentially abundant taxa by MAP infection. (A) Cladogram generated by LefSe demonstrating differential abundances of taxa (LDA > 3.0); (B) bar graph showing LDA scores for the negative-farm group, the negative group cohoused on the positive farm, and the positive group (positive farm). Only taxa meeting the LDA significance threshold are shown in the bar chart (LDA > 2.0).

the data set of relative abundance values for all taxa (referred to as the raw set). Two additional data sets, each of which consists of original values to the power of 1.5 (1.5 power set) and constant e to the original values (Exp set), were prepared to increase the variances of the original data set and were used together for further analysis.

First, several taxa with constant and quasi-constant values among 52 samples were removed from each data set. The threshold for the quasi-constant was set to 0.00005 based on the variance among the samples. By that, from 588 taxa, the number of remaining taxa for the raw set was 478, while those for the 1.5 power and Exp sets were 387 and 480, respectively (Fig. S4). Subsequently, using the five feature selection methods, feature selection of each data set was conducted from the remaining taxa to classify individuals into three groups: MAP positive, MAP negative cohoused on the positive farm, and MAP negative on the negative farm. LDA and principal-component analysis (PCA) plots were generated from the selected features of each case to visualize its discriminating pattern. The clustering performances among them were quantitatively compared using the Calinski-Harabasz index and the silhouette score (Table S2). In the cases of the raw and 1.5 power sets, ElasticNet showed the highest scores for both indicators using 143 and 124 features, while LASSO showed the highest scores in the Exp set with 124 features. Meanwhile, although Feature Selector mostly showed the lowest scores for all cases except the Calinski-Harabasz index of the raw set, it showed apparent clusters by groups with only 12, 13, and 13 features in the raw, 1.5 power, and Exp sets, respectively (Fig. 4A). That is, only approximately a dozen features were required to explain 98% of the whole data set, since the parameter threshold for the cumulative importance of Feature Selector was 0.98. Indeed, distributions of dots in the PCA plots based on Feature Selector-originated features were similar to those of the original data set. The selected important features for each data type were plotted (Fig. 4B). Surprisingly, the most important feature for all data types was the relative abundance data of *C. difficile*, with normalized importance values of 0.475, 0.190, and 0.364 for the raw, 1.5 power, and Exp sets, respectively. The values for *Bacilli* (raw, 0.017; 1.5 power, 0.121; Exp, 0.056) and *Ruminococcus* (raw, 0.068; 1.5 power, 0.033; Exp, 0.021) were commonly observed from all data types as well. In addition, for the raw set, *Clostridiaceae* (0.119), *Peptostreptococcaceae* (0.051), and *Ruminococcaceae*

A



B

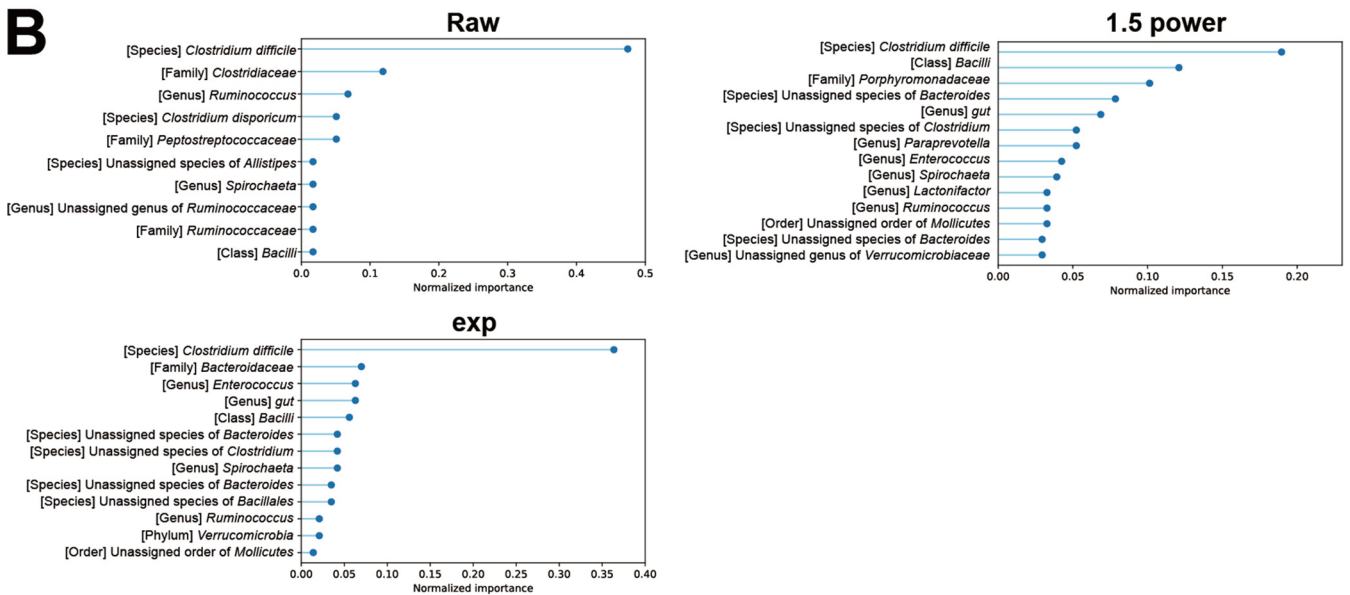


FIG 4 Feature selection of microbial abundance data sets by Feature Selector with three different types of transformed values. (A) LDA and PCA plots for the original relative abundance data set and different transformation types (raw, 1.5 power, and Exp). Blue, MAP positive; green, MAP negative cohoused on a positive farm; red, MAP negative on a negative farm. The values in parentheses indicate the number of selected features. (B) Variable-importance plots for the selected microbial features of each transformation type.

