RESEARCH ARTICLE

Rapid detection of six *Oceanobacillus* species in *Daqu* starter using single-cell Raman spectroscopy combined with machine learning

Lei Xu[1,2] | Yuan Liang[1] | Wei E Huang[3,4] | Lin-Dong Shang[5] |
Li-Juan Chai[2] | Xiao-Juan Zhang[2] | Jin-Song Shi[6] | Bei Li[5] | Yun Wang[3] |
Zheng-Hong Xu[1,2,7] | Zhen-Ming Lu[1,2,7]

[1]Key Laboratory of Industrial Biotechnology of Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi, China

[2]National Engineering Research Center of Cereal Fermentation and Food Biomanufacturing, Jiangnan University, Wuxi, China

[3]Oxford Suzhou Centre for Advanced Research, Suzhou, China

[4]Department of Engineering Science, University of Oxford, Oxford, UK

[5]State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China

[6]School of Life Sciences and Health Engineering, Jiangnan University, Wuxi, China

[7]National Engineering Research Center of Solid-State Brewing, Luzhou, China

Correspondence
Yun Wang, Oxford Suzhou Centre for Advanced Research, Suzhou 215000, P.R. China.
Email: yun.wang@oxford-oscar.cn

Zheng-Hong Xu and Zhen-Ming Lu, Key Laboratory of Industrial Biotechnology of Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi 214122, China.
Email: zhenghxu@jiangnan.edu.cn and zmlu@jiangnan.edu.cn

Abstract

Many traditional fermented foods and beverages industries around the world request the addition of multi-species starter cultures. However, the microbial community in starter cultures is subject to fluctuations due to their exposure to an open environment during fermentation. A rapid detection approach to identify the microbial composition of starter culture is essential to ensure the quality of the final products. Here, we applied single-cell Raman spectroscopy (SCRS) combined with machine learning to monitor *Oceanobacillus* species in *Daqu* starter, which plays crucial roles in the process of Chinese *baijiu*. First, a total of six *Oceanobacillus* species (*O. caeni*, *O. kimchii*, *O. iheyensis*, *O. sojae*, *O. oncorhynchi* subsp. *Oncorhynchi* and *O. profundus*) were detected in 44 *Daqu* samples by amplicon sequencing and isolated by pure culture. Then, we created a reference database of these *Oceanobacillus* strains which correlated their taxonomic data and single-cell Raman spectra (SCRS). Based on the SCRS dataset, five machine-learning algorithms were used to classify *Oceanobacillus* strains, among which support vector machine (SVM) showed the highest rate of accuracy. For validation of SVM-based model, we employed a synthetic microbial community composed of varying proportions of *Oceanobacillus* species and demonstrated a remarkable accuracy, with a mean error was less than 1% between the predicted result and the expected value. The relative abundance of six different *Oceanobacillus* species during *Daqu* fermentation was predicted within 60 min using this method, and the reliability of the method was proved by correlating the Raman spectrum with the amplicon sequencing profiles by partial least squares regression. Our study provides a rapid, non-destructive and label-free approach for rapid identification of *Oceanobacillus* species in *Daqu* starter culture, contributing to real-time monitoring of fermentation process and ensuring high-quality products.

INTRODUCTION

The technique using *Daqu* to enrich functional microorganisms for comprehensive fermentation of starchy grains has been used over thousands of years. Microorganisms inhabiting *Daqu* not only participate in fermentation of *baijiu* (Chinese distilled spirits) but also provide abundant enzymes for substrate degradation and flavour compound production (Wang et al., 2019; Zou et al., 2018). Meanwhile, microbial metabolites formed during *Daqu* fermentation may be further converted into intermediates for further *baijiu* fermentation of end flavour components, or dissolve in the final product, which is referred to as '*Daqu* aroma' in practice (He et al., 2019; Yang et al., 2021). Thus, the diversity and abundance of functional microorganisms in *Daqu* are vital for *baijiu* fermentation.

Since the fermentation process is carried out in an open environment, the structure of *Daqu* microbiota can be easily affected by raw materials (Zhang et al., 2022), environmental factors (e.g., tools, indoor ground, room temperature) (Li et al., 2016; Zhang et al., 2022) and fermentation technology (e.g., ventilation, packing method) (Kang et al., 2022), resulting in different *Daqu* types. Recent studies have revealed that fermenting *Daqu* with different raw materials leads to varying microbial characteristics (Mao et al., 2022). *Daqu*-making environments, such as indoor ground and tools, are the main sources of fungal communities in *Daqu*, while raw materials primarily contribute to bacterial communities (Du et al., 2019). By controlling the maximum temperature at the centre of the *Daqu* brick (threshold limit) during the fermentation process, three types of *Daqu* can be produced: low-temperature (40–50°C), medium-temperature (50–60°C) and high-temperature *Daqu* (60–65°C). They are used in the production of *baijiu* with different flavours (Wang et al., 2011). However, quality of *Daqu* even in the same batch on same maximum temperature of the *Daqu* brick will be different due to spatial heterogeneity (Shi et al., 2022). Therefore, rapid detection of microorganisms in *Daqu* fermentation is crucial for making operational decisions and ensuring quality control.

Detection of specific species within community contexts is a research hotspot in the field of microbial ecology. Culture-dependent methods have isolated many culturable species from *Daqu*, such as *Bacillus*, *Lactobacillus* and *Saccharomyces* (Li et al., 2015; Wu et al., 2013). However, these methods are labour and time-consuming and have limited ability to detect rare species or non-culturable species (Zheng et al., 2012). On the other hand, PCR-based culture-independent methods such as amplicon sequencing and PCR-DGGE have been widely used to assess the microbial community structure (Chai et al., 2019, 2020; Ling et al., 2020). Most of these methods require high-quality DNA extraction, specific primers and long sequencing and data assembly times (>1 day). Consequently, these factors make these approaches less suitable for rapid detection of microbes during the fermentation process.

Raman spectroscopy is a photonic technique based on vibrational Raman scattering, which indicates molecular vibration information through calculating the difference in wavelength between excitation and emission upon interaction with a sample (Raman & Krishnan, 1928). Single-cell Raman spectra (SCRS) can reveal physiological and biochemical information at the single-cell level by collecting molecular vibration profiles from cells (Cui et al., 2022; Wakisaka et al., 2016). With the advantages of being rapid (a few seconds per measurement), label-free and non-destructive, SCRS is highly attractive for studying different biological objects in fields such as cancer diagnosis (Xu et al., 2021), plant science (Barańska et al., 2012), toxicology (Arend et al., 2020) and food safety (Wang et al., 2020). When combined with machine learning, SCRS can identify various single cells, such as infected and non-infected cells from peripheral blood (Arend et al., 2020), and urinary tract infection strains (Kloss et al., 2013). As a result, these characteristics make SCRS coupled with machine learning an ideal rapid method for monitoring microbial succession in brewing and detecting potential spoilage microorganisms.

Oceanobacillus is an alkaliphilic and extremely halotolerant genus that is widely found in traditional fermented foods such as kimchi (Nam et al., 2008), soy sauce (Tominaga et al., 2009) and shrimp paste (Namwong et al., 2009). In these food fermentation processes, *Oceanobacillus* can secrete various enzymes and organic acids, which play important roles in the formation of unique food flavour (Nam et al., 2008; Namwong et al., 2009; Thanapun, 2013; Tominaga et al., 2009; Zhang et al., 2021). In high-temperature *Daqu* microbiota, *Oceanobacillus* has recently been identified as a dominant genus (relative abundance >1%) correlated with liquefaction and saccharification enzyme activity (Chen et al., 2020; Shi et al., 2022). Metaproteomics analysis showed that *O. iheyensis* could regulate the metabolism of five-member heterocyclic amino acids in *Daqu* microbiota (Yang et al., 2023). Furthermore, *Oceanobacillus* functions as a significant contributor to characteristic compounds during *Daqu* fermentation (Zhang et al., 2021). Therefore, detecting the composition and content of *Oceanobacillus* in *Daqu* fermentation is helpful for controlling the fermentation process and product quality. In this study, we applied SCRS combined with machine learning model to identify six *Oceanobacillus* species in the microbial community of high-temperature *Daqu*. This study provides a label-free, non-destructive and rapid approach for monitoring functional microorganisms in *Daqu* fermentation.

