Journal of Microbiological Methods 165 (2019) 105693

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Identification of lactic acid bacteria *Enterococcus* and *Lactococcus* by near-infrared spectroscopy and multivariate classification

Sylvain Treguier[a,*], Christel Couderc[a], Helene Tormo[a], Didier Kleiber[a], Cecile Levasseur-Garcia[b]

[a] Université de Toulouse, Ecole d'Ingénieurs de Purpan, INPT, 75 voie du T.O.E.C., F-31076 Toulouse, Cedex 03, France
[b] Université de Toulouse, Ecole d'Ingénieurs de Purpan, INPT, LCA, 75 voie du T.O.E.C., F-31076 Toulouse, Cedex 03, France

ARTICLE INFO

Keywords:
Discrimination
Enterococcus
Identification
Lactic acid bacteria
Lactococcus
Near-infrared spectroscopy

ABSTRACT

Lactic acid bacteria are important in numerous biological processes. The fabrication of cheese, for example, uses the lactic acid bacteria found in raw milk such as *Lactococcus lactis* as starters to improve the organoleptic properties of milk. Conventional methods to determine the genus and species of lactic acid bacteria isolated from raw milk involve genotyping and phenotyping, which require specific preparation and sample destruction. To improve on this situation, we present herein a simple and non-destructive screening method to discriminate between the *Lactococcus* and *Enterococcus* species most commonly found in raw milk (*L. lactis*, *E. durans*, *E. faecalis*, and *E. faecium*). The bacteria are grown on agar plates and assessed by using near-infrared spectroscopy in a spectral range from 800 to 2777 nm. Principle component analysis loading line plots highlight the inter-genus and inter-species differences at various wavelengths, which are mostly assigned to cell-wall compounds such as polysaccharides. The best artificial neural network identification models give 98.8% and 86.3% classification rates at the genus and species level, respectively, for an external validation set made of 80 samples. These results suggest that near-infrared spectroscopy may be used to identify lactic acid bacteria on agar medium.

1. Introduction

To fabricate cheese, the lactic acid bacteria (LAB)[1] present in milk are completed with lactic ferments to give these dairy products precise characteristics regarding their preservation, fermentation, and aroma (Bachmann et al., 1996; Centeno et al., 1996; Demarigny et al., 2006). The most common bacterial species isolated from raw milk include the natural lactic starter *Lactococcus lactis* and several *Enterococcus* spp., which can easily be confused because their genotypes and phenotypes are closely related (Badis et al. (2004a); Badis et al. (2004b); Callon et al., 2007; Cheriguene et al., 2007; Edalatian Dovom et al., 2012; Guessas and Kihal, 2004; Mas et al., 2002; Tormo et al., 2015). Thus, the development of starter cultures requires the assessment of isolates at the genus and species level.

The most frequently used diagnoses for bacterial identification include genotypic techniques such as polymerase chain reaction, phenotyping and analysis of ribosomal proteins by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF)[2] mass spectrometry or sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)[3] (Barreiro et al., 2012; Davis, 2014; Sandrin et al., 2013; Soomro and Masud, 2007; Yang et al., 2010). Unfortunately, these conventional methods require sample destruction because of method-specific sample preparations.

As a result, increasing demand has arisen for non-destructive high-throughput screening methods for bacteria present in milk products. Near-infrared spectroscopy (NIRS)[4] is a potential technique for identifying these microorganisms. NIRS is fast, inexpensive, easy to use, and has many applications in the agri-food industry, both *in situ* and in the laboratory.

The use of NIRS to identify bacterial strains was initially explored by Rodriguez-Saona et al. (2001). Membrane-filtered bacterial films were measured and clustered at the genus and strain level by using a

* Corresponding author.
E-mail addresses: sylvain.treguier@purpan.fr (S. Treguier), christel.couderc@purpan.fr (C. Couderc), helene.tormo@purpan.fr (H. Tormo), didier.kleiber@purpan.fr (D. Kleiber), cecile.levasseur@purpan.fr (C. Levasseur-Garcia).

[1] LAB: lactic acid bacteria.
[2] MALDI-TOF: matrix-assisted laser desorption ionization time-of-flight.
[3] SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis.
[4] NIRS: near-infrared spectroscopy.

https://doi.org/10.1016/j.mimet.2019.105693
Received 27 June 2019; Received in revised form 14 August 2019; Accepted 16 August 2019
Available online 19 August 2019

principal component analysis (PCA). Alexandrakis et al. (2008) and de Sousa Marques et al. (2013) investigated whether soft independent modeling of class analogy (SIMCA)[5] and partial least-squares discriminant analysis (PLS-DA)[6] can be used to classify NIR absorbance spectra of bacterial suspensions. Both concluded that PLS-DA coupled with NIRS is well suited for bacterial identification. Alexandrakis et al. obtained 100% correct classification rates in prediction at the genus and species level (including three different Pseudomonas species) and de Sousa Marques et al. obtained 87.5% and 88.3% for *Escherichia coli* and *Salmonella Enteritidis*, respectively. Feng et al. (2015) and Mu et al. (2018) compared linear and nonlinear classification methods coupled with NIRS to identify bacterial strains resuspended in tryptic soy broth. Both found that nonlinear methods yield better results; by using support vector machines (SVM),[7] Feng et al. correctly classified 81.5% of prediction samples from three *E. coli* strains, and Mu et al. used a competitive adaptive reweighted sampling method (CARS)[8] with SVM to correctly classify 100% of the spectra from six strains of various genus and species in cross validation.