(0.017) at the family level, *Spirochaeta* (0.017) and an unassigned genus of *Ruminococcaceae* (0.017) at the genus level, and *Clostridium disporicum* (0.051) and an unassigned species of *Allistipes* (0.017) at the species level were additionally selected. For the 1.5 power set, these features were 1 order (unassigned order of *Mollicutes*, 0.033), 1 family (*Porphyromonadaceae*, 0.101), 6 genera (*Lactonifactor*, 0.033; unassigned genus of *Verrucomicrobiaceae*, 0.029; *Paraprevotella*, 0.052; *gut*, 0.069; *Enterococcus*, 0.042; *Spirochaeta*, 0.039), and 3 species (unassigned species of *Clostridium*, 0.052; two unassigned species of *Bacteroides*, 0.029 and 0.078). For Exp, these features were 1 phylum (*Verrucomicrobia*, 0.021), 1 order (unassigned order of

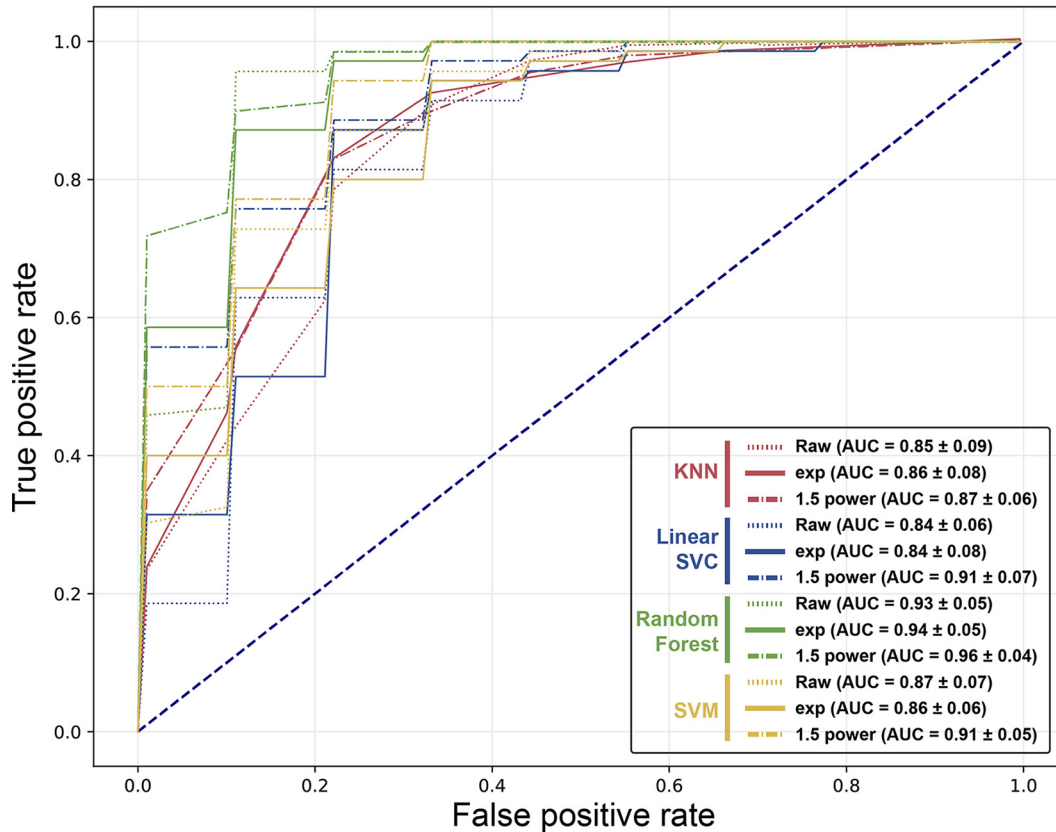


FIG 5 ROC curves and AUC for MAP infection-predictive models. The models were constructed using four different algorithms (KNN, LinearSVC, random forest, and SVM) with three types of data. The numbers in parentheses indicate the means and standard deviations (SD) of the AUC for each case. All ROC curves and their AUC values were averaged over 10 repetitions of 10-fold cross-validation.

Mollicutes, 0.014), 1 family (*Bacteroidaceae*, 0.070), 3 genera (gut, 0.063; *Enterococcus*, 0.063; *Spirochaeta*, 0.042), and 4 species (unassigned species of *Bacillales*, 0.035; unassigned species of *Clostridium*, 0.042; two unassigned species of *Bacteroides*, 0.035 and 0.042). Considering the number of selected features and the clustering performances among tools, Feature Selector was chosen to build a predictive model for MAP infection.

The classification model based on the selected features has the potential to be a good predictor of MAP infection. Using the set of selected features by Feature Selector, machine learning models for classification of the infection state were designed and their performances were investigated. The models were built by applying four different algorithms: KNN ($k = 3$), LinearSVC, random forest classifier, and SVM. The data were split in a ratio of 80:20 for training and testing purposes for the algorithms with 10-fold cross-validation. To assess the performances, the precision, recall, F1 score, and receiver operating characteristic-area under the curve (ROC-AUC) among the models were monitored (Fig. 5 and Table 1). In terms of predictive performance, it was observed that all indices reached at least 0.75. Notably, the best prediction results were achieved with random forest. When focusing on the AUC values, the values of 1.5 power were higher than the other data for all algorithms. Collectively, the best performance was obtained with random forest using 1.5 power data with an AUC of 0.96 for its mean value of 10-fold cross-validation. The model accuracy and AUC values of random forest models combined with the diagnosis results of fecal PCR and serum ELISA were additionally investigated, although there was no significant difference among the cases (Table S3).

TABLE 1 Classification performance metrics for each MAP infection-predictive model^a

Algorithm	Data set	Precision	Recall	F1 score	AUC
KNN ($k = 3$)	Raw	0.81 ± 0.11	0.79 ± 0.11	0.79 ± 0.11	0.86 ± 0.09
	1.5 power	0.80 ± 0.06	0.78 ± 0.05	0.77 ± 0.06	0.87 ± 0.06
	Exp	0.83 ± 0.11	0.82 ± 0.11	0.82 ± 0.11	0.86 ± 0.08
LinearSVC	Raw	0.76 ± 0.11	0.75 ± 0.10	0.75 ± 0.10	0.84 ± 0.06
	1.5 power	0.86 ± 0.09	0.84 ± 0.09	0.84 ± 0.09	0.91 ± 0.07
	Exp	0.78 ± 0.06	0.76 ± 0.07	0.75 ± 0.07	0.84 ± 0.08
Random forest	Raw	0.90 ± 0.07	0.89 ± 0.08	0.89 ± 0.08	0.93 ± 0.05
	1.5 power	0.88 ± 0.05	0.84 ± 0.08	0.84 ± 0.08	0.96 ± 0.04
	Exp	0.88 ± 0.07	0.86 ± 0.07	0.85 ± 0.08	0.94 ± 0.05
SVM	Raw	0.80 ± 0.09	0.77 ± 0.08	0.77 ± 0.08	0.87 ± 0.07
	1.5 power	0.82 ± 0.07	0.80 ± 0.08	0.79 ± 0.09	0.91 ± 0.05
	Exp	0.80 ± 0.04	0.79 ± 0.04	0.79 ± 0.05	0.86 ± 0.06

^aFor each experiment, the precision, recall, F1 score, and AUC value of the ROC curves were considered to quantify the performance. Values are means and SD for the predictive model that applied 10-fold cross-validation (training set, $n = 49$; testing set, $n = 5$) based on the labeled information for each sample.

