Rapid identification of *Lactobacillus* strains using near-infrared spectroscopy and chemometrics

Jiyong Shi[a], Xuetao Hu[a], Fang Zhang[a], Xiaobo Zou[a, *], Mel Holmes[b], Zhihua Li[a], Yiwei Xu[a], Wen Zhang[a], Xiaowei Huang[b]

[a] School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

[b] School of Food Science and Nutrition, the University of Leeds, Leeds LS2 9JT, United Kingdom

*Corresponding author. Tel: +86 511 88780085; Fax: +86 511 88780201

Email address: zou_xiaobo@ujs.edu.cn (xiaobo Zou)

Highlights

NIR spectroscopy was applied for identification of *Lactobacillus* strains according to the discrepancy of NIR acquired from bacterial colonies.

PCA and HCA were investigated for general discriminant analysis of *Lactobacillus*.

UVE and GA were performed for selection of characteristic wavenumbers.

Sensitivity in identification of *Lactobacillus* strains was excellent (92.857%) using UVE-GA-LS-SVM with 10 wavenumbers.

IR spectra of bacterial colonies and cells were researched and compared in order to demonstrate the feasibility of using NIR spectra acquired from bacterial colonies for identification of *Lactobacillus*.

Abstract:

Lactobacillus (*L.*) plays an important role in food fermentation, while the presence of some particular *Lactobacillus* species may decrease the quality of fermented food products. In this study, near-infrared (NIR) spectral features of *Lactobacillus* species were extracted and the feasibility in rapid identification of *Lactobacillus* based on these NIR spectral features was investigated. Bacterial colonies of four *Lactobacillus* species (*L. breris*, *L. casei*, *L. fermentum*, *L. reuteri*) were cultured using spread-plate technique with MRS agar medium. Raw NIR spectral data of bacterial colonies were acquired in the wavelength range of 4,000-10,000 cm^{-1}. After pre-processing, uninformative variables elimination (UVE) and genetic algorithm (GA) were used to select the characteristic wavelengths correlated with the four species of *Lactobacillus*. The NIR data corresponding to the selected characteristic wavelengths were employed to build identification models for discriminating the four species of *Lactobacillus* using Least Squares Support Vector Machine (LS-SVM). The recognition rates of calibration set and prediction set using the optimal identification model were 100% and 92.857%, respectively. In order to explain scientifically the good results of NIR, mid infrared (MIR) spectra of bacterial cells were collected and analyzed. Analytical results of MIR indicated that (1) significant differences were observed in MIR spectral data collected from the four species of *Lactobacillus;* (2) the selected NIR wavenumbers were quite correlated with the MIR wavelengths that could reflect changes of components and structure among four *Lactobacillus* cells. This explained why the four species of *Lactobacillus* were reasonably well identified based on its NIR data. It is concluded that NIR spectroscopy combining with chemometrics methods proposed in this paper could be applied for rapid discrimination of *Lactobacillus*.

Keywords: NIR spectroscopy; *Lactobacillus*; Uninformative variables elimination: Genetic algorithm; Least Squares support vector machine

1. Introduction

The lactic acid bacteria (LAB) are rod-shaped bacilli or cocci characterized by an increased tolerance to a lower pH range(*1*). Various genera of lactic acid bacteria used as probiotics can improve the intestinal immune status, and maintain microbial balance during gastrointestinal disturbances(*2*). *Lactobacillus* (*L.*) is one of the most important genera in LAB fermenting glucose primarily to lactic acid, CO_2 and ethanol(*3*). *Lactobacillus* are widely used in food fermentation of yogurt(*4*), cheese(*5*), soybean meal(*6*), etc.(*7, 8*). However, some *Lactobacillus* species*,* such as *L. brevis*, have been recognized as spoilage bacteria in food processes such as beer fermentation(*9*). Uncontrolled growth of *Lactobacillus* in beer fermentation process may decrease the organoleptic quality of the beer (turbidity, sediment, acidification, off-flavor and ropiness), and even affect its safety(*9, 10*). Therefore, it is necessary to identify the specie of *Lactobacillus* in food production.

There is differentiation in chemical and physical properties between bacterial strains so that they can be differentiated from one another. Traditional approaches for identifying *Lactobacillus* are phenotypic method (*5, 11, 12*) and molecular biological method (*13-15*). In the phenotypic method, a lot of phenotypic bacteria characteristics are used to identify an isolate of *Lactobacillus*. These phenotypic bacteria characteristics include morphological features(*16*) (shape, size, color, dimensions, form, etc.), physiological features(*5*) (modes of fermentation, acid/alkaline/salt tolerance, content of lactic acid dehydrogenases, etc.), and biochemical features(*17*) (the type of interbridging in the peptidoglycan, type of fatty acids and proteins, etc.). In general, about 17 phenotypic tests are required to acquire these phenotypic bacteria characteristics, which makes the phenotypic method is time-consuming, tedious, and involve numerical preparation procedures(*18*). Researchers also pointed out that the reliability of these tests has been questioned(*16*). The exact identifications of these closely related species were not reliable; some were doubtful or unacceptable and some strains were misidentified with a good identification level(*19*). In the molecular biological method, different molecular features based on molecular characterization techniques (genotyping, multilocus sequence typing (MLST)(*20*), pulsed-field gel electrophoresis (PFGE)(*21*) , and ribotyping(*22*)) are used to identify an isolate of *Lactobacillus(19)*. The molecular biological method has been a powerful tool that have helped microbiologists to detect the smallest variations within microbial species even within individual strains. However, the molecular biological method requires advanced instruments and very well trained hands for elaborated sample preparation (*14, 23*). Therefore, there is a need to investigate a rapid and simple technology for identification of bacterial strains.

Usually, optical methods, such as mid infrared (MIR) spectroscopy and near infrared (NIR) spectroscopy, can obtain bacteria characteristics more quickly than phenotypic method and molecular biological method. Optical methods have been employed to rapidly identify bacterial strains (*24-27*). Recently, MIR and NIR were used to record the spectral features of different bacterial strains, and bacterial strains were identified based on these spectral features due to the fact that these spectral features could reflect differences in chemical components (proteins, fatty acids, nucleic acids, etc.)(*28-30*) between bacterial strains. Various researches have shown that MIR can be used to differentiate and identify a number of microorganism at different taxonomic levels, such as, *Lactobacilli* strains(*31*), *Filamentous Fungi* strains(*26*), *Brettanomyces bruxellensis* strains(*32*) et.al. However, it is difficult to acquire MIR spectra of bacteria directly owing to bacterial cell which could not be seen by naked eye. In their researches, the following procedures were applied for preparation of bacterial sample in order to obtain the corresponding MIR spectra of strains: (1) strain was placed into MRS broths and incubated; (2) bacterial broth was centrifuged and supernatant discarded; (3) the bacterial pellet was dried under moderate vacuum to obtain a transparent bacterial film before spectral acquisition. Due to the complicated and time-consuming bacterial preparation which are indispensable before acquisition of MIR spectra, it is no longer a simple and fast method to identify strains using MIR spectroscopy. NIR spectroscopy technique, as an alternative rapid detection method, has been investigated the feasibility in identification and classification of *Escherichia coli* strains(*30*), *Pseudomonas* *aeruginosa* strains(*33*), *Bacillus amyloliquifaciens* strains(*34*), *Bacillus cereus* strains(*34*), *Listeria innocua* strains*(30)* et.al. In order to overcome the difficulties of spectral acquisition using the NIR spectroscopy on account of strains with small size, the NIR spectra were acquired from bacterial suspension after resulting pellets re-suspended into a series of serial dilutions. The resulting pellets were obtained from the centrifugation of bacterial broths after inoculation and culture. Although the sample preparation for acquisition of NIR spectra is simpler than the sample preparation for acquisition of MIR spectra. It is difficult to identify all species at the same time when more than one specie simultaneously exist in the bacterial suspension. Therefore, it would restrict the application of NIR in rapid identification of strains.