IDENTIFICATION OF OCEANOBACILLUS IN DAQU STARTER

MICROBIAL
BIOTECHNOLOGY

Applied
Microbiology
International

3 of 12

EXPERIMENTAL PROCEDURES

Daqu sampling and amplicon sequencing

Forty-four high-temperature *Daqu* samples of different qualities were collected from several fermentation rooms in Guizhou province, China. The samples were ground into powder in liquid nitrogen with pestles. The total DNA of each *Daqu* sample was extracted using the soil DNA isolation kit (QIAGEN, Germany) according to the manufacturer's instructions. Primers 338F (5′-ACTCCTACGGGAGGCAGCAG-3′)/806R (5′-GAC TACHVGGGTWTCTAAT-3′) were used to amplify the V3-V4 hypervariable region of bacterial 16S rRNA genes. The amplicon sequencing was completed by using the Illumina MiSeq platform (Shanghai Majorbio Bio-pharm Technology Co., Ltd.). Sequencing reads were grouped into amplicon sequence variants (ASVs) as a function of their pairwise sequence similarities. The ASVs of *Oceanobacillus* were collected, and the relative abundances of these ASVs in *Daqu* samples were calculated. A phylogenetic tree displaying the genotypic heterogeneity of ASVs belonging to the genus *Oceanobacillus* was constructed using the neighbour-joining method. MEGA 7 software was employed to construct the ASVs of *Oceanobacillus* phylogenetic tree with the neighbour-joining method.

Isolation of *Oceanobacillus* strains

The Nutrient Agar Medium was used to isolate *Oceanobacillus* strains from *Daqu*, including 1.0% peptone, 0.3% beef extract, 0.5% NaCl and 1.5% agar. The taxonomic classification of isolates was performed by sequencing the 16S rRNA genes. In total, six *Oceanobacillus* species were isolated from *Daqu*, namely *O. caeni* ZY111012, *O. kimchii* ZY10094H, *O. iheyensis* ZS10902A, *O. sojae* ZS102924, *O. oncorhynchi* subsp. *oncorhynchi* ZS102907 and *O. profundus* ZQ100018. The 16S rRNA gene sequences of the six *Oceanobacillus* strains were submitted to GenBank (ON935606, ON935612, ON935611, ON935610, ON935607 and ON935609). A BLAST analysis was conducted to compare the similarity of the 16S rRNA gene sequence of the isolated strain and the ASVs of *Oceanobacillus* from amplicon sequencing by using DNAman software.

Raman measurement

Six *Oceanobacillus* strains were cultivated at 37°C using Nutrient Medium to reach their log phases. Each strain was cultured independently in at least three different batches as biological replicates. Cells in the fermentation broth (1 mL) were obtained by centrifugation at 7000 rpm for 2 min and resuspended in sterile water

three times. Then, 3 μL of the suspension was dried for 15 min on a sterile aluminium-coated substrate. Raman spectra of the six *Oceanobacillus* strains were measured across monolayer regions of the dried samples using the auto-measure mode of confocal Raman microscopy (HOOKE Instruments Ltd., China). The spectrometer response was calibrated against a source with some silicon, standard reference materials 2242a, Ne, Ar and Hg samples. Meanwhile, 532 nm illumination at 3 mW was used with a 600 g/mm grating to generate spectra with 4 cm^{-1} dispersion to maximize signal strength while minimizing background signal from autofluorescence. To acquire adequate SCRS data to cover different kinds of morphological or physiological features, we randomly picked 100 single cells from each *Oceanobacillus* species. Next, Raman spectra of the six *Oceanobacillus* species were treated with the same spectrum processing steps, such as removing cosmic ray, correcting baseline with adaptive iteratively reweighted penalized least-squares (airPLS) (Zhang et al., 2010), smoothing spectra with Savitzky–Golay and normalization. Finally, the Raman spectra range between 500 and 1800 cm^{-1} was used for further analysis.

Chemometric analysis of *Oceanobacillus* Raman spectrum

To analyse the spectral differences in different bacterial species, dimension reduction techniques such as linear discriminant analysis (LDA) and hierarchical cluster analysis (HCA) were used to analyse SCRS. LDA was performed using the MASS package (Ripley, 2015). The significances of LD1 and LD2 values in different groups were compared by Duncan's multiple-range test. Peaks marked with different letters possess significantly different values ($p < 0.05$). HCA was performed by using the Euclidean distance of preprocessed average spectral data. The assignment of the major Raman bands in biological samples has been investigated by a number of studies, which serves as a reference library for spectral interpretation in microbiological analysis (De Gelder et al., 2007; Talari et al., 2014; Wang et al., 2015). Raman peaks observed in the spectra of *Oceanobacillus* species and their tentative assignments were summarized. Raman peaks from six *Oceanobacillus* strains were analysed by Kruskal–Wallis tests to reveal the statistical significance of the interpretation of each *Oceanobacillus* strain.

Model training and validation

In order to discriminate different species of *Oceanobacillus*, five machine learning models including k-nearest neighbour (kNN) (Zhang et al., 2018),

4 of 12

MICROBIAL
BIOTECHNOLOGY

Applied
Microbiology
International

XU ET AL.

logistic regression (LR) (Stoltzfus, 2011), random forest (RF) (Ho, 1998), support vector machine (SVM) (Tolstik et al., 2014) and Gaussian Naive Bayesian (GNB) (Bhargava et al., 2006) were established. The collected Raman spectra of six *Oceanobacillus* strains were used for model establishment, of which 70% were used for training and 30% for testing. The scikit-learn package was used to perform a grid search function to adjust the hyperparameters based on the area under curve (AUC) value. The five classification models were evaluated by calculating the precision, recall and F1 score, with the optimal model possessing the highest score. The confusion matrix of the model was used to visualize the classification and prediction results. The row corresponds to the bacterial species identified by the standard biological real class, while the column corresponds to the bacterial identification predicted by the algorithm.

The applicability of the classification model was verified with synthetic microbial communities (SynComs) containing different proportions of *Oceanobacillus* species, *Oceanobacillus*-free vinegar microbiota (microbiota within the fermented grains in vinegar brewing) or *Oceanobacillus*-containing *Daqu* microbiota. The designed combination of *Oceanobacillus* species is shown in Figure 4A. The sample pretreatment and Raman detection were carried out according to previous conditions. The number of collected SCRS was 100–200 for each SynCom. The Raman spectra were input into the classification model to generate predicted results. The prediction score threshold of the classification model was 0.9 (that is, if the prediction score of SCRS is less than 0.9, the single cell is classified as another species; if the prediction score is higher than 0.9, the cell is classified as the highest score strain). The relative abundance of *Oceanobacillus* in each SynCom was calculated, and the predicted ratio to the real ratio was calculated by mean absolute error. The recovery rate was calculated by the relative abundance of *O. sojae* in SynComs containing *Daqu* microbiota.