DISCUSSION

Since MAP has been suspected as a productivity-reducing and/or zoonotic agent of various diseases in both ruminants and humans (24–26), accurate and rapid detection of its infection is crucial for controlling MAP-related diseases. Although several diagnostic tools, such as bacterial culture, fecal PCR, and ELISA, have been used to identify MAP infection, the alternative methods are still needed due to their time-consuming or false-negative nature. Meanwhile, the representative symptoms of the infection, such as chronic diarrhea and impairment of nutrition absorption, occur in the intestine, and the microbial community living there may be intrinsically involved in the occurrence, symptoms, and outcome of the infection. In this study, the fecal microbiota of MAP-infected cattle was investigated, and the prediction of MAP infection was carried out with machine learning models using the relative abundance values of the assigned microbial taxa.

The fecal microbiotas of MAP-infected cattle revealed significant changes in microbial richness, diversity, microbial taxon composition, and predicted metagenomes compared to those of noninfected cattle. Considering that numerous studies have reported that the microbial diversity in microbiotas is a representative characteristic of gut health status (27–29), the decreased values of overall alpha diversity indices and distinct clusters corresponding to MAP infection imply that the pathogen may induce dysbiosis. Additionally, several taxa showed distinct modulation of their population by the pathogen. Indeed, several studies reported that some taxa have significant changes in their abundance in the gut of MAP-infected individuals. For instance, *C. difficile*, *Bacilli*, and *Ruminococcus* were the representative taxa identified as distinct taxa according to MAP infection by feature selection. The close association of *C. difficile* with MAP infection was confirmed by LEfSe, by the fact that the bacterium was differentially abundant in the MAP-infected group, and by the results selected as the most important variables by feature selection. Interestingly, it was reported that both MAP and *C. difficile* provoked CD, whose symptoms are similar to those of JD (30, 31). Moreover, it was reported that the genus *Clostridium* showed a positive correlation with histopathology scores in MAP-infected calves (32).

In the case of *Bacilli*, although there are few studies investigating the relationship between *Bacilli* and JD, the increased population (percent relative abundance) of its lower taxonomic levels, such as the family *Bacillaceae* (Table S1) (negative, 0.68 ± 0.93 ; positive, 2.80 ± 1.89 ; Mann-Whitney U test, $P < 0.001$), was reported in the case of dextran sulfate sodium-induced colitis (33).

Meanwhile, *Ruminococcus*, belonging to the family *Lachnospiraceae*, was significantly enriched in the MAP-infected group (Table S1) (negative, 0.55 ± 0.85 ; positive, 2.24 ± 1.41 ; Mann-Whitney U test, $P < 0.001$). This genus has been reported for its

mucin-degrading ability, and several murolytic species belonging to the genus were enriched in the gut of CD and ulcerative colitis patients (34, 35). Therefore, it is suggested that these three taxa, which were selected by feature selection using all three different values of relative abundance and their transformants, may be the keystone taxa for MAP infection.

In addition, other taxa reported as being specific to MAP infection were also observed to have distinct modulation in this study. A previous study reported that a logistic model was built using four distinct taxa to distinguish the fecal microbiota of MAP-infected calves from the noninfected group (32). Similar to that study, *Paraprevotellaceae* were significantly enriched in MAP-infected animals (negative, 1.09 ± 0.82 ; positive, 1.66 ± 0.62 ; Mann-Whitney U test, $P < 0.01$) and positively correlated with the ELISA sample/positive (S/P) ratio (Pearson's correlation coefficient [r] = 0.670, $P < 0.001$) in this study. Although there was no significance in statistical analyses, *Faecalibacterium* was detected in only four individuals in the MAP-infected group and was depleted in all noninfected individuals. *Akkermansia* had a tendency to decrease with MAP infection (negative, 1.00 ± 0.64 ; positive, 0.75 ± 0.92 ; Mann-Whitney U test, $P = 0.0584$). In the case of *Planococcaceae*, no population was observed in any MAP-infected animals in this study. In addition, the genera *Alistipes* (negative, 4.65 ± 1.52 ; positive, 2.94 ± 0.86 ; Mann-Whitney U test, $P < 0.001$) and *Paraprevotella* (negative, 0.13 ± 0.15 ; positive, 0.08 ± 0.11 ; Mann-Whitney U test, $P < 0.001$) were significantly decreased by MAP infection in this study, which is consistent with other studies (36, 37), whereas the overrepresentation of *Firmicutes* (negative, 56.18 ± 7.09 ; positive, 62.83 ± 7.33 ; Mann-Whitney U test, $P < 0.01$), *Enterococcus* (negative, 0.03 ± 0.06 ; positive, 0.07 ± 0.05 ; Mann-Whitney U test, $P < 0.001$), and *Streptococcus* (negative, 0.03 ± 0.09 ; positive, 0.08 ± 0.12 ; Mann-Whitney U test, $P < 0.05$) in the MAP-infected group was observed, which was also reported for MAP-infected animals and humans (37, 38). In the case of *Actinobacteria*, there is controversy regarding its population change due to MAP infection. In this study, there was a significant increase in the MAP-infected group, although the highest value for its relative abundance was under 3% (negative, 0.25 ± 0.40 ; positive, 0.79 ± 0.62 ; Mann-Whitney U test, $P < 0.001$), contrary to the result of up to 30% abundance in other studies (36, 39). Other dysbiosis-associated taxa identified in CD patients or diarrheic calves, such as *Enterobacteriaceae* (negative, 0.00 ± 0.1 ; positive, 0.11 ± 0.21 ; Mann-Whitney U test, $P < 0.001$) and *Fusobacteriaceae* (detected in only two individuals in the MAP-positive group), were also overrepresented or detected only in MAP-positive groups, while the underrepresentation of *Porphyromonadaceae* (negative, 0.28 ± 0.13 ; positive, 0.17 ± 0.18 ; Mann-Whitney U test, $P < 0.05$) was also observed (40, 41).