Several studies have focused on LABs. Onda et al. (2001) assessed the fermentation type of numerous LAB strains inoculated in de Man, Rogosa, and Sharpe broth (MRS)[9] by applying a linear discriminant analysis with Mahalanobis distances at 2272 and 2311 nm—the former possibly being associated with lactic acid. Based on the first three principal components of a PCA, Cámara-Martos et al. (2011) discriminated between *Leuconostoc mesenteroides*, *Lactobacillus sakei*, and *Lactobacillus plantarum* strains measured by Fourier-transform NIRS (FT-NIRS).[10] Levasseur-Garcia et al. (2017) used FT-NIRS to discriminate between various *Enterococcus* and *Lactococcus* species grown on Elliker agar plates and obtained an 87% correct classification rate on an independent validation set. LAB fermentation was also monitored by NIRS in several studies using PCA or PLS regression to quantify fermentation-related parameters such as sugar and lactic acid production (Grassi et al., 2013; Liu et al., 2016; Macedo et al., 2002; Svendsen et al., 2017).

Many of these works acquired spectra from bacterial pellets obtained by resuspension and centrifugation of cultures. Spectral effect of nutrient media to the spectra were not reduced for measurements carried out directly on cultures. The objective of the present study is thus to determine whether NIRS may be used to identify bacteria on agar medium and whether decreasing the contribution of agar to the spectra can improve this identification. To do this, we inoculated collection strains originating mainly from cheese and raw milk onto agar plates, and then measured the spectrum of resulting growth under repeatable conditions to limit external biases. Finally, we built chemometric models based on the acquired data to discriminate between the genus and species of the given strains.

2. Materials and methods

2.1. Strain collection

This study used 40 LAB strains, 20 of which belonged to 3 different species from the *Enterococcus* genus: *E. durans*, *E. faecalis*, and *E. faecium*, with the remaining 20 strains belonging to *Lactococcus lactis* (Table A). All strains were acquired from two distinct culture collections: CIRM-BIA (Centre International de Ressources Microbiennes dédié aux Bactéries d'Intérêt Alimentaire, France),[11] and CRBIP (Centre de Ressources Biologiques de l'Institut Pasteur, France),[12] The strains were isolated mostly from dairy products.

Table A
Genus, species and ID of the 40 strains used in this study.

Genus	Species	Collection	Strain collection ID
Enterococcus	*durans*	CIRM-BIA	61, 743 T, 1485
		CRBIP	55,125 T
	faecalis	CIRM-BIA	256, 567, 739 T, 1328 T, 1486
		CRBIP	103,630, 104,055, 104,056
	faecium	CIRM-BIA	259, 499, 502, 503, 504, 505, 506, 1487
Lactococcus	*lactis*	CIRM-BIA	53, 79 T, 80 T, 81 T, 84, 235, 236, 238, 239, 241, 242, 244, 245,247, 248, 633, 644, 1562, 1973, 2008

2.2. Strain storage

Upon reception, all strains were stored at −80 °C. Liquid cultures were made from the freeze-dried CRBIP strains. Each lyophilizate was first rehydrated in 500 µL of MRS broth (Grosseron, France), then revived by inoculating 50 µL of culture into 5 mL of MRS broth and incubated in aerobic conditions at 30 °C for the *Lactococci* strains and at 37 °C for the *Enterococci* strains. A total of 500 µL of each culture was then added to a vial with 500 µL of an 85%–15% milk-glycerol mix and stored again at −80 °C.

Aliquots were made from each strain for this study by inoculating 50 µL of each inoculum in 5 mL of MRS broth and incubating for 12 h in aerobic conditions at their respective optimum growth temperature. Next, 500 µL of each culture was added to three vials with 500 µL of an 85%–15% milk-glycerol mix and stored at −20 °C. Daughter strains were used for all subsequent experiments.