It is noted that the colony is enriched with a large number of bacterial cells owing to self-replication of single strain and the classification of different strains maybe performed by the unique NIR spectra of each strain. When the bacterial colonies at different species are cultured on agar plate simultaneously, NIR spectra of each colony are obtained. Therefore, rapid and simultaneous classification of multiple strains could be performed by the NIR spectra. To validate the feasibility of this idea, the following researches were investigated: (1) NIR spectra of four *Lactobacillus* were acquired from bacterial colonies by NIR spectroscopy; (2) methodology for rapid identification of *Lactobacillus* was developed by NIR spectral data and chemometrics; (3) MIR spectra were obtained to prove the validation of identification using the NIR spectra acquired from bacterial colony.

2. Materials and methods

2.1 Strains and preparation of samples

Four species of *Lactobacillus* (*L. breris*, *L. casei*, *L. fermentum*, *L. reuteri*) from Fermentation Laboratory of the school of Food and Biological Engineering, Jiangsu University were used in this study. Four *Lactobacillus* were stored in glycerol at -80°C prior to use. *Lactobacillus* grew on a tube containing 25 mL deMan-Rogosa-Sharpe (MRS) broth and incubated at 37°C for 24h. Each broth was diluted at a ratio of 1:9 (broth: distilled water). Following repeated dilution, 1 ml of the each dilution was extracted and smeared with a sterilized spreader onto MRS agar plates, which was incubated at 37°C for 36h. Three replicates from each *Lactobacillus* were cultivated and prepared in dependent assays to produce three independent sample. Therefore, 21 plates for each *Lactobacillus* were obtained. Bacterial colonies in plates were used to collect NIR spectral data for further analysis.

2.2 NIR spectral acquisition

The Antaris^{TM} II near-infrared spectrophotometer (Thermo Electron Corporation, USA) was employed for spectral acquisition over the 10,000 cm^{-1} to 4000 cm^{-1} at intervals of 3.856 cm^{-1} which resulted in 1557 variables. The plates, containing individual *Lactobacillus* colonies, was placed on sample stage of NIR spectroscopy for acquisition of bacterial spectra. Five spectra were collected from five discontinuous region in individual cultured plate with the diffuse reflectance integrating sphere mode. The average of five spectra from same plate was computed and used in further data processing. Finally, 84 NIR spectra (21 spectra for each *Lactobacillus*) from 84 plates respectively can be considered independent samples and used for identification of *Lactobacillus*.

2.3 Chemometrics methods

Due to the NIR spectra significantly influenced by non-linearities and baseline shift introduced by light scatter, suitable pre-processing should be applied to eliminate these effects largely. For spectral data compression and information extraction, PCA were performed to transform the spectral data after pre-processing into new uncorrelated variable called principal components (PCs). HCA was applied with the PCs as input variables to investigate similarities between different bacterial samples and reveal if there are natural clustering in bacterial samples. Employing the full spectral data does not always yield optimal results as the full variables may include variables not correlated with colony features, partially correlated with colony features and highly correlated with colony features. Therefore, UVE combined with GA called UVE-GA is employed for characteristic variables selection in which UVE is employed for elimination of uninformative variables and GA is for selection of characteristic variables. LS-SVM was implemented in this study for classification purposes. After the LS-SVM models were developed, they were applied to classify bacterial samples. The data analysis mentioned above were performed by MATLAB 2010a after NIR spectral collection and steps of data analysis were shown in Fig.1.

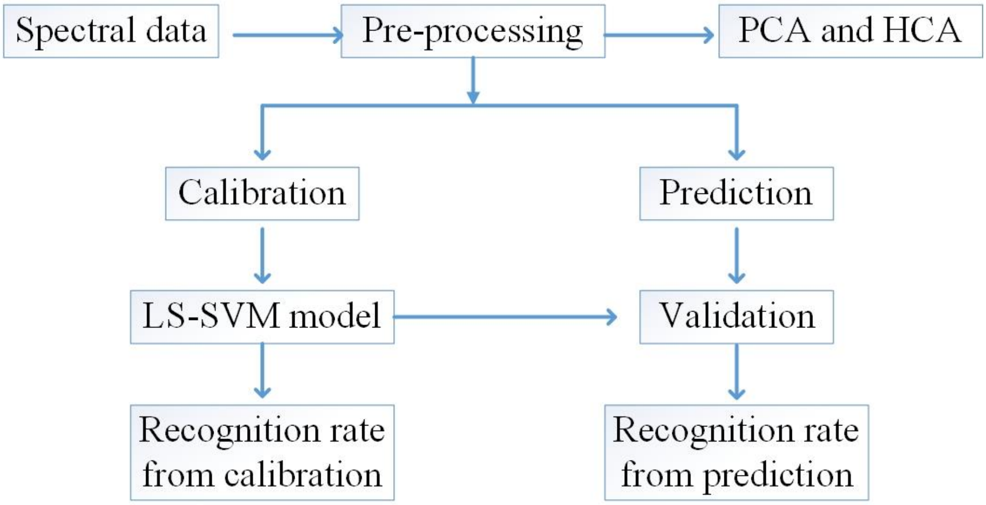

Fig.1 Flowchart of data analysis in identification of *Lactobacillus*

2.3.1 Pre-processing methods

All pre-processing technique have the goal of reducing the un-modeled variability noise in data in order to enhance the feature in the spectra. However, there is always the danger of applying the wrong type or applying a severe processing that will remove the valuable information(*35*). In this study, standard normal variate (SNVT), multiplicative scatter correction (MSC), first derivative (1D) and second derivative (2D) were performed to correct for light scattering, modify the additive and multiplicative effects, and remove the influence of any baseline variation(*36*).