Microbial composition-based metagenome prediction suggested that MAP infection might modulate the gut metagenome to upregulate pathways for sensing and responding to the extracellular environment. Furthermore, the pathway representing amino acid-related metabolism was downregulated by the pathogen. Likewise, it was reported that there was a significant alteration in the metabolism of amino acids within the MAP-infected group (32). These results suggest that the biosynthesis and degradation of various amino acids may be crucial factors in dysbiosis, such as JD. Collectively, the microbiota perturbation associated with MAP infection and its subsequent metagenomic modulation corresponded well to the aforementioned previous studies. Thus, it is likely that the microbial community of MAP-infected cattle in this study might accurately reflect the general topological state of gut microbiotas infected with MAP and that, importantly, the microbiota signature identified by feature selection is reliable as well. This signature can serve as a biomarker to compensate for the drawbacks of existing tools, since the microbial community may maintain traces of MAP infection during the latent period or intermittent shedding.

While the best scenario of the predictive model based on the microbiota signature resulted in a high AUC value of approximately 0.96, a combination of the signature and other diagnostic results, such as the cell number of MAP estimated by fecal PCR and the ELISA S/P ratio, showed a decrease in the values of model accuracy and AUC (Table S3). This might be related to the low sensitivity of traditional diagnosis methods.

Indeed, 50% of individuals in the positive group (11/20) showed low ELISA S/P ratios under a threshold of 50 (Table S1). The low sensitivity of ELISA using serum or milk samples has been pointed out as a limitation during the early stage of MAP infection (42, 43). Conversely, seven individuals in the positive group were not identified by fecal PCR, while the ELISA S/P ratio was over the threshold (Table S1). This result may be caused by the variable intermittent shedding in feces at different time points, which was already reported in other studies (44, 45). Furthermore, there were cases (e.g., sample 181.2, from the same animal as samples 181 and 181.3, and sample 188.2, from the same animal as 188 but collected in a different year) that continued to be judged MAP positive because at least one diagnostic tool identified them as positive, even though the diagnostic result for an individual was negative (Table S1). These inconsistent patterns of diagnostic results may weaken the discriminative power. These findings suggest that machine learning using microbial population data as input can be a novel approach to develop novel noninvasive biomarkers, thereby compensating for the low sensitivity of current diagnostic methods due to intermittent shedding or subclinical infection. Meanwhile, it was not possible to detect the abundance of MAP by 16S rRNA sequencing, which is due to its low proportion in the gut microbiota and the inadequate efficiency of genomic DNA extraction from lipopeptide-coated MAP (46, 47).

We acknowledge several limitations in our study. First, we used fecal samples, so the bacterial population of ileal mucosa in which MAP proliferates could not be directly identified (48). However, the purpose of this study was to discover a novel noninvasive biomarker for MAP infection that can be practically used on farms without the need for sacrificing animals. Fortunately, the microbiome signature identified in this study showed high accuracy for predicting the MAP infection state. Second, only two sites were used to collect the samples. Since the environment, including diet, has a great impact on the structure of the gut microbiota (49), differences among farms may contribute to the generation of distinct clusters of microbiotas. Indeed, it was observed that the cattle microbiota was clearly distinguished according to the farm in this study (Fig. S5) (PERMANOVA, 999 permutations; $P = 0.001$ for both unweighted and weighted UniFrac distances). Nonetheless, the microbiota signature found by statistical analyses and feature selection was trustworthy, since negative-farm samples and samples from negative animals cohoused on the positive farm were combined and then compared with those from the MAP-infected group. Although samples were collected from animals with various ages, parities, lactation periods, and breed types (Table S1), further study using large quantities of novel samples obtained from multiple sites to minimize the effect of within-animal variability is needed to validate the robustness of the microbiota signature and its derived predictive model, which was made using a relatively small number of samples in this study. Moreover, comparing distinct microbial features of the gut microbiome of MAP-infected cattle and those of other enteric infections, such as colibacillosis or salmonellosis, will provide an opportunity to clearly investigate the taxonomic signatures for MAP-specific pathology.

Last, although machine learning has broadened the current limited understanding of the complexity of microbiotas by detecting informative patterns in the microbial community system, the overfitting issue of the model should be pointed out. The generalizability of the model cannot be verified, since to our current knowledge there are no available sequence read archives of MAP-infected animals with accurate metadata. This issue may be solved by additive 16S rRNA sequencing of novel samples. Nevertheless, several studies on the application of machine learning models for disease diagnosis have been already reported (50, 51). In particular, machine learning with bacterial taxon information regarding the gut microbiome showed its potential for detecting cardiovascular disease (52). Likewise, the distinct microbial taxa identified by machine learning without human subjectivity and the subsequent generation of predictive models in this study showed another case for a microbiome-based

machine learning approach for diagnostic screening, although comparative studies with large-scale follow-up experiments are needed.

In summary, the microbiota signature of MAP-infected cattle was investigated using both statistical analyses and machine learning algorithms. The results indicate that several specific microbial taxa that distinguish the infection state have the potential to be noninvasive biomarkers for classifying MAP infection to support the current diagnostic tools. In addition, the machine learning-based investigation of major features in the microbiota can be applied to other biomarker discovery studies for prophylactic or diagnostic use.