Predicting the abundance and numbers of *Oceanobacillus* in *Daqu*

Three parallel *Daqu* bricks were taken out from the same incubating room on days 2, 4, 6, 10, 20 and 40, respectively. *Daqu* bricks were separately mashed into powders, and *Daqu* (1 g) were homogenized in 5 mL of phosphate buffer saline. After being vigorously vortexed for 10 min, the cell suspension was centrifuged at 500 rpm for 5 min. Then, the supernatant was centrifuged at 7000 rpm for 2 min. The precipitate was washed repeatedly with sterile water and centrifuged three times. The cell pellets were resuspended in 0.2 mL sterile water. Finally, an appropriate amount of bacterial

suspension was placed on the Raman chip and allowed to air dry. The SCRS of *Daqu* was collected by a confocal Raman spectrometer, with an excitation wavelength 532 nm, scanning spectrum range of 500~3750 cm^{-1}, laser power range of 3 mW and scanning time of 5 s. The determination condition setting of the Raman spectrum and the method of Raman data processing were the same as that of single cell. More than 500 SCRS of microorganisms in *Daqu* were collected. The taxonomic information and relative abundance of six *Oceanobacillus* species in *Daqu* microbiota were determined by using the *Oceanobacillus* Raman classification model. To evaluate the linearity between Raman detection and amplicon sequencing, we analysed the relative abundance of *Oceanobacillus* using a PLS-R model. The model performance was evaluated using the coefficient of determination (R^2) and the root mean square error (RMSEcv).
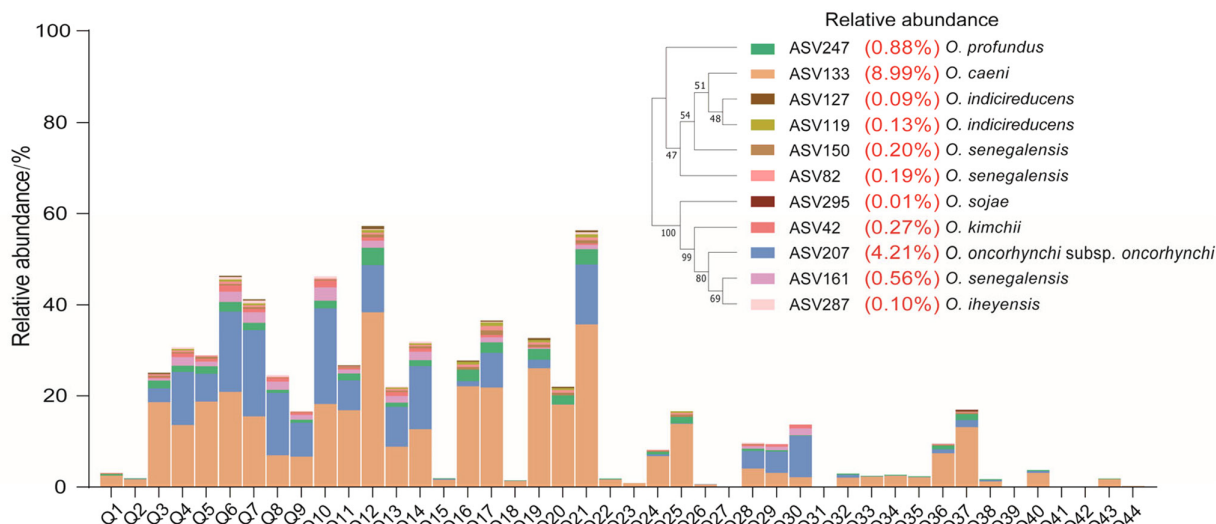
RESULTS

Identification of *Oceanobacillus* in starter culture in *Daqu*

In order to investigate whether *Oceanobacillus* can be used as an indicator strain for monitoring *Daqu* of varying quality, we explored the diversity of *Oceanobacillus* community in *Daqu* with different levels of quality. As revealed by 16S rRNA gene sequencing from difference quality *Daqu* samples, *Oceanobacillus* was a common bacteria (relative abundances from 0% to 57.3%, average >10%) in the bacterial community of high-temperature *Daqu* (Figure 1A). At the species level, a total of 410 ASVs were annotated by SILVA (v132) databases from 44 *Daqu* samples, in which 11 ASVs distributed in eight *Oceanobacillus* species, including *O. caeni* (ASV133, 8.99%), *O. oncorhynchi* subsp. *oncorhynchi* (ASV207, 4.21%), *O. senegalensis* (ASV161, 0.56%; ASV150, 0.2%; ASV82, 0.19%), *O. profundus* (ASV247, 0.88%), *O. kimchii* (ASV42, 0.27%), *O. indicireducens* (ASV119, 0.13%; ASV127, 0.09%), *O. iheyensis* (ASV287, 0.1%) and *O. sojae* (ASV295, 0.01%). The relative abundances of these 11 ASVs were significantly different among the 44 *Daqu* samples ($p < 0.05$), even undetectable in some *Daqu* samples. The *Oceanobacillus* community content and species-level structure exhibit variations in *Daqu* with different flavour and functional nutrients, resulting in the differences in *Daqu* quality (Figure 1A).

Isolation of *Oceanobacillus* strains

A total of 280 strains were isolated from *Daqu* sample by pure culture with Nutrient Agar Medium. Among these isolates, 50 strains were classified to the genus

IDENTIFICATION OF OCEANOBACILLUS IN DAQU STARTER

MICROBIAL
BIOTECHNOLOGY

Applied
Microbiology
International

5 of 12

(A)

(B)

The closest strain of ASV sequence alignment with EZBiocloud	Identity	*Oceanobacillus* ASV	Similarity	16S rRNA sequence of isolated strain	Identity	The closest strain of 16S rRNA sequence of isolated strain alignment with EZBiocloud
Oceanobacillus caeni S-11	99.76%	ASV133	100%	ON935606	98.63%	*Oceanobacillus caeni* S-11
Oceanobacillus oncorhynchi subap. *onc.* R-2	98.72%	ASV207	98.86%	ON935607	99.71%	*Oceanobacillus oncorhynchi* subsp. *onc.* R-2
Oceanobacillus profundus S5	97.86%	ASV247	99.33%	ON935609	100%	*Oceanobacillus profundus* S5
Oceanobacillus kimchii X50	98.52%	ASV42	98.35%	ON935612	99.18%	*Oceanobacillus kimchii* X50
Oceanobacillus iheyensis HTE831	98.51%	ASV287	98.35%	ON935611	99.18%	*Oceanobacillus iheyensis* HTE831
Oceanobacillus sojae Y27	100%	ASV295	100%	ON935610	99.18%	*Oceanobacillus sojae* Y27

FIGURE 1 Targeted isolation of *Oceanobacillus* species from the high-temperature *Daqu* microbiota. (A) Species-level diversity of *Oceanobacillus* community in different high-temperature *Daqu* samples (n = 44) from different *baijiu* factories, among which Q1–Q15 belonged to black *Daqu*, Q16–Q30 belonged to yellow *Daqu* and Q31–Q44 belonged to white *Daqu*. The phylogenetic tree displayed the genotypic heterogeneity of amplicon sequence variants (ASVs) belonging to the genus *Oceanobacillus* by the neighbour-joining method. Percentage in the bracket means the relative abundance of ASV in the bacterial community of *Daqu*. (B) Identification of the isolated strains. The 16S rRNA gene sequence of the isolated strains was aligned with ASVs of *Oceanobacillus* determined from the *Daqu* data set and annotated in the EZBioCloud database to match the closest model strain. The similarity was assessed between the ASV sequence and 16S rRNA gene, both in comparison to the GenBank database, respectively.

Oceanobacillus (identity >97%), by comparing with the 16S rRNA gene sequence of model strains in the EZBioCloud database. These isolates were further categorized into six species, *O. caeni*, *O. kimchii*, *O. iheyensis*, *O. sojae*, *O. oncorhynchi* subsp. *oncorhynchi* and *O. profundus* (Figure 1B).

In order to identify the representative strain with the highest similarity to the dominant ASV sequence among the *Oceanobacillus* species, the 16S rRNA gene sequence of the isolated strains was compared with the ASV sequence determined by amplicon sequencing. A total of six strains, including *O. caeni* ZY111012, *O. kimchii* ZY10094H, *O. iheyensis* ZS10902A, *O. sojae* ZS102924, *O. oncorhynchi* subsp. *oncorhynchi* ZS102907 and *O. profundus* ZQ100018, were respectively matched with ASV133, ASV42, ASV287, ASV295, ASV207 and ASV247, with the similarity from 98% to 100% (Figure 1B). The results indicated that the main *Oceanobacillus* species were isolated from the microbial community of *Daqu*.

Characterizing and distinguishing SCRS of *Oceanobacillus* species

Under the bright field microscope, the cells of different *Oceanobacillus* species were oval, accompanied by some spores (Figure 2A) and could not be distinguished by the naked eye. More than 100 SCRS for each *Oceanobacillus* species were collected and processed by de-background noise, smoothing and normalization. The collected SCRS were linked to their taxonomic data, thereby constituting a reference database of *Daqu Oceanobacillus*. In general, all the six *Oceanobacillus* species had strong Raman spectral signals and their average Raman spectral curves and intensities were basically similar, with SCRS in the range of 500–1800 cm^{-1} (Figure 2B).