2.3.2 PCA and HCA

The PCA is a technique that is used for spectral data compression and information extraction by the reduction of the data dimensionality and generation of new uncorrelated variables called PCs(*37*). The PCs that account for a large percentage of total variance reveal most information of *Lactobacillus* (*30*). HCA is a unsupervised pattern recognition technique that can be used for clustering of strains based on the similarities between spectra (*37*). The HCA was performed with the PCs as input data using Ward’s clustering algorithm and the Squared Euclidean Distance Measure to generate a dendrogram(*38, 39*).

2.3.3 Variable selection methods (UVE, GA)

Uninformative variables elimination (UVE) is employed to eliminate uninformative variables that clearly have no information about *Lactobacillus* in this study. The artificial random variables are added to the calibration data as a reference so that those spectral variables which play a less important role in model than the random variables are eliminated(*40*). Compared to other methods, it keeps a large number of the partially relevant variables for finally modeling. Therefore, GA was employed to reduce the number of variables in order to achieve a simpler, more robust model. GA is a popular heuristic optimization technique that employs a probabilistic, non-local search process. Many studies have demonstrated the importance of GA for characteristic wavelength selection(*41*).

2.3.4 Pattern recognition method (LS-SVM)

Before building classification models, the spectra data were randomly assigned to calibration and test set according to the proportion of 2:1. For supervised classification models, least squares support vector machine (LS-SVM) was implemented. By applying LS-SVM classifiers, the empirical classification error can be minimized while ensuring maximization of the interclass geometric boundary(*37*).

2.3 Development of identification models

The full-spectrum LS-SVM model using calibration set was obtained based on full spectrum and LS-SVM. The UVE-LS-SVM model was obtained based on the characteristic wavelengths selected by UVE. The characteristic wavelengths selected by UVE and GA were applied to build UVE-GA-LS-SVM model. And then models are validated by test sets. The performance of model are evaluated via recognition rates of calibration and test sets(*30*).

3. Result and discussion

3.1 NIR spectra investigation

The spectra in the range of 4000-10000 cm^{-1} illustrated that the absorption decrease with the wavenumber increasing expect some spectral peaks (Fig.1). As can be seen in Fig.1 (a), average NIR spectra of each *Lactobacillus* has no significant discrepancy instead of the intensity. Three obvious absorption peaks around 5350 cm^{-1}, 7150 cm^{-1} and 8650 cm^{-1} are all based on molecular overtone and combination vibrations. MSC, 1D, 2D, SNVT, MSC+1D, MSC+2D, SNV+1D and SNV+2D were applied for the multivariate data to eliminate spectrum interference factors due to the influence of noise or the diffuse reflectance. LS-SVM based on raw spectral data and the spectral data after eight kinds of pre-processing methods was employed to select most suitable pre-processing method according to the recognition rates showing in Table 1. It was found that the MSC performed relatively better than the other noted methods for separation of the *Lactobacillus* (Table 1). Therefore, MSC was selected to be employed for further process in this study. The average spectra of each *Lactobacillus* after MSC was shown in Fig.1 (b).

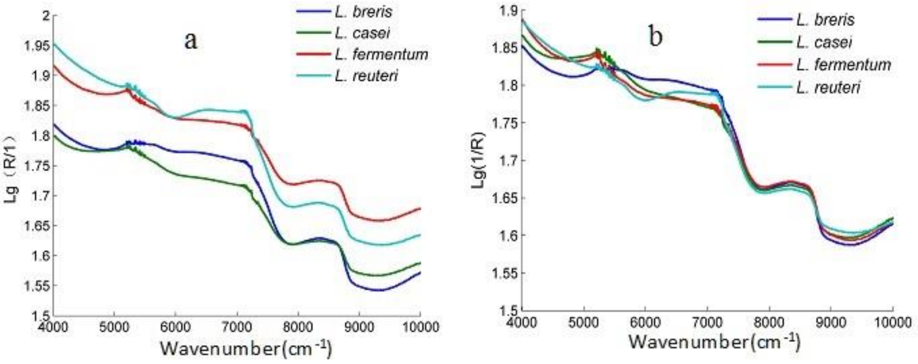

Fig.1. the raw average NIR spectra of four strains (a) and the NIR spectra after MSC (b)

3.2 Clustering analysis using PCA and HCA

The NIR spectral data of four *Lactobacillus* (81 objects × 1557 variables) after MSC were submitted to PCA to extract the effective information. New uncorrelated variables were created, and the scores plot of PC1 versus PC2 display clustering for *Lactobacillus* (Fig.2a). The first two principal components together accounted for approximately 83.81% of the total variance (PC1: 54.13% and PC2: 29.68%). Based on Fig.2a, a number of spectra for each *Lactobacillus* are overlapped. This weak separation was obtained and maybe owing to the first two PCs that contained insufficient information for classification. Therefore, more PCs and more chemometrics, such as variable selection, supervised pattern recognition, should be attempted to identify *Lactobacillus*. The loading plots of the first two PCs were investigated to show wavenumbers making great contribution to the variation in the data set. In Fig.2 (b), the peaks at 5905cm^{-1}, 7471cm^{-1}, 8354cm^{-1} influencing the PC1 significantly could be the important wavenumbers that have high correlation with the features of each *Lactobacillus*. Furthermore, the PC2 also was responsible for the separation of *L. breris* and *L. fermentum*. The corresponding loading plot of PC2 shown in Fig.2 (c) manifest that 5153cm^{-1}, 5427cm^{-1}, 5839cm^{-1}, 6808cm^{-1}, 7633cm^{-1}, 8616cm^{-1}, associated with O-H, N-H structure, are the critical wavenumbers for such separation.

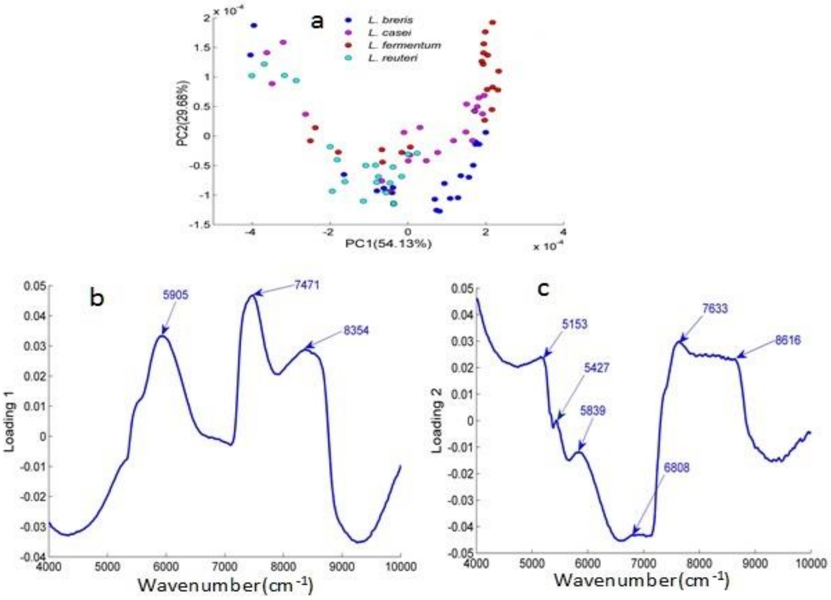

Fig.2. PCA results: (a) score plot of PC1 versus PC2, (b) the loading plot for PC1, and (c) the loading plot for PC2