MATERIALS AND METHODS

Sample collection. All specimens were obtained from farms that were referred to our laboratory for diagnostic testing for Johne's disease. ELISA and fecal real-time PCR tests were used to cross-validate, and the diagnostic tests were conducted a total of three times, periodically from 2019 to 2021. To determine the infection state of animals, the animals were labeled negative only when all tests were consistently negative, and those whose tests were positive at least once were considered infected. The negative farm was initially selected because no cases of MAP infection had been reported in the quarantine system of South Korea, and additional tests were conducted for every new animal that came to the farm. Based on the results, two farms, positive and negative, were selected for this study, and a total of 52 animals were selected for analysis according to their infection status. Twenty-two dairy cows that were positive on diagnostic tests along with 10 negative cows with evenly distributed ages and parities were selected on the positive farm. On the negative farm, a total of 20 negative animals, 10 Korean native cattle (Hanwoo) and 10 Holstein dairy cows, were selected using the same criteria.

Infection with MAP was diagnosed by a commercial ELISA kit (Idexx Laboratories Inc., Westbrook, ME, USA) and a fecal real-time PCR assay that was optimized in our laboratory. ELISA was performed using serum samples according to the manufacturer's instructions. For DNA extraction for real-time PCR, 2 g of feces was mixed with 35 mL of distilled water (DW), and the upper 5 mL was used after standing for 30 min. Fecal DNA extraction was conducted using a fecal microbe DNA extraction kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. Two target genes (*IS900* and *ISMap02*) were used for real-time PCR diagnosis. Real-time PCR was performed as previously described (53).

DNA extraction and sequencing. The feces used for diagnosis were delivered to the laboratory in a refrigerated state and then stored at -80°C . After diagnostic testing, fecal samples selected for sequencing were subjected to DNA extraction. DNA extraction was conducted using a DNeasy PowerSoil kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The extracted DNA was quantified using Quant-IT PicoGreen (Invitrogen). The DNA libraries were constructed according to the Illumina 16S metagenomic sequencing library protocols. Briefly, the V3/V4 region of the bacterial 16S rRNA gene was amplified from 2 ng of input genomic DNA (gDNA) using Herculase II fusion DNA polymerase (Agilent Technologies, Santa Clara, CA) with following primer pair: V3-F, 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGC CTACGGGNGGCWGCAG-3', and V4-R, 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGG TATCTAATCC-3'. Then, additional PCR amplification with the Nextera XT indexed primer was conducted to construct the final library. After every PCR step, purification was conducted using AMPure beads (Agencourt Bioscience, Beverly, MA). The quality of the final product was monitored with a qPCR quantification protocol guide (KAPA library quantification kits for Illumina sequencing platforms) for its quantification and the TapeStation D1000 ScreenTape (Agilent Technologies, Waldbronn, Germany) for its qualification. Finally, the amplicons were sequenced on an Illumina MiSeq 2 × 300-bp paired-end sequencing platform (Macrogen, Daejeon, South Korea).

Microbial community analysis. The microbial community was analyzed mainly using the QIIME (Quantitative Insights Into Microbial Ecology) 2 v2021.8 pipeline (54). Raw sequence reads were denoised and ASV tables were generated using DADA2 (55). Taxonomic assignment was conducted using a pretrained naive Bayes classifier on the GreenGenes database, and the relative abundances of bacterial taxa were expressed as percentages of total 16S rRNA sequences. For the microbial diversity (alpha and beta), the feature tables were rarefied to even depths based on the minimum number of features among the samples. The microbial diversity of the samples (alpha diversity) was determined using Pielou's evenness, observed features, Faith's phylogenetic diversity, Shannon's index, and Simpson's index. Principal-coordinate analysis was performed based on weighted and unweighted UniFrac distances, and the differences in the sample distances between groups were evaluated using PERMANOVA (56). LEfSe analysis was used to identify differentially abundant taxa between the groups with LDA scores of >3.0 and P values of <0.05 (57).

Subsequent prediction of the functions of the microbial communities was conducted using PICRUSt2 (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States, version 2.0) (58) and CowPI (23) to predict the functional profile of the microbial communities based on the 16S rRNA gene sequences obtained. Since the web server for CowPI is unavailable, the tool was reconstructed using precalculated files deposited on Zenodo (<https://zenodo.org/record/1252858>). The predicted metagenomes were obtained from the precalculated KEGG orthologs and classified in a hierarchy using the KEGG pathway metadata. LEfSe analysis was performed using a threshold of an LDA score of >3.0 and a P value of <0.05 to identify differentially abundant KEGG pathways.

Feature selection. To reduce the dimensionality of the data and select the most significant features in the data set of relative abundance for microbial taxa, five algorithms/tools (ridge regression, LASSO, ElasticNet, Feature Selector [<https://github.com/WillKoehrs/en/feature-selector>], and the filter method) were used. The parameters for each regularization (ridge regression, LASSO, and ElasticNet) were optimized by built-in cross-validation in the Python package Scikit-learn v1.1.1. The hyperparameter values (alpha) for ridge regression, LASSO, and ElasticNet were 0.1, 1.9, and 0.9, respectively. In the case of the filter method, the corr() method of the Python package pandas was used. The threshold value for the correlation was set to 0.5. The parameters for Feature Selector were optimized as follows: missing_threshold, 0.6; correlation_threshold, 0.98; task, classification; eval_metric, auc; and cumulative_importance, 0.95.

Construction of the machine learning-based classification model. To build classification models that differentiated MAP-infected individuals from negative individuals cohoused on the positive farm and those on the negative farm, the relative abundance values of the features selected by Feature Selector were used. Four different machine learning algorithms were implemented—LinearSVC (C = 1), KNN (n_neighbors = 3), random forest (n_estimators = 100), and SVM (kernel = linear; C = 1)—using “LinearSVC,” “KNeighborsClassifier,” “RandomForestClassifier,” and “SVC” in the Scikit-learn package. The discriminating performances of the models were measured and compared by values for accuracy and AUC.

Statistical analysis. Statistical analyses were conducted using the two-tailed Mann-Whitney U test for 2-group comparisons and the Kruskal-Wallis test for multiple-group comparisons.

Data availability. All of the sequence data obtained from the 52 samples in this study were deposited in the sequence read archives of the NCBI under accession number [SRP363929](https://www.ncbi.nlm.nih.gov/sra/SRP363929).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS

S.-M.L. and H.-T.P. designed and performed the experiments. S.-M.L. and S.P. conducted microbial community analysis via 16S rRNA sequencing and constructed machine learning models. H.-T.P., J.H.L., and Danil Kim acquired the fecal and blood samples for analysis and conducted fecal PCR and ELISA. S.-M.L. and H.-T.P. drafted the paper. S.-M.L., H.-T.P., S.P., J.H.L., Danil Kim, H.S.Y., and Donghyuk Kim edited the manuscript. H.S.Y. and Donghyuk Kim supervised the project.