Characteristic Raman peaks observed in the SCRS of six *Oceanobacillus* species and their tentative assignments are summarized in Table S1. Most of the peaks could be ascribed to the skeletal structure of

6 of 12

MICROBIAL
BIOTECHNOLOGY

Applied
Microbiology
International

XU ET AL.
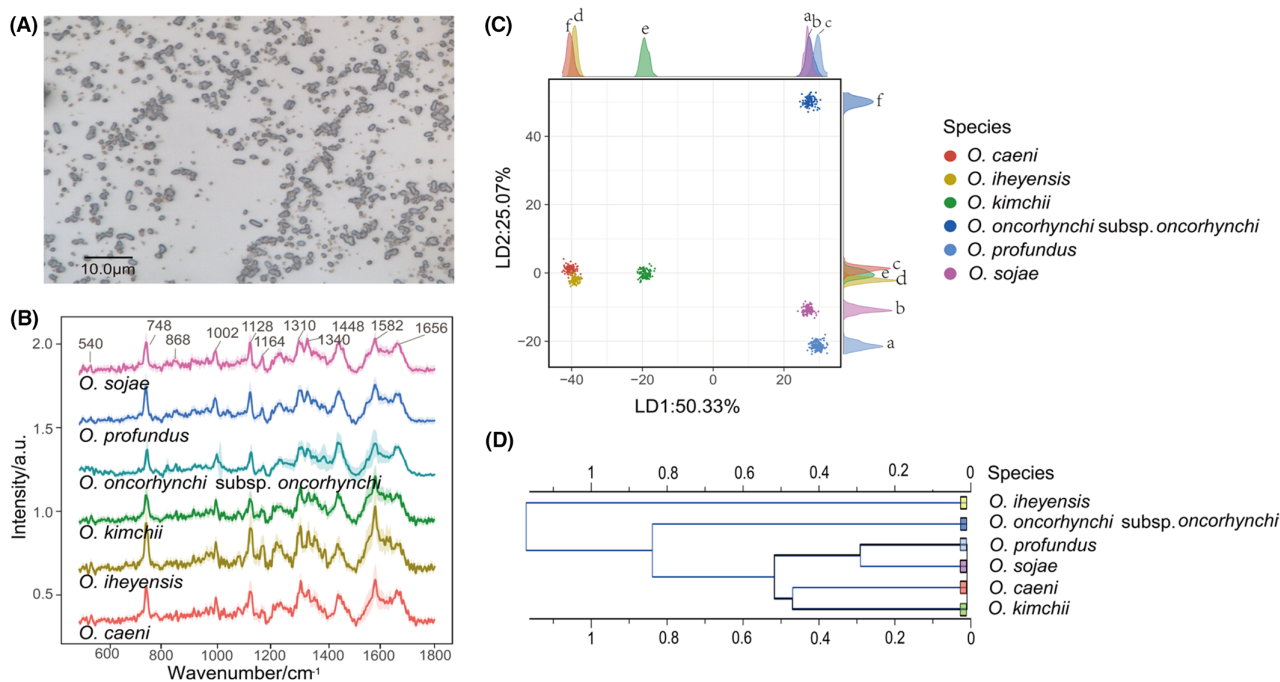
FIGURE 2 Multivariate analysis of the single-cell Raman spectra of six *Oceanobacillus* species. (A) The optical micrograph of the typical cell morphology of *O. profundus*. Under a 100-fold microscope, normal-sized cells and small spore cells were observed. Scale bar, 10 μm. (B) The averaged Raman spectra of 200 single cells belonging to the six *Oceanobacillus* species. The shadow areas represent the standard deviation of Raman spectra. (C) LDA visualization of the Raman spectra of different *Oceanobacillus*, in which six groups were identified (each point of the plot corresponds to a spectrum taken from one individual cell). LD1 and LD2 represent the results of LDA displayed on the *x*-axis and *y*-axis, respectively. Statistically significant differences ($p < 0.05$) are denoted by different letters (e.g., a and b). (D) The HCA results based on the averaged Raman spectra of different *Oceanobacillus*.

nucleic acids, proteins, lipids, cytochrome c and carbohydrates. The intensity of these Raman peaks showed differences among *Oceanobacillus* species (Figure S1). Specifically, the intensities of the characteristic Raman peaks at 540, 748, 964, 1128, 1164, 1310, 1340, 1582 cm^{-1} in single cells of *O. iheyensi* were significantly different from other *Oceanobacillus* species ($p < 0.05$) (Figure S1), indicating obvious metabolism differences.

The SCRS of six *Oceanobacillus* species were distinguished by LDA and HCA (Figure 2C,D). Different *Oceanobacillus* species were clustered separately by LDA. LD1 and LD2 were able to explain 70.35% of the variances in the data (50.33% by LD1 and 25.02% by LD2) (Figure 2C). Six *Oceanobacillus* species showed a significant distinction in the LDA scores plot for LD1 and LD2 by Duncan's multiple-range test ($p < 0.05$). The results of HCA were consistent with that of LDA (Figure 2D).

Establishing machine learning model for *Oceanobacillus* classification

In order to predict *Oceanobacillus* species in unknown samples, a reliable *Oceanobacillus* Raman spectral database is needed to reflect the biological

TABLE 1 Performance scores of k-nearest neighbour (kNN), logistic regression (LR), random forest (RF), support vector machine (SVM) and Gaussian Naive Bayesian (GNB) to predict six *Oceanobacillus* species.

Model	Precision[a]	Recall[b]	F1-score[c]
GNB	0.79	0.79	0.79
RF	0.91	0.92	0.91
kNN	0.92	0.92	0.92
LR	0.93	0.95	0.94
SVM	0.97	0.97	0.96

[a]Precision, the percentage of positive instances out of the total predicted positive instances.

[b]Recall encapsulates the percentage of positive instances out of the total real positive instances.

[c]F1 score is the mean of precision and recall.

heterogeneity of the species, followed by training and learning of the bacterial species identity using tools such as machine learning. Machine learning models can reduce the workload of operators and improve the efficiency and reliability of microbial detection. In this study, five machine learning models (kNN, LR, RF, SVM and GNB) have been used to distinguish six *Oceanobacillus* species. The precision, recall and F1 scores of SVM were 97%, 97% and 96%, which were the highest among the five models (Table 1). In

contrast, the precision of LR, kNN, RF and GNB models were 75.7%, 91%, 92% and 93%, respectively.

SVM is a widely used classification model that attempts to generate the optimal hyperplane or decision boundary to best separate different categories in a high-dimensional space. The confusion matrix of SVM reflected the misclassification probability among different *Oceanobacillus* species (Figure 3). The spectra of most species were classified into the correct category. A few mismatches of *O. sojae-O. profundus* and *O. iheyensis-O. kimchii* were observed, but their accuracies were acceptable (accuracy >0.8) (Figure 3).