Fig.3 shows the dendrogram obtained after carrying out the HCA. The first ten PCs, which account for 90% of total variance in dataset were used as the input data. Ward’s algorithm and squared Euclidian distance was selected to investigate the dissimilarities between the spectra of four strains. In Fig.3, the left vertical axis of dendrogram depicts the names of 40 bacterial samples. The second column on the left is the number from 1 to 40, in which from 1 to 10 represent *L. breris*, 11 to 20 represent *L. casei*, 21 to 30 represent *L. fermentum*, and 31 to 40 represent *L. reuteri*. The upper horizontal axis represents the distance between two bacterial samples or two cultures. The magnitude of this distance depends on the number of spectra in a cluster and the similarities between them. Two major clusters were illustrated in dendrogram. The first cluster (two clusters) only included *L. breris* and the second cluster was made up of two well distinguished subclusters, the first subcluster with *L. fermentum* (10 strain), the second subcluster with *L. reuteri* (10 strains) and *L. casei* (10 strains). The second subcluster is divided into two clusters, one is *L. reuteri*, and the other one is *L. casei*. Only single *L. reuteri* was wrongly assigned to the cluster of *L. casei*. The dendrogram in this paper does not completely follow the classical scheme (homofermentative and heterofermentative *Lactobacillus*) using FT-IR spectroscopy(*29*). In addition, the dendrogram provided in this study was also not similar with phylogenetic tree of *Lactobacillus* (*42, 43*),

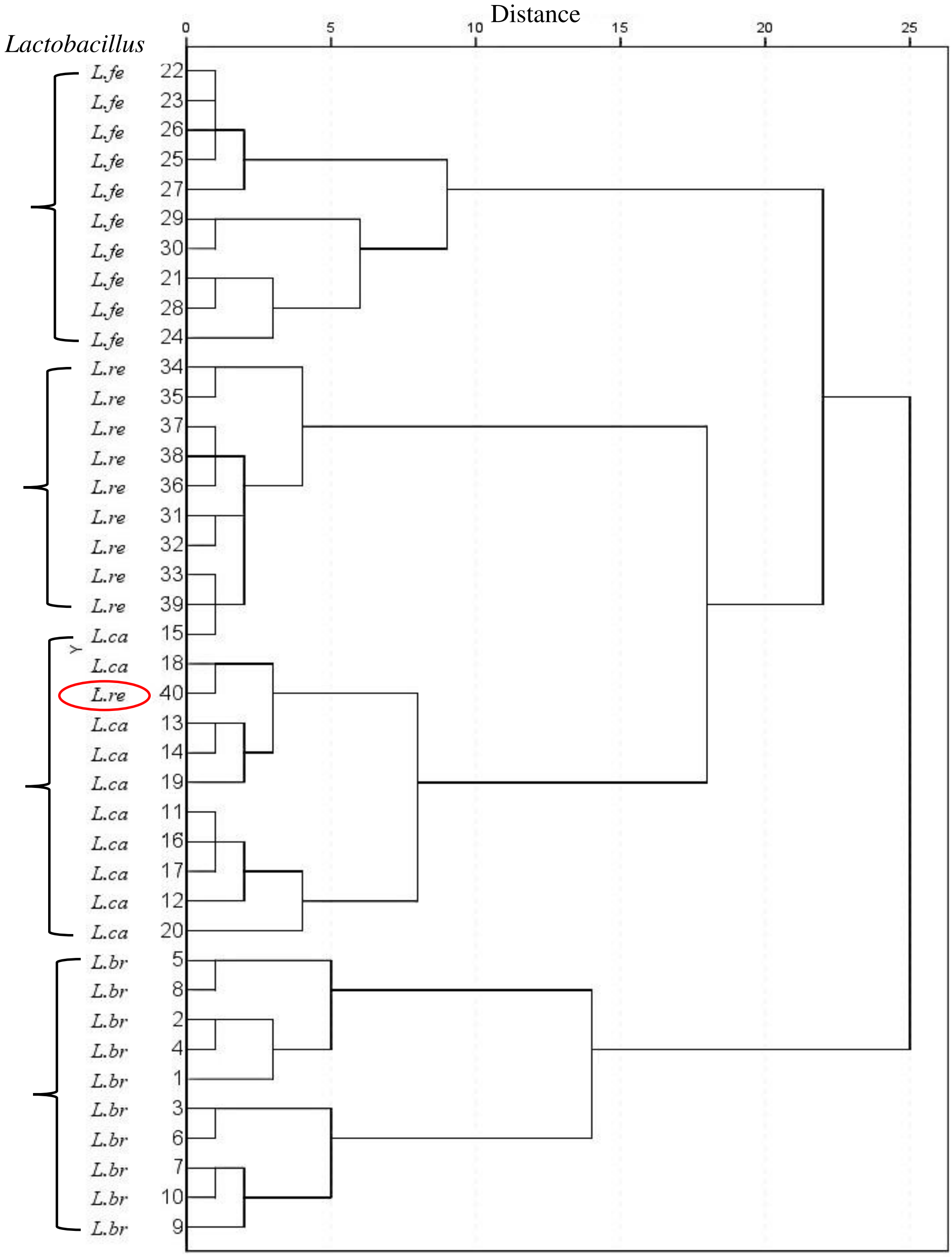

Fig.3 Dendrogram from hierarchical cluster analysis of four *Lactobacillus*

3.3 Development of identification models using UVE, GA, and LS-SVM

In this study, 56 samples were used for calibration and 28 samples were used for prediction data. Classification results based on full spectrum data after eight different pre-processing were illustrated in Tab.1. As indicated in Tab.1, the highest accuracy of LS-SVM model was based on MSC with recognition rate of 85.714% for test set and 98.214% for the calibration set. The second high recognition rate for prediction set is 82.143% when SNV is used. In the best model, four bacterial samples were mistakenly classified. One *L. fermentum* was mistakenly predicted as *L. casei*, one *L. breris* was mistakenly predicted as *L. fermentum*, one *L. fermentum* was mistaken for *L. reuteri*, and *one L. casei* was mistaken for *L. brevis*.