This work was carried out with the support of the Strategic Initiative for Microbiomes in Agriculture and Food (no. IPET918020-4); the Agriculture, Food and Rural Affairs Convergence Technologies Program for Educating Creative Global Leader (no. 320005-4); the Ministry of Agriculture, Food and Rural Affairs, BK21 FOUR Future Veterinary Medicine Leading Education and Research Center and Research Institute for Veterinary Science, Seoul National University, Seoul, Republic of Korea; and the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) (2022M3A9I5018934).

We appreciate Toby J. Wilkinson, a corresponding author of the CowPI paper (23), for sharing his source code and giving advice for reconstructing CowPI in a local environment.

We 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.

REFERENCES

- Rathniah G, Zinniel DK, Bannantine JP, Stabel JR, Grohn YT, Collins MT, Barletta RG. 2017. Pathogenesis, molecular genetics, and genomics of *Mycobacterium avium* subsp. *paratuberculosis*, the etiologic agent of Johne's disease. *Front Vet Sci* 4:187. <https://doi.org/10.3389/fvets.2017.00187>.
- Donat K, Soschinka A, Erhardt G, Brandt HR. 2014. Paratuberculosis: decrease in milk production of German Holstein dairy cows shedding *Mycobacterium avium* ssp. *paratuberculosis* depends on within-herd prevalence. *Animal* 8: 852–858. <https://doi.org/10.1017/S1751731114000305>.
- Cechova M, Beinhauerova M, Babak V, Kralik P. 2022. A viability assay combining palladium compound treatment with quantitative PCR to detect viable *Mycobacterium avium* subsp. *paratuberculosis* cells. *Sci Rep* 12:4769. <https://doi.org/10.1038/s41598-022-08634-x>.
- Dow CT, Alvarez BL. 2022. *Mycobacterium paratuberculosis* zoonosis is a One Health emergency. *Ecohealth* 19:164–174. <https://doi.org/10.1007/s10393-022-01602-x>.
- Kuenstner L, Kuenstner JT. 2021. *Mycobacterium avium* ssp. *paratuberculosis* in the food supply: a public health issue. *Front Public Health* 9: 647448. <https://doi.org/10.3389/fpubh.2021.647448>.
- Whittington R, Donat K, Weber MF, Kelton D, Nielsen SS, Eisenberg S, Arrigoni N, Juste R, Sáez JL, Dhand N, Santi A, Michel A, Barkema H, Kralik P, Kostoulas P, Citer L, Griffin F, Barwell R, Moreira MAS, Slana I, Koehler H, Singh SV, Yoo HS, Chávez-Gris G, Goodridge A, Oceppek M, Garrido J, Stevenson K, Collins M, Alonso B, Cirone K, Paolicchi F, Gavey L, Rahman MT, de Marchin E, Van Praet W, Bauman C, Fecteau G, McKenna S, Salgado M, Fernández-Silva J, Dzedzinska R, Echeverría G, Seppänen J, Thibault V, Fridriksdóttir V, Derakhshandeh A, Haghkhah M, Ruocco L, Kawaji S, et al. 2019. Control of paratuberculosis: who, why and how. A review of 48 countries. *BMC Vet Res* 15:198. <https://doi.org/10.1186/s12917-019-1943-4>.
- Links IJ, Denholm LJ, Evers M, Kingham LJ, Greenstein RJ. 2021. Is vaccination a viable method to control Johne's disease caused by *Mycobacterium avium*