Verification with synthetic microbial community

The applicability of SVM-based classification model was verified with SynComs containing designed combination of *Oceanobacillus* species (Figure 4A). The predicted relative abundances of six *Oceanobacillus* species were compared with the expected results in SynComs 1–8 (Figure 4B). In SynComs 1–4, containing different proportions of *Oceanobacillus* species, the predicted relative abundances were almost identical to the expected results, with the mean absolute error less than 0.2%. In SynCom 5, containing *Oceanobacillus*-free vinegar microbiota, the mean absolute error between the predicted and expected results was 0.75%. In SynComs 6–8, containing *Daqu* microbiota, the predicted relative abundances of *Oceanobacillus* were almost the same with expected results, with the standard recovery rates ranging from 95% to 112%. The above results showed that SCRS
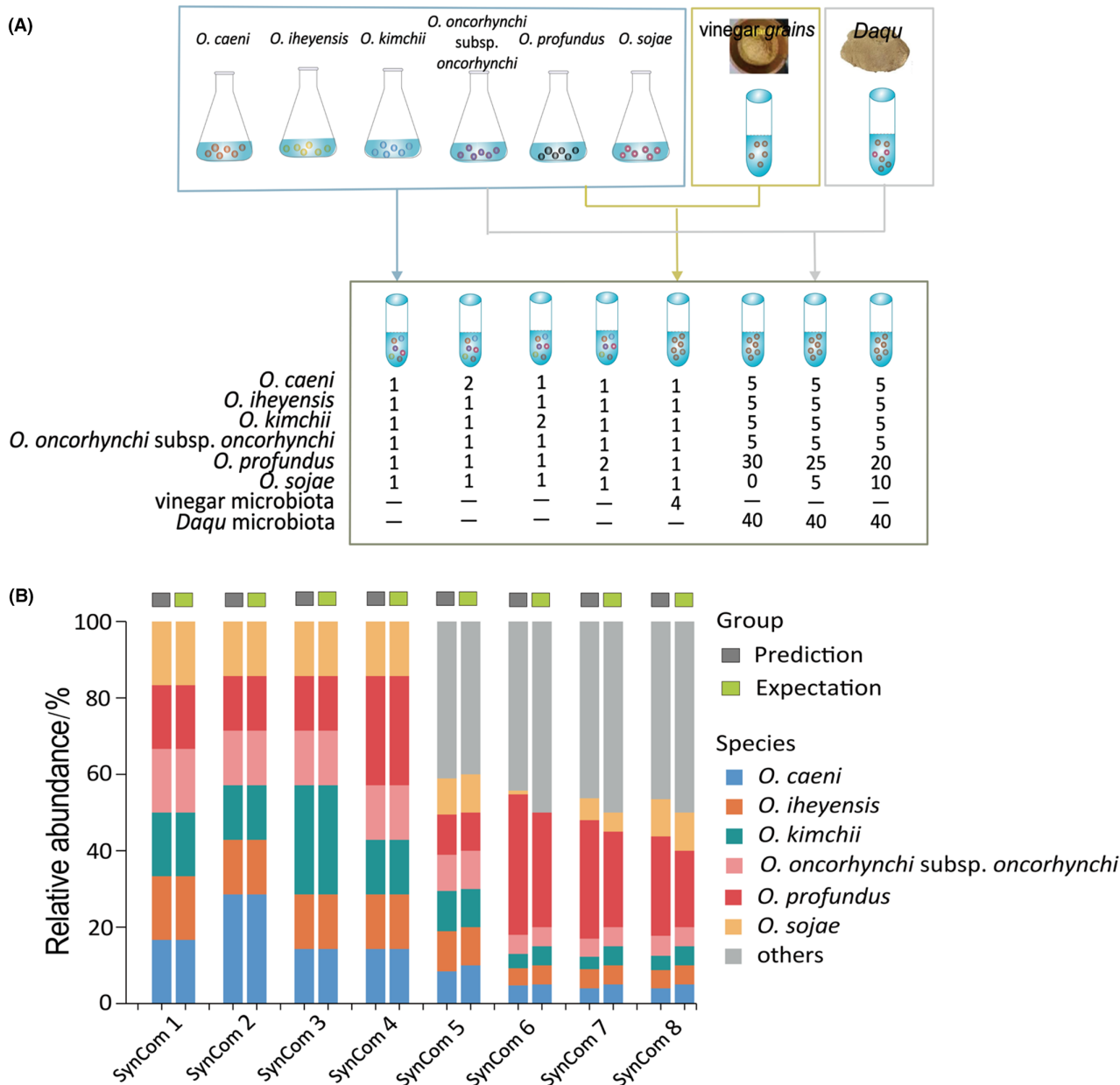
combined with the SVM-based model can accurately identify different *Oceanobacillus* species in the microbial community of *Daqu*.

Monitoring of *Oceanobacillus* species during *Daqu* fermentation

After validating the performance of SCRS combined with the SVM model in the detection of *Oceanobacillus* in vitro, we further investigated their feasibility for real-time *Oceanobacillus* detection during the fermentation of high-temperature *Daqu*. The results of Raman detection were similar to those of amplicon sequencing. T-test results showed no significant difference in *Oceanobacillus* diversity between Raman detection and amplicon sequencing ($p > 0.1$). *Oceanobacillus* was a dominant bacterial genus throughout fermentation (relative abundance varied from 26% to 50%), but the structure of *Oceanobacillus* community varied significantly (Figure 5A). Based on Raman analysis, *O. profundus* dominated the *Oceanobacillus* community in *Daqu* bricks on days 2–4, while *O. caeni* became the dominant *Oceanobacillus* species after 4 days of fermentation (Figure 5A). These results were consistent with those obtained from the amplicon sequencing. To test whether Raman spectra can be correlated with amplicon sequencing analysis at the genotype level, the PLS-R model demonstrated that a simple linear transformation linked these two types of high-dimensional data (Figure 5B). The R^2 of PLS-R model was 0.83 and the RMSEcv of PLS-R model was 7.48 (Figure 5B). It suggests that Raman analysis could be used as a rapid method to classify different *Oceanobacillus* and indicate their relative abundance.
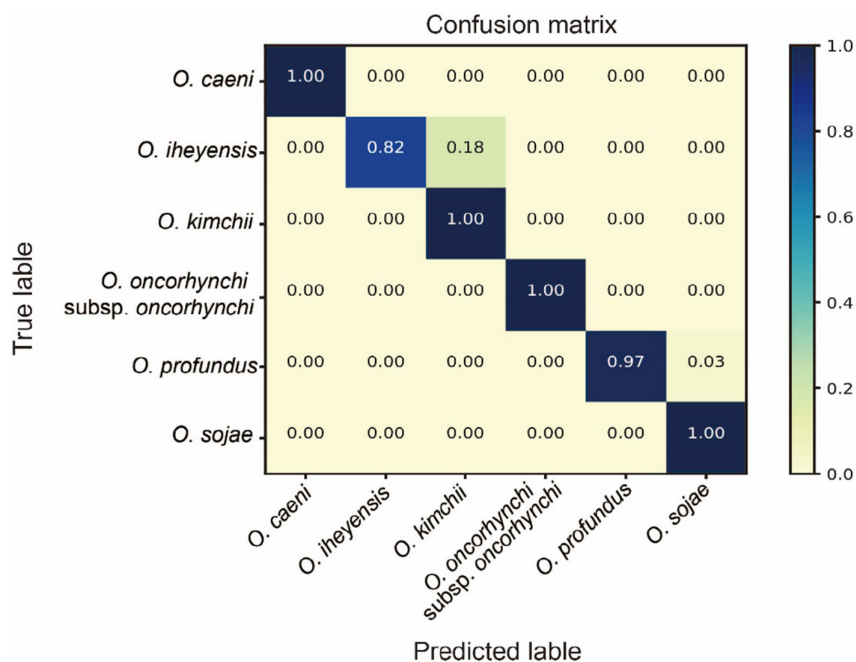
FIGURE 3 Binary confusion matrix for classification of the six *Oceanobacillus* species based on support vector machine.

MICROBIAL BIOTECHNOLOGY Applied Microbiology International

(A)

Species								
O. caeni	1	2	1	1	1	5	5	5
O. iheyensis	1	1	1	1	1	5	5	5
O. kimchii	1	1	2	1	1	5	5	5
O. oncorhynchi subsp. oncorhynchi	1	1	1	1	1	5	5	5
O. profundus	1	1	1	1	2	30	25	20
O. sojae	1	1	1	2	1	0	5	10
vinegar microbiota	—	—	—	—	4	—	—	—
Daqu microbiota	—	—	—	—	—	40	40	40

(B)

FIGURE 4 Evaluation of the feasibility of Raman spectroscopy coupled with the SVM-based classification model for classification of six *Oceanobacillus* species within synthetic microbial communities (SynComs). (A) Schematic illustration of SynComs with or without *Oceanobacillus* species (*O. caeni*, *O. kimchii*, *O. iheyensis*, *O. sojae*, *O. oncorhynchi*, *O. profundus*). The *Daqu* microbiota contains *Oceanobacillus* species, and the vinegar microbiota (microbiota within the fermented grains in vinegar brewing) contains no *Oceanobacillus* species. (B) The predicted results by the SVM-based classification model describe the relative abundance of six *Oceanobacillus* species in SynComs.