Owing to the high-dimensional data containing highly correlated variables, UVE and GA were employed in this research for selecting characteristic wavenumbers reflecting the features of *Lactobacillus*. It could be advantageous to use only few variables for accurate, simple and robust classification. The variables that do not contain more information than the random variables will be regarded as uninformative variables and eliminated using UVE. To minimize the uninformative variables the random variables would be optimized with several random variables in different orders of magnitude, such as 10^{-8}, 10^{-9}, 10^{-10}, 10^{-11}, 10^{-12}. After UVE in conjunction with LS-SVM, the optimal order of magnitude in random variables was 10^{-11}. As indicated in Fig.4 (a), the stability range of the added random variables with order of magnitude of 10^{-11} (the last 1557 wavenumbers) is from 3 to -3. From the observation of first 1557 variables, 700 variables whose stability lies out of the two dot lines will be remained for LS-SVM models and else variables whose stability lies within the dot lines will be eliminated. The remnant 700 wavenumbers were mainly located between 4000 and 7000cm^{-1}. Most of the variables between 7000 and 10000 were eliminated. As a result, recognition rate of UVE-LS-SVM model achieved 96.428% shown in Fig.4 (b). Only one *L. breris* was mistaken for *L. casei*. The UVE-LS-SVM model based on 700 wavenumbers selected by UVE is more stable and accurate than model based on full spectrum

To further eliminate most of insignificant variables, GA was utilized to select the characteristic wavenumbers highly correlated with the features of *Lactobacillus*. 10 characteristic wavenumbers (4119cm^{-1}, 4427cm^{-1}, 4317cm^{-1}, 5292cm^{-1}, 5905cm^{-1}, 6184cm^{-1}, 6527cm^{-1}, 6969cm^{-1}, 8531cm^{-1}, 8601cm^{-1}) were selected by GA based on remaining 700 wavelengths. The 10 wavenumbers were labeled on the spectrum shown in Fig.5. With the 10 characteristic wavelengths, UVE-GA-LS-SVM model achieved 92.857% for the recognition rate of prediction and 100% for the recognition rate of calibration. The recognition rates of UVE- GA-LS-SVM model were between that of the LS-SVM model in full-spectrum and that of UVE-LS-SVM model. But UVE- GA-LS-SVM with the least wavenumbers (10 wavenumbers) is simpler and more robust than UVE-LS-SVM and LS-SVM models with 700 and 1557 wavenumbers, respectively. As shown in Fig.5 (b), the classification result of UVE-GA-LS-SVM model demonstrates that one *L. breris* was mistaken for *L. casei* and one *L. reuteri* was mistakenly predicted as *L. fermentum*.

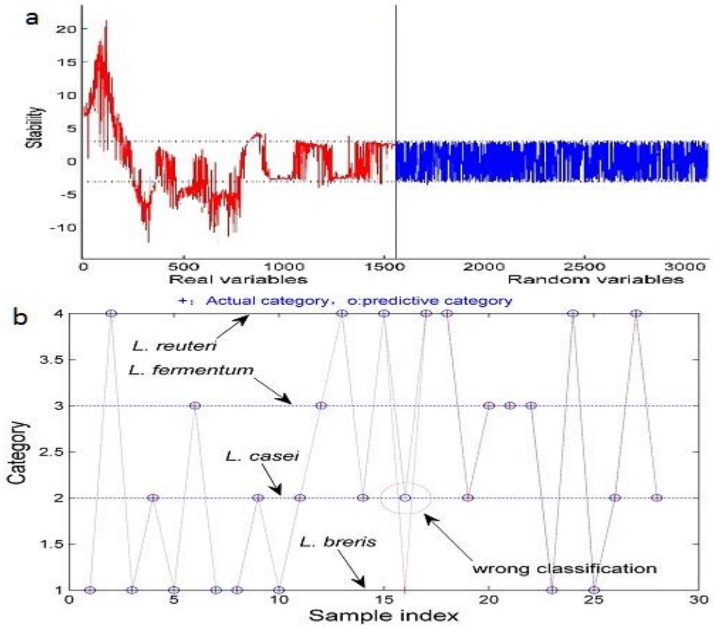

Fig.4. (a): the stability distribution of each variable for identification of strains by UVE. The two dot lines in (a) indicate the lower and upper threshold; (b): model performance for prediction.

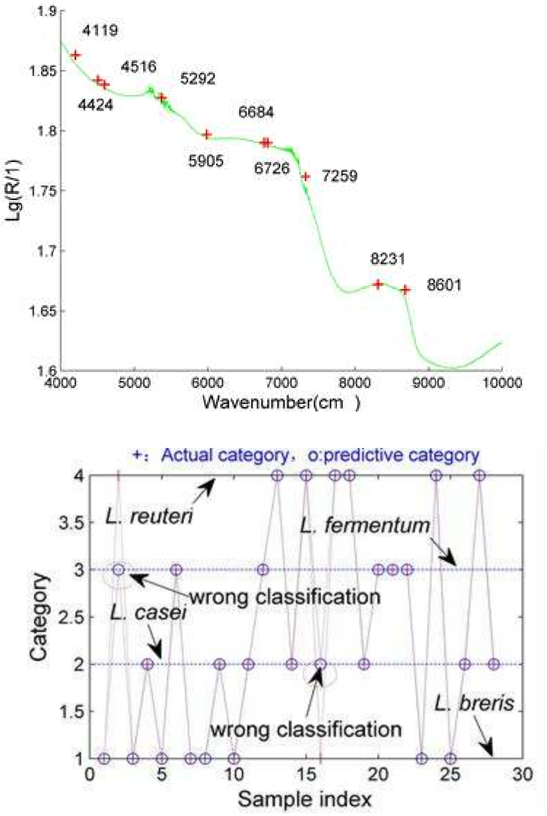

Fig.5. (a) the ten red plus signs indicative the position of 10 characteristic wavenumbers in spectra by UVE and GA. (b) model performance for prediction based on UVE-GA-LS-SVM.

Tab.1. overall recognition rates (%) for identification of four *Lactobacillus* based on full spectrum LS-SVM using different pre-processing methods

Chemometrics methods	LS-SVM	
	calibration	prediction
None	85.714	82.857
SNV	98.214	88.571
MSC	98.214	91.429
1D	89.286	82.857
2D	89.286	82.857
SNV-1D	92.857	85.714
SNV-2D	92.857	85.714
MSC-1D	96.429	85.714
MSC-2D	98.214	85.714
MSC-UVE	98.214	96.428
MSC-UVE-GA	98.214	92.857

3.4 Scientific explanation of established identification models

In this study, MIR spectroscopy was employed to investigate the components and structure of bacterial cells and the absorption bands of each *Lactobacillus* was used to validate the difference of components and structure among *Lactobacillus*. MIR spectra of 4 bacterial cells were displayed in Fig.6a. Absorption bands of four *Lactobacillus* were located at the region of 1040-1090cm^{-1}, 1230-1240 cm^{-1},1390-1400, 1440-1450, 1530-1560 cm^{-1}, 1630-1640 cm^{-1}, 2850-1990 cm^{-1}, 3250-3260 cm^{-1} owing to the inconsistent peaks of each spectrum. The characteristic absorption in frequency at 3250-3260 cm^{-1} in the frequency region of 3150 cm^{-1}-3500cm^{-1} is attributed to the O-H bonded stretching vibration. 2850-1990 cm^{-1} are in the range of 3000-2800cm-1 influenced by C-H stretching vibrations in fatty acid and some amino acid. The C=O stretching vibration of amides linked to proteins make contribution to the absorption peak at 1630-1640 cm^{-1} (amide I) and the absorption peak at 1530-1560cm^{-1} is likely due to the N-H deformation of amides linked to proteins (amide II). The peaks between 1500~1200 cm^{-1}, namely 1230-1240 cm^{-1}, 1390-1400 and 1440-1450 cm^{-1} were possibly influenced by CH_2 and CH_3 bending modes of proteins, fatty acids and phosphate-carrying compounding. The 1040-1090cm^{-1} between 1200 cm^{-1} and 900 cm^{-1} were due to the symmetric stretching vibration of PO_2^- groups found in nucleic acids (*44, 45*). Bacterial cells maybe all contain proteins, peptides, fatty acid, and polysaccharides et al based on these absorption peaks. However, intensity of peaks and the highest point of peaks among the four spectra was different. That means the groups, components and the content of them are different. Therefore, *Lactobacillus* could be classified based on the difference of groups, components and the content of them.