- subsp. paratuberculosis? Data from 12 million ovine vaccinations and 7.6 million carcass examinations in New South Wales, Australia from 1999–2009. *PLoS One* 16:e0246411. <https://doi.org/10.1371/journal.pone.0246411>.
8. Foddai ACG, Grant IR. 2020. Methods for detection of viable foodborne pathogens: current state-of-art and future prospects. *Appl Microbiol Biotechnol* 104:4281–4288. <https://doi.org/10.1007/s00253-020-10542-x>.
 9. Kirkpatrick BW, Cooke ME, Frie M, Sporer KRB, Lett B, Wells SJ, Coussens PM. 2022. Genome-wide association analysis for susceptibility to infection by *Mycobacterium avium* ssp. paratuberculosis in US Holsteins. *J Dairy Sci* 105:4301–4313. <https://doi.org/10.3168/jds.2021-21276>.
 10. Acharya KR, Dhand NK, Whittington RJ, Plain KM. 2017. PCR inhibition of a quantitative PCR for detection of *Mycobacterium avium* subspecies paratuberculosis DNA in feces: diagnostic implications and potential solutions. *Front Microbiol* 8:115. <https://doi.org/10.3389/fmicb.2017.00115>.
 11. Hosseiniporham S, Rebechesu L, Pintore P, Lollai S, Dattena M, Russo S, Ruiu A, Sechi LA. 2022. A rapid phage assay for detection of viable *Mycobacterium avium* subsp. paratuberculosis in milk. *Sci Rep* 12:475. <https://doi.org/10.1038/s41598-021-04451-w>.
 12. Salgado M, Steuer P, Troncoso E, Collins MT. 2013. Evaluation of PMS-PCR technology for detection of *Mycobacterium avium* subsp. paratuberculosis directly from bovine fecal specimens. *Vet Microbiol* 167:725–728. <https://doi.org/10.1016/j.vetmic.2013.09.009>.
 13. Zuo T, Kamm MA, Colombel JF, Ng SC. 2018. Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 15:440–452. <https://doi.org/10.1038/s41575-018-0003-z>.
 14. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JL. 2005. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 102:11070–11075. <https://doi.org/10.1073/pnas.0504978102>.
 15. Magne F, Gotteland M, Gauthier L, Zazueta A, Pesoa S, Navarrete P, Balamurugan R. 2020. The Firmicutes/Bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients? *Nutrients* 12:1474. <https://doi.org/10.3390/nu12051474>.
 16. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. 2017. *Faecalibacterium prausnitzii*: from microbiology to diagnostics and prognostics. *ISME J* 11:841–852. <https://doi.org/10.1038/ismej.2016.176>.
 17. Xu C, Jackson SA. 2019. Machine learning and complex biological data. *Genome Biol* 20:76. <https://doi.org/10.1186/s13059-019-1689-0>.
 18. Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. 2018. Next-generation machine learning for biological networks. *Cell* 173:1581–1592. <https://doi.org/10.1016/j.cell.2018.05.015>.
 19. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. 2019. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf Fusion* 50:71–91. <https://doi.org/10.1016/j.inffus.2018.09.012>.
 20. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. 2016. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput Biol* 12:e1004977. <https://doi.org/10.1371/journal.pcbi.1004977>.
 21. Mossotto E, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. 2017. Classification of paediatric inflammatory bowel disease using machine learning. *Sci Rep* 7:2427. <https://doi.org/10.1038/s41598-017-02606-2>.
 22. Verhaar BJH, Hendriksen HMA, de Leeuw FA, Doorduijn AS, van Leeuwenstijn M, Teunissen CE, Barkhof F, Scheltens P, Kraaij R, van Duijn CM, Nieuwdorp M, Muller M, van der Flier WM. 2021. Gut microbiota composition is related to AD pathology. *Front Immunol* 12:794519. <https://doi.org/10.3389/fimmu.2021.794519>.
 23. Wilkinson TJ, Huws SA, Edwards JE, Kingston-Smith AH, Siu-Ting K, Hughes M, Rubino F, Friedersdorff M, Creevey CJ. 2018. CowPI: a rumen microbiome focussed version of the PICRUSt functional inference software. *Front Microbiol* 9:1095. <https://doi.org/10.3389/fmicb.2018.01095>.
 24. Scanu AM, Bull TJ, Cannas S, Sanderson JD, Sechi LA, Dettori G, Zanetti S, Hermon-Taylor J. 2007. *Mycobacterium avium* subspecies paratuberculosis infection in cases of irritable bowel syndrome and comparison with Crohn's disease and Johne's disease: common neural and immune pathogenecities. *J Clin Microbiol* 45:3883–3890. <https://doi.org/10.1128/JCM.01371-07>.
 25. Sisto M, Cucci L, D'Amore M, Dow TC, Mitolo V, Lisi S. 2010. Proposing a relationship between *Mycobacterium avium* subspecies paratuberculosis infection and Hashimoto's thyroiditis. *Scand J Infect Dis* 42:787–790. <https://doi.org/10.3109/00365141003762306>.
 26. Frau J, Cossu D, Coghe G, Loreface L, Fenu G, Melis M, Paccagnini D, Sardu C, Murru MR, Tranquilli S, Marrosu MG, Sechi LA, Cocco E. 2013. *Mycobacterium avium* subsp. paratuberculosis and multiple sclerosis in Sardinian patients: epidemiology and clinical features. *Mult Scler* 19:1437–1442. <https://doi.org/10.1177/1352458513477926>.
 27. Backhed F, Fraser CM, Ringel Y, Sanders ME, Sartor RB, Sherman PM, Versalovic J, Young V, Finlay BB. 2012. Defining a healthy human gut microbiome: current concepts, future directions, and clinical applications. *Cell Host Microbe* 12:611–622. <https://doi.org/10.1016/j.chom.2012.10.012>.
 28. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HM, Coakley M, Lakshminarayanan B, O'Sullivan O, Fitzgerald GF, Deane J, O'Connor M, Harnedy N, O'Connor K, O'Mahony D, van Sinderen D, Wallace M, Brennan L, Stanton C, Marchesi JR, Fitzgerald AP, Shanahan F, Hill C, Ross RP, O'Toole PW. 2012. Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488:178–184. <https://doi.org/10.1038/nature11319>.
 29. Infante-Villamil S, Huerlimann R, Jerry DR. 2021. Microbiome diversity and dysbiosis in aquaculture. *Rev Aquacult* 13:1077–1096. <https://doi.org/10.1111/raq.12513>.
 30. Zhang L, Liu F, Xue J, Lee SA, Liu L, Riordan SM. 2022. Bacterial species associated with human inflammatory bowel disease and their pathogenic mechanisms. *Front Microbiol* 13:801892. <https://doi.org/10.3389/fmicb.2022.801892>.
 31. Naser SA, Sagrainsingh SR, Naser AS, Thanigachalam S. 2014. *Mycobacterium avium* subspecies paratuberculosis causes Crohn's disease in some inflammatory bowel disease patients. *World J Gastroenterol* 20:7403–7415. <https://doi.org/10.3748/wjg.v20.i23.7403>.
 32. Derakhshani H, De Buck J, Mortier R, Barkema HW, Krause DO, Khafipour E. 2016. The features of fecal and ileal mucosa-associated microbiota in dairy calves during early infection with *Mycobacterium avium* subspecies paratuberculosis. *Front Microbiol* 7:426. <https://doi.org/10.3389/fmicb.2016.00426>.
 33. De Fazio L, Cavazza E, Spisni E, Strillacci A, Centanni M, Candela M, Pratico C, Campieri M, Ricci C, Valerii MC. 2014. Longitudinal analysis of inflammation and microbiota dynamics in a model of mild chronic dextran sulfate sodium-induced colitis in mice. *World J Gastroenterol* 20:2051–2061. <https://doi.org/10.3748/wjg.v20.i8.2051>.
 34. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobetti M, De Angelis M. 2020. The controversial role of human gut Lachnospiraceae. *Microorganisms* 8:573. <https://doi.org/10.3390/microorganisms8040573>.
 35. Png CW, Linden SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin TH. 2010. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* 105:2420–2428. <https://doi.org/10.1038/ajg.2010.281>.
 36. Fecteau ME, Pitta DW, Vecchiarelli B, Indugu N, Kumar S, Gallagher SC, Fyock TL, Sweeney RW. 2016. Dysbiosis of the fecal microbiota in cattle infected with *Mycobacterium avium* subsp. paratuberculosis. *PLoS One* 11:e0160353. <https://doi.org/10.1371/journal.pone.0160353>.
 37. Matthews C, Cotter PD, O'Mahony J. 2021. MAP, Johne's disease and the microbiome: current knowledge and future considerations. *Anim Microbiome* 3:34. <https://doi.org/10.1186/s42523-021-00089-1>.
 38. Elmagzoub WA, Idris SM, Isameldin M, Arabi N, Abdo A, Ibrahim M, Khan MAA, Tanneberger F, Bakht SM, Okuni JB, Ojok L, Gameel AA, Abd El Wahed A, Bekaert M, Mukhtar ME, Amanzada A, Eltom KH, Eltayeb E. 2022. *Mycobacterium avium* subsp. paratuberculosis and microbiome profile of patients in a referral gastrointestinal diseases centre in the Sudan. *PLoS One* 17:e0266533. <https://doi.org/10.1371/journal.pone.0266533>.
 39. Kim ET, Lee SJ, Kim TY, Lee HG, Atikur RM, Gu BH, Kim DH, Park BY, Son JK, Kim MH. 2021. Dynamic changes in fecal microbial communities of neonatal dairy calves by aging and diarrhea. *Animals (Basel)* 11:1113. <https://doi.org/10.3390/ani11041113>.
 40. Kim HS, Whon TW, Sung H, Jeong YS, Jung ES, Shin NR, Hyun DW, Kim PS, Lee JY, Lee CH, Bae JW. 2021. Longitudinal evaluation of fecal microbiota transplantation for ameliorating calf diarrhea and improving growth performance. *Nat Commun* 12:161. <https://doi.org/10.1038/s41467-020-20389-5>.
 41. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, Gonzalez A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. 2014. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15:382–392. <https://doi.org/10.1016/j.chom.2014.02.005>.
 42. Laurin EL, Sanchez J, Chaffer M, McKenna SLB, Keefe GP. 2017. Assessment of the relative sensitivity of milk ELISA for detection of *Mycobacterium avium* ssp. paratuberculosis infectious dairy cows. *J Dairy Sci* 100:598–607. <https://doi.org/10.3168/jds.2016-11194>.
 43. Kohler H, Wichert A, Donat K. 2022. Variation in the performance of different batches of two *Mycobacterium avium* subspecies paratuberculosis