DISCUSSION

Complicated microorganisms are beneficial to flavour formation in natural food fermentation, but they also pose challenges to maintain the batch-to-batch uniformity of the fermentation process and the quality of the final products. *Daqu* harbours intricate microbial communities (Figure S2), posing a challenge for quality assurance in production. Comprehensively understanding the dynamic changes in dominant microorganisms presents a challenge due to their abundance and difficult to separate. Therefore, the identification of distinct functional microorganisms in various qualities of *Daqu*, particularly the diverse contents and species of *Oceanobacillus* (Figure 1), proves valuable in utilizing *Oceanobacillus* as marker organisms for monitoring *Daqu* fermentation quality. Moreover, the importance of *Oceanobacillus* in food fermentation has been highlighted, as it demonstrates the capacity to produce acid and α-glucosidase, and the activities of liquefaction and saccharification enzyme in *Daqu* were positively correlated with *Oceanobacillus*, but few studies have

IDENTIFICATION OF OCEANOBACILLUS IN DAQU STARTER

MICROBIAL
BIOTECHNOLOGY

Applied
Microbiology
International

9 of 12

FIGURE 5 Comparison of the *Oceanobacillus* community based on SCRS and 16S rRNA gene sequencing. (A) Detection of *Oceanobacillus* at the species level in the fermentation process of high-temperature *Daqu* by Raman spectroscopy coupled with SVM and amplicon sequencing. (B) Use the PLS-R model to fit a linear regression model between SCRS and 16S rRNA gene sequencing of *Oceanobacillus* in *Daqu* fermentation. The horizontal axis represents the predicted value of Raman spectroscopy, and the vertical axis represents the measured value amplicon sequencing by explained.
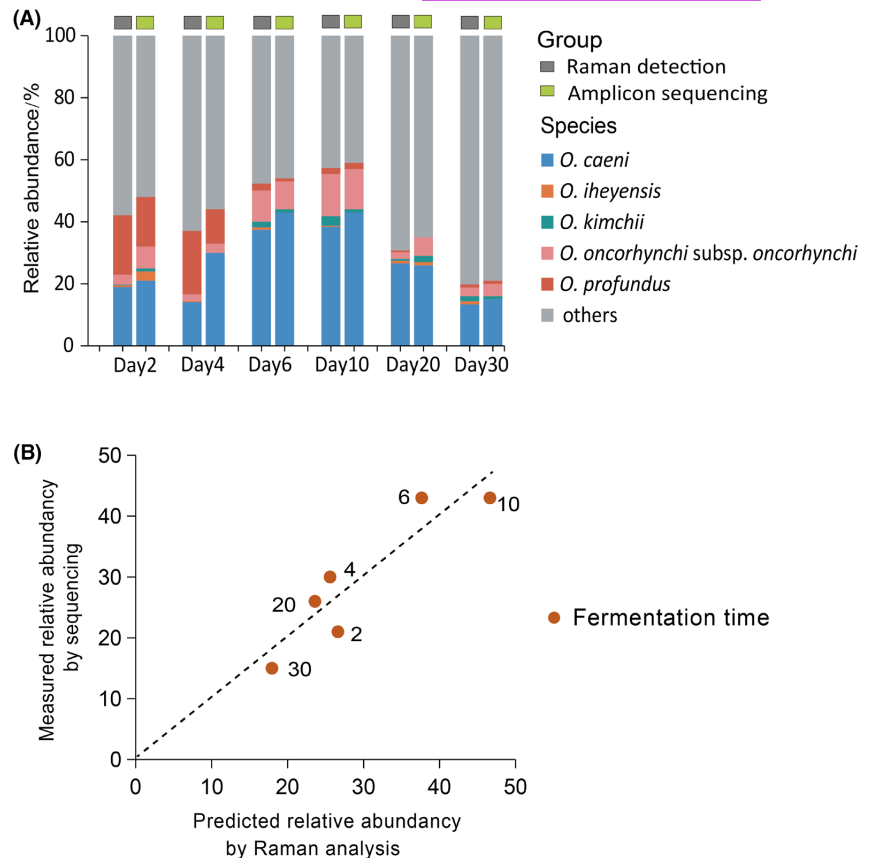
identified *Oceanobacillus* at the species level. To date, more than 35 species in the genus *Oceanobacillus* have been recorded in the EzBioCloud database. PCR-based assays have limited discrimination power in identifying a large number of *Oceanobacillus* species simultaneously, due to the specificity of primers or the preference of amplification sequence (Kumar et al., 2017). Meanwhile, the majority of genotypic methods require high-quality DNA extraction, and long sequencing and data assembly time (>1 day). Thus, we evaluated Raman spectroscopy as a bacterial typing tool for identifying *Oceanobacillus* to the species level that can be eventually used as an approach for *Daqu* fermentation.

To ensure the coverage of the established reference database of *Daqu Oceanobacillus*, 44 high-temperature *Daqu* samples of different qualities were collected from different factories for amplicon sequencing. A total of 11 ASVs distributed in 9 *Oceanobacillus* species were revealed in the microbial community of *Daqu*, demonstrating obvious species diversity of *Oceanobacillus* (Figure 1A). To isolate the representative strains belonging to all the *Oceanobacillus* species in *Daqu* samples, the ASV sequences from amplicon sequencing were aligned with the 16S rRNA gene sequence of isolated strains. The results showed that the dominant *Oceanobacillus* species have been isolated from *Daqu* microbiota (Figure 1B). It is worth noting that the annotated results of ASV sequences with the RDA

and EzBioCloud databases might be different due to inconsistent information (Figure 1). With the update of the database content, the information of target species would be more accurate.

By comparison, it is found that SVM is the optimal machine learning algorithm for classification of SCRS of six *Oceanobacillus* species. Significantly, among the characteristic Raman peaks of six *Oceanobacillus* strains, some 540 and 1128 cm^{-1} represent glucose chemical bonds, and some 748, 1128, 1310 and 1582 cm^{-1} represent cytochrome c bonds. Much research have shown that most halophilic and tolerant microorganisms accumulate high concentrations of organic osmotic solutes (glycerol, monosaccharides and amino acids) and need necessary energy across ion pumps of cell membrane to protect cells from extreme osmotic pressure (Gunde et al., 2018; Jehlička et al., 2012). This can explain why there are a large number of characteristic peaks of monosaccharides and cytochrome c. It should be pointed out that the SCRS of *Oceanobacillus* with different growth periods are not collected in this study. Previous studies have shown that the internal composition of bacteria will be different under different culture conditions or physiological conditions, but this change will not affect identification of bacteria using Raman spectroscopy at the species level (Huang et al., 2004; Hutsebaut et al., 2004; Vossenberg et al., 2013).

According to the results of amplicon sequencing, two species that may be in a viable but non-culturable

(VBNC) state (Lewis et al., 2020), including *O. senegalensis* (ASV82, ASV150 and ASV161) and *O. indicireducens* (ASV119 and ASV127), were not isolated (Figure 1). Although these species are not dominant in the genus *Oceanobacillus*, there could generate some degree of interference in the test results of *Oceanobacillus* in *Daqu* samples. Hence, further studies will focus on isolating more *Oceanobacillus* species from *Daqu* microbiota, by using culturomics and single-cell sorting technologies such as Raman-activated microbial cell sorting (RACS), optical tweezers and microfluidic cell sorting technology to isolate similar types for cultivation or mini-metagenomics (Gross et al., 2015; Jing et al., 2018). Moreover, with the assistance of this technology, the isolation of other VBNC microorganisms remains an achievable objective. In the future, establishing a dual microbial index or a multi-microbial index across different genera can enhance the monitoring of the microbial community. This approach holds promise for evaluating the fermentation quality of *Daqu*. Overall, this study provides a rapid, non-destructive and label-free approach to detect the composition of *Oceanobacillus* species in *Daqu* starter. The detection process of a *Daqu* sample can be completed within 60 min, from the sample processing to the output of result data analysis. This method has great potential for application during the fermentation process of *Daqu*.