For acquisition of MIR spectroscopy, bacterial samples, potassium bromide (KBr) pellets were prepared. KBr pellets were the most common sample for acquisition of MIR spectra and described by D.J.M. Mouwen et.al(*24*). About 2 mg colonies and 200 mg KBr powder were homogenized in an agate-stone mortar. The mixture was made into a coin shaped pellet. Finally, six KBr pellets for each *Lactobacillus* were obtained and used for acquisition of MIR spectra. The MIR spectra were measured with Nicolet 380 FT-IR spectrometer (Thermo Electron Corporation, USA). MIR spectra were acquired in the spectral range of 500 to 4000 cm^{-1} at resolution of 2 cm^{-1}. The colonies and coin shaped pellets were positioned and directly contacted with an infrared attenuated total reflection diamond. Six spectra were acquired for each sample and each spectrum were composed of an average of 36 separate scans. Finally, six spectra were averaged. Although FT-IR has been well known as a technique for identification of bacteria Bacterial samples preparation for MIR spectroscopy requires a large number of bacterial cells, centrifugation, lyophilization and pressing. On account of the complicated and time-consuming sample preparation, the application of NIR in rapid identification of strains were restricted.

In this study, the NIR spectra were acquired from colonies on agar plates not the KBr pellets. This sample preparation leave out most of complicated procedures (centrifugation, lyophilization and pressing) except bacterial culture. The selected ten characteristic wavenumbers for identification of strains were also influenced by functional groups in fatty acid, peptides and proteins et al. 5905 cm^{-1} is the first overtone of Carbon-hydrogen (C-H); 8531 and 8601 are the weaker second overtone of C-H; 6184 and 6527cm^{-1} are the first overtone of Nitrogen–hydrogen (N-H); 6969 cm^{-1} is the first overtone of Oxygen–hydrogen (O-H); 5292cm^{-1} is the second miscellaneous overtone band of carbonyl group in peptides; 4119, 4427 and 4317 cm^{-1} are the C-H combination bands. The bands of C-H may be linked with fatty acid and some amino acid. The bands of O-H may be linked with water. The bands of N-H may be linked with proteins and peptides. The bands of miscellaneous overtone maybe linked to peptides. Relationship between the ten characteristic wavenumbers in NIR spectra and the absorption bands in MIR spectra was also investigated. The frequency of 5905 cm^{-1} is included in frequencies of two times of the region of 2850-2990 cm^{-1}. The frequencies of 8531 and 8601 cm^{-1}are included in frequencies of three times of the region of 1210-1260cm^{-1}. The frequencies of 6184 and 6527 cm^{-1} are included in frequencies of three times of the region of 1530-1560 and 1630-1640 cm^{-1}respectively. The frequency of 5292 cm^{-1} is included in frequencies of three times of the region of 1630-1640 cm^{-1}. The frequencies of 4119, 4427 and 4317 cm^{-1} are the combination of 1230-1240 cm^{-1}, 1390-1400 cm^{-1}, 1440-1450 and 2850-2990 cm^{-1}. So we consider that of 5905, 8531 and 8601 cm^{-1} are influenced by functional groups of membrane fatty acid and by some amino acid on account of C-H stretching vibrations. 6184 and 6527 cm^{-1} are affected by amide I and amide II groups belong to proteins and peptides due to N-H stretching. 5292cm^{-1} is the second miscellaneous overtone band of carbonyl group in peptides owing to C=O stretching. 4119 cm^{-1}, 4427 cm^{-1} and 4317 cm^{-1} are the C-H combination bands influenced by vibration of C-H. The NIR spectra at ten wavenumbers, influenced by the components and structure in bacterial cells, are unique for each *Lactobacillus* owing to the different components and structure in diverse *Lactobacillus*. Therefore, single strain could be identified used the unique NIR spectra at ten wavenumbers.

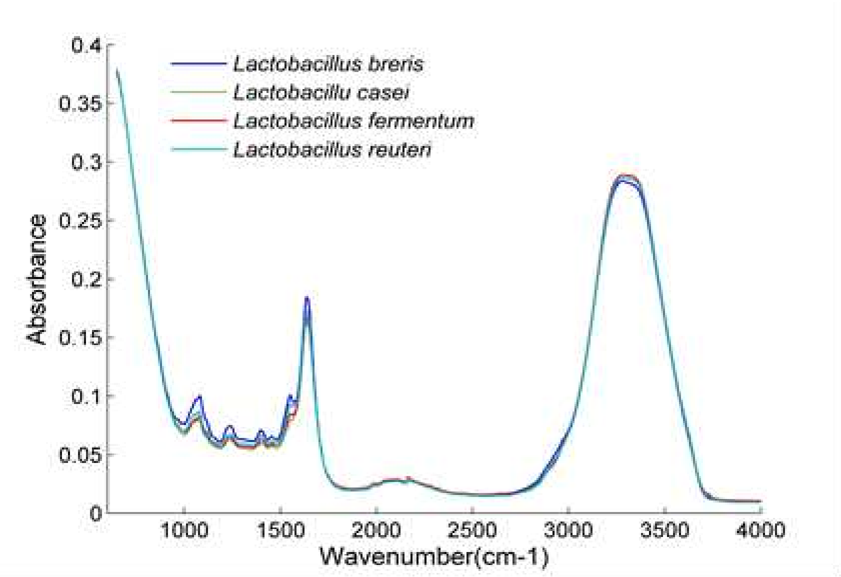

Fig.6. The average IR spectra of bacterial cells (a) and bacterial colonies (b)

Table 2 Tentative assignment of characteristic wavenumbers found in NIR spectra

Frequency (cm^{-1})	Related frequency in MIR	Assignment
5905, 8531, 8601	2850-2990, 1210-1260	CH, CH_2, CH_3
6184, 6527	1530-1560	Amide II band, C=O
5292	1630-1640	Amide I, N-H
4119, 4427, 4317	1230-1240, 1390-1400	Combination of C-H