- antibody ELISAs used for pooled milk samples. *Animals (Basel)* 12:442. <https://doi.org/10.3390/ani12040442>.
44. Mathevon Y, Foucras G, Falguieres R, Corbiere F. 2017. Estimation of the sensitivity and specificity of two serum ELISAs and one fecal qPCR for diagnosis of paratuberculosis in sub-clinically infected young-adult French sheep using latent class Bayesian modeling. *BMC Vet Res* 13:230. <https://doi.org/10.1186/s12917-017-1145-x>.
 45. Koets A, Ravesloot L, Ruuls R, Dinkla A, Eisenberg S, Lievaart-Peterson K. 2019. Effects of age and environment on adaptive immune responses to *Mycobacterium avium* subsp. paratuberculosis (MAP) vaccination in dairy goats in relation to paratuberculosis control strategies. *Vet Sci* 6:62. <https://doi.org/10.3390/vetsci6030062>.
 46. Timms VJ, Mitchell HM, Neilan BA. 2015. Optimisation of DNA extraction and validation of PCR assays to detect *Mycobacterium avium* subsp. paratuberculosis. *J Microbiol Methods* 112:99–103. <https://doi.org/10.1016/j.mimet.2015.03.016>.
 47. Biet F, Bay S, Thibault VC, Euphrasie D, Grayon M, Ganneau C, Lanotte P, Daffe M, Gokhale R, Etienne G, Reytrat JM. 2008. Lipopentapeptide induces a strong host humoral response and distinguishes *Mycobacterium avium* subsp. paratuberculosis from *M. avium* subsp. *avium*. *Vaccine* 26: 257–268. <https://doi.org/10.1016/j.vaccine.2007.10.059>.
 48. Over K, Crandall PG, O'Bryan CA, Ricke SC. 2011. Current perspectives on *Mycobacterium avium* subsp. paratuberculosis, Johne's disease, and Crohn's disease: a review. *Crit Rev Microbiol* 37:141–156. <https://doi.org/10.3109/1040841X.2010.532480>.
 49. Sonnenburg ED, Sonnenburg JL. 2019. The ancestral and industrialized gut microbiota and implications for human health. *Nat Rev Microbiol* 17: 383–390. <https://doi.org/10.1038/s41579-019-0191-8>.
 50. Cammarota G, Ianiro G, Ahern A, Carbone C, Temko A, Claesson MJ, Gasbarrini A, Tortora G. 2020. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat Rev Gastroenterol Hepatol* 17:635–648. <https://doi.org/10.1038/s41575-020-0327-3>.
 51. Marcos-Zambrano LJ, Karadzovic-Hadziabdic K, Turukalo TL, Przymus P, Trajkovic V, Aasmets O, Berland M, Gruca A, Hasic J, Hron K, Klammssteiner T, Kolev M, Lahti L, Lopes MB, Moreno V, Naskinova I, Org E, Paciencia I, Papoutsoglou G, Shigdel R, Stres B, Vilne B, Yousef M, Zdravetski E, Tsamardinos I, Pau ECD, Claesson MJ, Moreno-Indias I, Truu J. ML4Microbiome. 2021. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. *Front Microbiol* 12:634511. <https://doi.org/10.3389/fmicb.2021.634511>.
 52. Aryal S, Alimadadi A, Manandhar I, Joe B, Cheng X. 2020. Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. *Hypertension* 76:1555–1562. <https://doi.org/10.1161/HYPERTENSIIONAHA.120.15885>.
 53. Park H-T, Ha S, Park H-E, Shim S, Hur TY, Yoo HS. 2020. Comparative analysis of serological tests and fecal detection in the diagnosis of *Mycobacterium avium* subspecies paratuberculosis infection. *Korean J Vet Res* 60: 117–122. <https://doi.org/10.14405/kjvr.2020.60.3.117>.
 54. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwards CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
 55. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
 56. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71:8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>.
 57. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol* 12:R60. <https://doi.org/10.1186/gb-2011-12-6-r60>.
 58. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. 2020. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 38:685–688. <https://doi.org/10.1038/s41587-020-0548-6>.