AUTHOR CONTRIBUTIONS

Lei Xu: Conceptualization (equal); data curation (equal); methodology (equal); writing – original draft (equal). **Yuan Liang:** Formal analysis (equal). **Wei E Huang:** Writing – original draft (equal); writing – review and editing (equal). **Lin-Dong Shang:** Methodology (equal); software (equal). **Li-Juan Chai:** Writing – original draft (equal); writing – review and editing (equal). **Xiao-Juan Zhang:** Writing – original draft (equal); writing – review and editing (equal). **Jin-Song Shi:** Writing – original draft (equal); writing – review and editing (equal). **Bei Li:** Methodology (equal); software (equal). **Yun Wang:** Conceptualization (equal); project administration (equal); writing – review and editing (equal). **Zheng-Hong Xu:** Conceptualization (equal); project administration (equal); writing – review and editing (equal). **Zhen-Ming Lu:** Conceptualization (equal); project administration (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Wei E Huang https://orcid.org/0000-0003-1302-6528
Zheng-Hong Xu https://orcid.org/0000-0003-1340-6838
Zhen-Ming Lu https://orcid.org/0000-0002-6956-2867

REFERENCES

Arend, N., Pittner, A., Ramoji, A., Mondol, A.S., Dahms, M., Ruger, J. et al. (2020) Detection and differentiation of bacterial and fungal infection of neutrophils from peripheral blood using Raman spectroscopy. *Analytical Chemistry*, 92(15), 10560–10568. Available from: https://doi.org/10.1021/acs.analchem.0c01384

Barańska, M., Roman, M., Dobrowolski, J.C., Schulz, H. & Baranski, R. (2012) Recent advances in Raman analysis of plants: alkaloids, carotenoids, and Polyacetylenes. *Current Analytical Chemistry*, 9, 108–127. Available from: https://doi.org/10.2174/1573411011309010108

Bhargava, R., Fernandez, D.C., Hewitt, S.M. & Levin, I.W. (2006) High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data. *Biochimica et Biophysica Acta*, 1758(7), 830–845. Available from: https://doi.org/10.1016/j.bbamem.2006.05.007

Chai, L.J., Lu, Z.M., Zhang, X.J., Ma, J., Xu, P.X., Qian, W. et al. (2019) Zooming in on butyrate-producing clostridial consortia in the fermented grains of baijiu via gene sequence-guided microbial isolation. *Frontiers in Microbiology*, 10, 1397. Available from: https://doi.org/10.3389/fmicb.2019.01397

Chai, L.J., Shen, M.N., Sun, J., Deng, Y.J., Lu, Z.M., Zhang, X.J. et al. (2020) Deciphering the D−/L-lactate-producing microbiota and manipulating their accumulation during solid-state fermentation of cereal vinegar. *Food Microbiology*, 92, 103559–103569. Available from: https://doi.org/10.1016/j.fm.2020.103559

Chen, Y., Li, K., Liu, T., Li, R., Fu, G., Wan, Y. et al. (2020) Analysis of difference in microbial community and physicochemical indices between surface and central parts of Chinese special-flavor *baijiu Daqu*. *Frontiers in Microbiology*, 11, 592421–592433. Available from: https://doi.org/10.3389/fmicb.2020.592421

Cui, D., Kong, L., Wang, Y., Zhu, Y. & Zhang, C. (2022) In situ identification of environmental microorganisms with Raman spectroscopy. *Environmental Science and Ecotechnology*, 11, 10087–10099. Available from: https://doi.org/10.1016/j.ese.2022.100187

De Gelder, J., De Gussem, K., Vandenabeele, P. & Moens, L. (2007) Reference database of Raman spectra of biological molecules. *Journal of Raman Spectroscopy*, 38(9), 1133–1147. Available from: https://doi.org/10.1002/jrs.1734

Du, H., Wang, X., Zhang, Y. & Xu, Y. (2019) Exploring the impacts of raw materials and environments on the microbiota in Chinese Daqu starter. *International Journal of Food Microbiology*, 297, 32–40. Available from: https://doi.org/10.1016/j.ijfoodmicro.2019.02.020

Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R. & Koltay, P. (2015) Technologies for single-cell isolation. *International Journal of Molecular Sciences*, 16(8), 16897–16919. Available from: https://doi.org/10.1007/978-981-10-4857-9_9-1

Gunde, C.N., Plemenitaš, A. & Oren, A. (2018) Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiology Reviews*, 42(3), 353–375. Available from: https://doi.org/10.1093/femsre/fuy009

He, G., Huang, J., Zhou, R., Wu, C. & Jin, Y. (2019) Effect of fortified *Daqu* on the microbial community and flavor in Chinese

strong-flavor liquor brewing process. *Frontiers in Microbiology*, 10, 56. Available from: https://doi.org/10.3389/fmicb.2019.00056

Ho, T.K. (1998) The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. Available from: https://doi.org/10.1109/34.709601

Huang, W.E., Griffiths, R.I., Thompson, I.P., Bailey, M.J. & Whiteley, A.S. (2004) Raman microscopic analysis of single microbial cells. *Analytical Chemistry*, 76(15), 4452–4458. Available from: https://doi.org/10.1021/ac049753k

Hutsebaut, D., Maquelin, K., de Vos, P., Vandenabeele, P., Moens, L. & Puppels, G.J. (2004) Effect of culture conditions on the achievable taxonomic resolution of Raman spectroscopy disclosed by three *bacillus* species. *Analytical Chemistry*, 76(21), 6274–6281. Available from: https://doi.org/10.1021/ac049228l

Jehlička, J., Oren, A. & Edwards, H.G.M. (2012) Raman spectra of osmotic solutes of halophiles. *Journal of Raman Spectroscopy*, 43(8), 1134–1140. Available from: https://doi.org/10.1002/jrs.3136

Jing, X., Gou, H., Gong, Y., Su, X., Xu, J., Ji, Y. et al. (2018) Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. *Environmental Microbiology*, 20(6), 2241–2255. Available from: https://doi.org/10.1111/1462-2920.14268

Kang, J., Xue, Y., Chen, X. & Han, B. (2022) Integrated multi-omics approaches to understand microbiome assembly in *Jiuqu*, a mixed-culture starter. *Comprehensive Reviews in Food Science and Food Safety*, 21(5), 4076–4107. Available from: https://doi.org/10.1111/1541-4337.13025

Kloss, S., Kampe, B., Sachse, S., Rösch, P., Straube, E., Pfister, W. et al. (2013) Culture independent Raman spectroscopic identification of urinary tract infection pathogens: a proof of principle study. *Analytical Chemistry*, 85(20), 9610–9616. Available from: https://doi.org/10.1021/ac401806f

Kumar, M.A., Kumar, J., Pandey, R., Gupta, S., Kumar, M., Bansal, G. et al. (2017) Comparative genomics of host-symbiont and free-living *Oceanobacillus* species. *Genome Biology and Evolution*, 9(5), 1175–1182. Available from: https://doi.org/10.1093/gbe/evx076

Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z. & Ettema, T.J.G. (2020) Innovations to culturing the uncultured microbial majority. *Nature Reviews Microbiology*, 19, 225–240. Available from: https://doi.org/10.1038/s41579-020-00458-8

Li, P., Lin, W.T., Liu, X., Wang, X. & Luo, L. (2016) Environmental factors affecting microbiota dynamics during traditional solid-state fermentation of Chinese *Daqu* starter. *Frontiers in Microbiology*, 7, 1237. Available from: https://doi.org/10.3389/fmicb.2016.01237

Li, Z., Chen, L., Bai, Z., Wang, D., Gao, L. & Hui, B. (2015) Cultivable bacterial diversity and amylase production in two typical light-flavor Daqus of Chinese spirits. *Frontiers in Life Science*, 8(3), 264–270. Available from: https://doi.org/10.1080/21553769.2015.1041188

Ling, Y., Li, W., Tong, T., Li, Z., Li, Q., Bai, Z. et al. (2020) Assessing the microbial communities in four different Daqus by using PCR-DGGE, PLFA, and biolog analyses. *Polish Journal of Microbiology*, 69(1), 27–37. Available from: https://doi.org/10.33073/pjm-2020-004