4. Conclusion

This paper presented a rapid method to discriminate four *Lactobacillus* strains based on NIR spectroscopy technique aided by chemometric methods. The NIR spectra were acquired from bacterial colonies on MRS agar medium. Rapid identification of *Lactobacillus* were developed with the pre-treatment (MSC), variables selection (UVE and GA), and supervised discriminant analysis (LS-SVM) performed. MSC based on full wavenumber and LS-SVM showed the best performance in all pre-processing methods with the recognition rate of 91.429%. Utilization of UVE and GA, resulting in 10 wavenumber that have high correlation with the features of *Lactobacillus*, could simplify the identification models and improved the performance. The recognition rates of UVE- GA-LS-SVM model (92.857%) were between that of the LS-SVM model in full-spectrum (91.429%) and that of UVE-LS-SVM model (96.428%). The UVE-GA-LS-SVM model with 10 wavenumbers is simpler and more robust than the full-spectrum LS-SVM model with 1557 wavenumbers and UVE-LS-SVM model with 700 wavenumbers. By comparing with absorption bands in MIR spectra, the ten characteristic wavenumbers are influenced by functional groups in components and structure. Each bacteria has a unique NIR spectra due to the stretching and bending vibrations of molecular bends or functional groups presented in cellular components (proteins, nucleic, lipids, etc.). Therefore, single strain could be identified for the unique NIR spectrum. This methodology may become a powerful tool for identification of strains due to timely spectral collection and high sensitivity for identification strains.

Acknowledgements

The authors would like to acknowledge the financial support from the national natural science foundation of China (Grant No.61301239), the national science and technology support program (2015BAD19B03), the natural science foundation of Jiangsu province (BK20130505), China postdoctoral science foundation (2013M540422, 2014T70483), Science foundation for postdoctoral in Jiangsu province (1301051C), Chinese 863 Program (Grant Nos. 2011AA108007), the Jiangsu province science fund for distinguished young scholars (BK20130010), New century excellent talents in university (NCET-11-0986), Specially-appointed professors by universities in Jiangsu province, Research foundation for advanced talents in Jiangsu University (13JDG039), Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Reference

1. Patel, S. J., A comprehensive review on Probiotics. *Int. J. Pure App. Biosci* **2015,** *3*, 286-290.
2. Zhai, Q.; Yin, R.; Yu, L.; Wang, G.; Tian, F.; Yu, R.; Zhao, J.; Liu, X.; Chen, Y. Q.; Zhang, H., Screening of lactic acid bacteria with potential protective effects against cadmium toxicity. *Food Control* **2015,** *54*, 23-30.
3. Hammes, W. P.; Vogel, R. F., The genus lactobacillus. In *The genera of lactic acid bacteria*, Springer: 1995; pp 19-54.
4. Ng, E. W.; Yeung, M.; Tong, P. S., Effects of yogurt starter cultures on the survival of Lactobacillus acidophilus. *International Journal of Food Microbiology* **2011,** *145*, 169-175.
5. Cogan, T. M.; Barbosa, M.; Beuvier, E.; BIANCHI-SALVADORI, B.; COCCONCELLI, P. S.; FERNANDES, I.; GOMEZ, J.; GOMEZ, R.; KALANTZOPOULOS, G.; LEDDA, A., Characterization of the lactic acid bacteria in artisanal dairy products. *Journal of Dairy Research* **1997,** *64*, 409-421.
6. Wang, L.; Zhou, H.; He, R.; Xu, W.; Mai, K.; He, G., Effects of soybean meal fermentation by Lactobacillus plantarum P8 on growth, immune responses, and intestinal morphology in juvenile turbot (Scophthalmus maximus L.). *Aquaculture* **2016,** *464*, 87-94.
7. Haghshenas, B.; Nami, Y.; Abdullah, N.; Radiah, D.; Rosli, R.; Khosroushahi, A. Y., Anticancer impacts of potentially probiotic acetic acid bacteria isolated from traditional dairy microbiota. *LWT-Food Science and Technology* **2015,** *60*, 690-697.
8. Waites, M. J.; Morgan, N. L.; Rockey, J. S.; Higton, G., *Industrial microbiology: an introduction*. John Wiley & Sons: 2009.
9. Geissler, A. J.; Behr, J.; von Kamp, K.; Vogel, R. F., Metabolic strategies of beer spoilage lactic acid bacteria in beer. *International journal of food microbiology* **2016,** *216*, 60-68.
10. García-Ruiz, A.; Crespo, J.; López-de-Luzuriaga, J.; Olmos, M.; Monge, M.; Rodríguez-Alfaro, M.; Martín-Alvarez, P.; Bartolome, B.; Moreno-Arribas, M., Novel biocompatible silver nanoparticles for controlling the growth of lactic acid bacteria and acetic acid bacteria in wines. *Food Control* **2015,** *50*, 613-619.
11. de Almeida Júnior, W. L. G.; da Silva Ferrari, Í.; de Souza, J. V.; da Silva, C. D. A.; da Costa, M. M.; Dias, F. S., Characterization and evaluation of lactic acid bacteria isolated from goat milk. *Food Control* **2015,** *53*, 96-103.
12. Nair, P. S.; Surendran, P. K., Biochemical characterization of lactic acid bacteria isolated from fish and prawn. **2005**.

13. Zhao, J.; Fleet, G., The effect of lactic acid bacteria on cocoa bean fermentation. *International journal of food microbiology* **2015,** *205*, 54-67.

14. Singh, S.; Goswami, P.; Singh, R.; Heller, K. J., Application of molecular identification tools for Lactobacillus, with a focus on discrimination between closely related species: a review. *LWT-Food Science and Technology* **2009,** *42*, 448-457.

15. Sohier, D.; Coulon, J.; Lonvaud-Funel, A., Molecular identification of Lactobacillus hilgardii and genetic relatedness with Lactobacillus brevis. *International Journal of Systematic and Evolutionary Microbiology* **1999,** *49*, 1075-1081.

16. Coeuret, V.; Dubernet, S.; Bernardeau, M.; Gueguen, M.; Vernoux, J. P., Isolation, characterisation and identification of lactobacilli focusing mainly on cheeses and other dairy products. *Le Lait* **2003,** *83*, 269-306.

17. Montville, T. J.; Matthews, K. R., *Food microbiology: an introduction*. ASM Press: 2005.

18. Moreira, J. L. S.; Mota, R. M.; Horta, M. F.; Teixeira, S. M.; Neumann, E.; Nicoli, J. R.; Nunes, Á. C., Identification to the species level of Lactobacillus isolated in probiotic prospecting studies of human, animal or food origin by 16S-23S rRNA restriction profiling. *BMC microbiology* **2005,** *5*, 1.

19. Adeyemo, S.; Onilude, A., Molecular identification of Lactobacillus plantarum isolated from fermenting cereals. *International Journal of Biotechnology and Molecular Biology Research* **2014,** *5*, 59-67.