2.3. Sample preparation

All strains were suspended in Bennett broth and incubated overnight at 30 °C in aerobic conditions. Next, 200 µL of each of the 40 precultures was inoculated onto three different Bennett agar plates by using an easySpiral Pro spiral plater (Interscience, Netherlands). Plating was done in constant deposition mode to obtain confluent growth on every dish so that most of the spectral integration surface was related to bacteria. All plates contained 20 mL of Bennett agar and were incubated at 30 °C in aerobic conditions for 72 h. To verify the absence of contaminations during the process, several tubes of Bennett broth were incubated along with precultures, and several control plates were inoculated with this control broth and incubated.

2.4. Near-infrared spectroscopy measurements

After incubation, absorbance spectra of confluent plates were acquired by using a spectrometer. Because NIRS is very sensitive to the presence of water (Dufour, 2009), all dishes were dried under a laminar flow for 5 min to remove condensation.

Spectral acquisitions were made plate by plate with an MPA Fourier-transform NIR spectrometer (Bruker GmbH, Germany). Measurements were done in transmission mode, with the bottom of the dish positioned on the side of the light source and the lid positioned on the side of the transmission unit. A cache was placed over the light source, the transmission unit, and the sample before each acquisition to avoid polluting the spectra with stray light.

For the 40 strains, two different spots were measured on each of the three plates and on the control plates to obtain the background from the

[5] SIMCA: soft independent modeling of class analogy.
[6] PLS-DA: partial least-squares discriminant analysis.
[7] SVM: support vector machines.
[8] CARS: competitive adaptive reweighted sampling method.
[9] MRS: de Man, Rogosa, and Sharpe.
[10] FT-NIRS: Fourier-transform near-infrared spectroscopy.
[11] CIRM-BIA: Centre International de Ressources Microbiennes dédié aux Bactéries d'Intérêt Alimentaire.
[12] CRBIP: Centre de Ressources Biologiques de l'Institut Pasteur.

pure medium.

2.5. Data analysis

The acquired spectra were preprocessed by applying several corrections or combinations of corrections: Savitzky-Golay smoothing with an 11-point window, Savitzky-Golay first derivative with a 61-point window, Savitzky-Golay second derivative with a 121-point window, Savitzky-Golay smoothing with an 11-point window followed by a multiplicative scatter correction (MSC)[13], and Savitzky-Golay smoothing with an 11-point window followed by an extended multiplicative scatter correction (EMSC)[14]. Smoothing with the Savitzky-Golay algorithm consists of averaging data in successive windows to eliminate noise in the spectra. With the Savitzky-Golay derivation algorithms, derivatives are calculated in each window after applying the smoothing filter (Savitzky and Golay, 1964). Derivatives can augment spectral information by increasing peak resolution. The MSC technique corrects additive effects such as spectral offset, and multiplicative effects such as light scattering. For each spectrum, a least-squares regression is first applied to the average spectrum, and then the entire spectrum is corrected by using the slope and intercept of the computed equation (Helland et al., 1995). EMSC is a similar approach except that it uses another spectrum rather than the averaged data for the regression (Martens & Stark, 1991). In this work, EMSC was done with a pure Bennett agar spectrum to reduce the impact of the medium on the spectra.

Raw and pretreated spectra were analyzed by using a PCA to determine whether the components tended to cluster according to genera, species, or any external factor. PCA loadings revealed wavelengths with the most significant impact on clustering (Martens et al., 1987).

Prior to applying discrimination models, 66.7% of the spectra were selected at random to build a calibration set, with the remaining 33.3% assigned to a validation set. Both sets were identical for the raw spectra and six pre-processed spectra datasets. Linear and nonlinear models were developed by applying a PLS-DA and artificial neural networks (ANN)[15]. In the case of ANN, 20% of the spectra were removed from the calibration set to build an internal validation set to train the network. The performance of all models was assessed based on the percent of correctly identified genera and species in the calibration set and validation set and their corresponding confusion matrices (Visa et al., 2011). The best model for each criterion was selected based on the highest percent of correctly identified samples in the validation set.

3. Results

3.1. Exploratory analysis

We acquired the absorbance spectra at two spots on each of the 120 plates to obtain a total of 240 spectra from the samples used in this study, in addition to 24 absorbance spectra from control dishes containing only Bennett medium. Because the extremities of the spectra have significant noise, only the 1150–2675 nm range was retained for data analysis (Fig. A.1).

3.1.1. Exploratory analysis of raw absorbance spectra

Two distinct groups of spectra appear in the raw spectra of the samples (Fig. A.2). These groups have the same spectral characteristics, but one group seems to be biased compared with the other one, with a large positive offset at the lower wavelengths that gradually decreases over the spectrum. This bias is due to differences in blank measurements, room temperature, and the preparation of Bennett agar. It will be eliminated through preprocessing to improve separation at both given levels of the bacterial phylum.