Mao, F., Huang, J., Zhou, R., Qin, H., Zhang, S.Y., Cai, X.B. et al. (2022) Effects of different *Daqu* on microbial community domestication and metabolites in nongxiang baijiu brewing microecosystem. *Frontiers in Microbiology*, 13, 939904. Available from: https://doi.org/10.3389/fmicb.2022.939904

Nam, J.H., Bae, W. & Lee, D.H. (2008) *Oceanobacillus caeni* sp. nov., isolated from a bacillus-dominated wastewater treatment system in Korea. *International Journal of Systematic and Evolutionary Microbiology*, 58(5), 1109–1113. Available from: https://doi.org/10.1099/ijs.0.65335-0

Namwong, S., Tanasupawat, S., Lee, K.C. & Lee, J.S. (2009) *Oceanobacillus kapialis* sp. nov., from fermented shrimp paste in Thailand. *International Journal of Systematic and Evolutionary Microbiology*, 59(9), 2254–2259. Available from: https://doi.org/10.1099/ijs.0.007161-0

Raman, C.V.S. & Krishnan, K.S. (1928) A new type of secondary radiation. *Nature*, 121, 501–502. Available from: https://doi.org/10.1039/C3AY40289D

Ripley, B.D. (2015) Support functions and datasets for Venables and Ripley's MASS. 7.3-53.

Shi, W., Chai, L.J., Fang, G.Y., Mei, J.L., Lu, Z.M., Zhang, X.J. et al. (2022) Spatial heterogeneity of the microbiome and metabolome profiles of high-temperature Daqu in the same workshop. *Food Research International*, 156, 111298. Available from: https://doi.org/10.1016/j.foodres.2022.111298

Stoltzfus, J.C. (2011) Logistic regression: a brief primer. *Academic Emergency Medicine*, 18(10), 1099–1104. Available from: https://doi.org/10.1111/j.1553-2712.2011.01185.x

Talari, A.C.S., Movasaghi, Z., Rehman, S. & Rehman, I.U. (2014) Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 50(1), 46–111. Available from: https://doi.org/10.1080/05704928.2014.923902

Thanapun, T. (2013) Screening and characterization of protease-producing *Virgibacillus*, *Halobacillus* and *Oceanobacillus* strains from Thai fermented fish. *Journal of applied pharmaceutical. Science*, 54(10), 1098–1109. Available from: https://doi.org/10.7324/japs.201300563

Tolstik, T., Marquardt, C., Matthäus, C., Bergner, N., Bielecki, C., Krafft, C. et al. (2014) Discrimination and classification of liver cancer cells and proliferation states by Raman spectroscopic imaging. *Analyst*, 139(22), 6036–6043. Available from: https://doi.org/10.1039/c4an00211c

Tominaga, T., An, S.Y., Oyaizu, H. & Yokota, A. (2009) *Oceanobacillus soja* sp. nov. isolated from soy sauce production equipment in Japan. *Journal of General and Applied Microbiology*, 55(3), 225–232. Available from: https://doi.org/10.2323/jgam.55.225

Vossenberg, J.V.D., Tervahauta, H.A., Maquelin, K., Blokker, K.C.H.W., Uytewaal, A.M., Kooij, D.V.D. et al. (2013) Identification of bacteria in drinking water with Raman spectroscopy. *Analytical Methods*, 5, 2679–2687. Available from: https://doi.org/10.2323/jgam.55.225

Wakisaka, Y., Suzuki, Y., Iwata, O., Nakashima, A., Ito, T., Hirose, M. et al. (2016) Probing the metabolic heterogeneity of live Euglena gracilis with stimulated Raman scattering microscopy. *Nature. Microbiology*, 1, 16124. Available from: https://doi.org/10.1038/nmicrobiol.2016.124

Wang, H.-Y., Gao, Y.-B., Fan, Q. & Xu, Y. (2011) Characterization and comparison of microbial community of different typical Chinese liquor Daqus by PCR–DGGE. *Letters in Applied Microbiology*, 53, 134–140. Available from: https://doi.org/10.1111/j.1472-765X.2011.03076.x

Wang, K., Chen, L., Ma, X., Ma, L., Chou, K.C., Cao, Y. et al. (2020) *Arcobacter* identification and species determination using Raman spectroscopy combined with neural networks. *Applied and Environmental Microbiology*, 86(20), e00924-20. Available from: https://doi.org/10.1128/AEM.00924-20

Wang, S., Wu, Q., Nie, Y., Wu, J. & Xu, Y. (2019) Construction of synthetic microbiota for reproducible flavor compound metabolism in Chinese light-aroma-type liquor produced by solid-state fermentation. *Applied and Environmental Microbiology*, 85(10), e03090-18. Available from: https://doi.org/10.1128/AEM.03090-18

Wang, Y., Song, Y., Thompson, I.P., Xu, J. & Huang, W.E. (2015) Single-cell metabolomics. In: *Hydrocarbon and lipid microbiology protocols*. Springer protocols handbooks, Vol. 151. Berlin: Springer. Available from: https://doi.org/10.1007/8623_2015_151

Wu, Q., Chen, L. & Xu, Y. (2013) Yeast community associated with the solid state fermentation of traditional Chinese Maotai-flavor

MICROBIAL BIOTECHNOLOGY

Applied Microbiology International

liquor. *International Journal of Food Microbiology*, 166(2), 323–330. Available from: https://doi.org/10.1016/j.ijfoodmicro.2013.07.003

Xu, J., Yu, T., Zois, C.E., Cheng, J.X., Tang, Y., Harris, A.L. et al. (2021) Unveiling cancer metabolism through spontaneous and coherent Raman spectroscopy and stable isotope probing. *Cancers*, 13(7), 1718. Available from: https://doi.org/10.3390/cancers13071718

Yang, L., Fan, W. & Xu, Y. (2023) Chameleon-like microbes promote microecological differentiation of *Daqu*. *Food Microbiology*, 109, 104144. Available from: https://doi.org/10.1016/j.fm.2022.104144

Yang, Y., Wang, S.T., Lu, Z.M., Zhang, X.J., Chai, L.J., Shen, C.H. et al. (2021) Metagenomics unveils microbial roles involved in metabolic network of flavor development in medium-temperature *daqu* starter. *Food Research International*, 140, 110037. Available from: https://doi.org/10.1016/j.foodres.2020.110037

Zhang, H., Wang, L., Tan, Y., Wang, H., Yang, F., Chen, L. et al. (2021) Effect of *Pichia* on shaping the fermentation microbial community of sauce-flavor *baijiu*. *International Journal of Food Microbiology*, 336, 108898. Available from: https://doi.org/10.1016/j.ijfoodmicro.2020.108898

Zhang, S., Li, X., Zong, M., Zhu, X. & Wang, R. (2018) Efficient kNN classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, 29, 1774–1785. Available from: https://doi.org/10.1109/TNNLS.2017.2673241

Zhang, Y., Xu, J., Ding, F., Deng, W., Wang, X., Xue, Y. et al. (2022) Multidimensional profiling indicates the shifts and functionality of wheat-origin microbiota during high-temperature *Daqu* incubation. *Food Research International*, 156, 111191. Available from: https://doi.org/10.1016/j.foodres.2022.111191

Zhang, Z.M., Chen, S. & Liang, Y.Z. (2010) Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135(5), 1138–1146. Available from: https://doi.org/10.1039/b922045c

Zheng, X.W., Yan, Z., Han, B.Z., Zwietering, M.H., Samson, R.A., Boekhout, T. et al. (2012) Complex microbiota of a Chinese "*fen*" liquor fermentation starter (*fen-Daqu*), revealed by culture-dependent and culture-independent methods. *Food Microbiology*, 31(2), 293–300. Available from: https://doi.org/10.1016/j.fm.2012.03.008

Zou, W., Zhao, C. & Luo, H. (2018) Diversity and function of microbial community in Chinese strong-flavor baijiu ecosystem: a review. *Frontiers in Microbiology*, 9, 671. Available from: https://doi.org/10.3389/fmicb.2018.00671

SUPPORTING INFORMATION

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