20. Anukam K C, O. E. O., Ahonkhai I, 16S rRNA gene sequence and phylogenetic tree of Lactobacillus species from the vagina of healthy Nigerian women. *African Journal of Biotechnology* **2005,** *4*, 1222-1227.

21. Tenover F C, A. R. D., Goering R V, Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *Journal of clinical microbiology* **1995,** *33*, 2233.

22. Bidet, P., Barbut, F., Lalande, V., Burghoffer, B., & Petit, J. C., Development of a new PCR-ribotyping method for Clostridium difficile based on ribosomal RNA gene sequencing. *FEMS microbiology letters* **1999,** *175*, 261-266.

23. Marty, E.; Buchs, J.; Eugster-Meier, E.; Lacroix, C.; Meile, L., Identification of staphylococci and dominant lactic acid bacteria in spontaneously fermented Swiss meat products using PCR–RFLP. *Food microbiology* **2012,** *29*, 157-166.

24. Mouwen, D.; Hörman, A.; Korkeala, H.; Alvarez-Ordóñez, A.; Prieto, M., Applying Fourier-transform infrared spectroscopy and chemometrics to the characterization and identification of lactic acid bacteria. *Vibrational Spectroscopy* **2011,** *56*, 193-201.

25. Grasso, E. M.; Yousef, A. E.; de Lamo Castellvi, S.; Rodriguez-Saona, L. E., Rapid detection and differentiation of Alicyclobacillus species in fruit juice using hydrophobic grid membranes and attenuated total reflectance infrared microspectroscopy. *Journal of agricultural and food chemistry* **2009,** *57*, 10670-10674.

26. Lecellier, A.; Gaydou, V.; Mounier, J.; Hermet, A.; Castrec, L.; Barbier, G.; Ablain, W.; Manfait, M.; Toubas, D.; Sockalingum, G., Implementation of an FTIR spectral library of 486 filamentous fungi strains for rapid identification of molds. *Food microbiology* **2015,** *45*, 126-134.

27. Driver, T.; Bajhaiya, A. K.; Allwood, J. W.; Goodacre, R.; Pittman, J. K.; Dean, A. P., Metabolic responses of eukaryotic microalgae to environmental stress limit the ability of FT-IR spectroscopy for species identification. *Algal research* **2015,** *11*, 148-155.

28. Alvarez-Ordóñez, A.; Mouwen, D. J. M.; López, M.; Prieto, M., Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria. *Journal of Microbiological Methods* **2011,** *84*, 369-378.

29. Bosch, A.; Golowczyc, M. A.; Abraham, A. G.; Garrote, G. L.; De Antoni, G. L.; Yantorno, O., Rapid discrimination of lactobacilli isolated from kefir grains by FT-IR spectroscopy. *International Journal of Food Microbiology* **2006,** *111*, 280-287.

30. Feng, Y.-Z.; Downey, G.; Sun, D.-W.; Walsh, D.; Xu, J.-L., Towards improvement in classification of Escherichia coli, Listeria innocua and their strains in isolated systems based on chemometric analysis of visible and near-infrared spectroscopic data. *Journal of Food Engineering* **2015,** *149*, 87-96.

31. Oust, A.; Møretrø, T.; Kirschner, C.; Narvhus, J. A.; Kohler, A., FT-IR spectroscopy for identification of closely related lactobacilli. *Journal of Microbiological Methods* **2004,** *59*, 149-162.

32. Oelofse, A.; Malherbe, S.; Pretorius, I. S.; Du Toit, M., Preliminary evaluation of infrared spectroscopy for the differentiation of Brettanomyces bruxellensis strains isolated from red wines. *International Journal of Food Microbiology* **2010,** *143*, 136-142.

33. Marques, A. S.; Castro, J. N.; Costa, F. J.; Neto, R. M.; Lima, K. M., Near-infrared spectroscopy and variable selection techniques to discriminate Pseudomonas aeruginosa strains in clinical samples. *Microchemical Journal* **2016,** *124*, 306-310.

34. Rodriguez-Saona, L. E.; Khambaty, F. M.; Fry, F. S., &; Calvey, E. M., Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy. *Journal of agricultural and food chemistry* **2001,** *49*, 574-579.

35. Rinnan, Å.; Berg, F. v. d.; Engelsen, S. B., Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **2009,** *28*, 1201-1222.

36. Guo, Y.; Ni, Y.; Kokot, S., Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2016,** *153*, 79-86.

37. Berrueta, L. A.; Alonso-Salces, R. M.; Héberger, K., Supervised pattern recognition in food analysis. *Journal of Chromatography A* **2007,** *1158*, 196-214.

38. Mordehai, J.; Ramesh, J.; Huleihel, M.; Cohen, Z.; Kleiner, O.; Talyshinsky, M.; Erukhimovitch, V.; Cahana, A.; Salman, A.; Sahu, R. K., Studies on acute human infections using FTIR microspectroscopy and cluster analysis. *Biopolymers* **2004,** *73*, 494-502.

39. Xie, C.; Mace, J.; Dinno, M.; Li, Y.; Tang, W.; Newton, R.; Gemperline, P., Identification of single bacterial cells in aqueous solution using confocal laser tweezers Raman spectroscopy. *Analytical chemistry* **2005,** *77*, 4390-4397.

40. Centner, V.; Massart, D.-L.; de Noord, O. E.; de Jong, S.; Vandeginste, B. M.; Sterna, C., Elimination of uninformative variables for multivariate calibration. *Analytical chemistry* **1996,** *68*, 3851-3858.

41. de Sousa Marques, A.; de Melo, M. C. N.; Cidral, T. A.; de Lima, K. M. G., Feature selection strategies for identification of Staphylococcus aureus recovered in blood cultures

using FT-IR spectroscopy successive projections algorithm for variable selection: a case study. *Journal of microbiological methods* **2014,** *98*, 26-30.

42. Pavlova, S. I.; Kilic, A.; Kilic, S.; So, J. S.; Nader ‐ Macias, M.; Simoes, J.; Tao, L., Genetic diversity of vaginal lactobacilli from women in different countries based on 16S rRNA gene sequences. *Journal of Applied Microbiology* **2002,** *92*, 451-459.

43. Song, Y.-L.; Kato, N.; Liu, C.-X.; Matsumiya, Y.; Kato, H.; Watanabe, K., Rapid identification of 11 human intestinal Lactobacillus species by multiplex PCR assays using group-and species-specific primers derived from the 16S–23S rRNA intergenic spacer region and its flanking 23S rRNA. *FEMS Microbiology Letters* **2000,** *187*, 167-173.

44. Alvarez-Ordonez, A.; Mouwen, D.; Lopez, M.; Prieto, M., Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria. *Journal of microbiological methods* **2011,** *84*, 369-378.

45. Man, Y. C.; Mirghani, M. E. S., Rapid method for determining moisture content in crude palm oil by Fourier transform infrared spectroscopy. *Journal of the American Oil Chemists' Society* **2000,** *77*, 631-637.