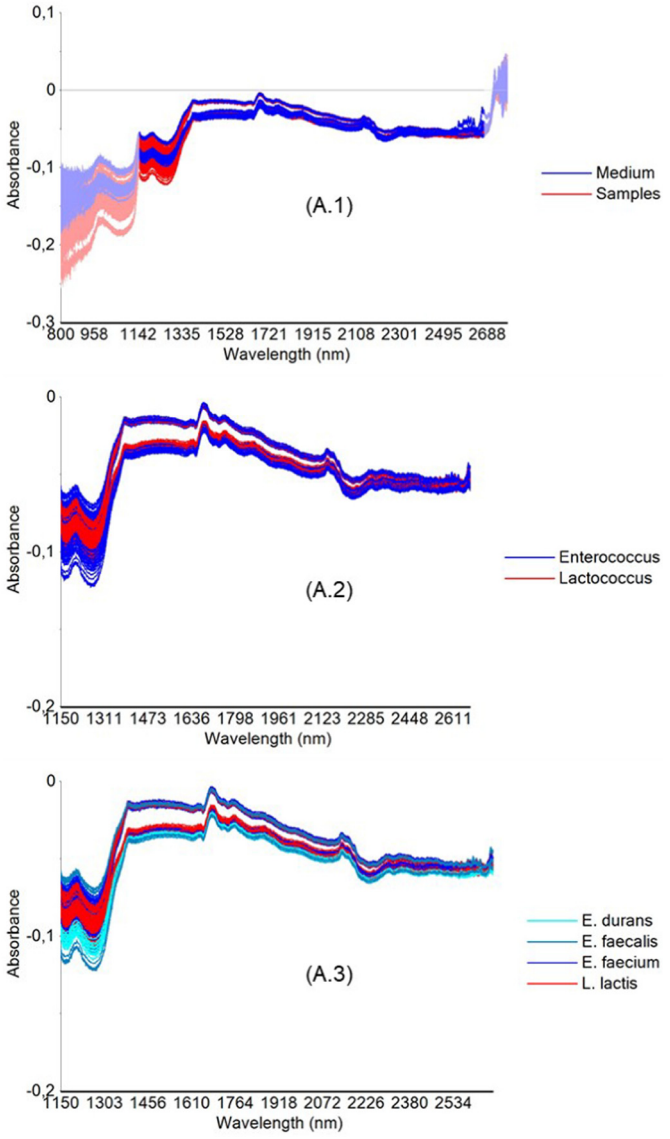

Fig. A. Raw absorbance spectra collected from lactic acid bacteria, and from medium, on entire spectral range (A.1); identification of genera on narrow spectral range: 1150–2675 nm (A.2); identification of species on narrow spectral range (A.3). [1.5 column].

Raw absorbance spectra show differences between the *Enterococcus* and *Lactococcus* genera as well as between the species (Fig. A.2 and 3). Clusters corresponding to species are distinguished on the two first PCs of the PCA (Fig. B).

3.1.2. Exploratory analysis of spectra preprocessed with Savitzky-Golay third derivative

The preprocessing that best separates the genera and the species of bacteria in the PCA is the Savitzky-Golay third derivative (Fig. C). The principal components contributing most to this clustering are PC1 and PC3, which explain 36% and 14% of the total variability of the model, respectively. *Enterococcus* and *Lactococcus* are significantly differentiated by PC1, while species are more differentiated by PC3. The respective loadings of PC1 and PC3 are shown on Fig. D.

[13] MSC: multiplicative scatter correction.
[14] EMSC: extended multiplicative scatter correction.
[15] ANN: artificial neural networks.

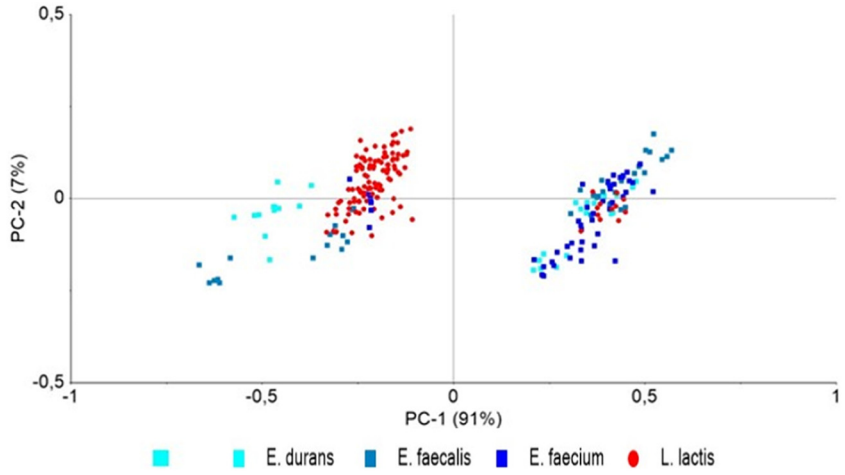

Fig. B. PCA of raw absorbance spectra from samples with different species. [single column].

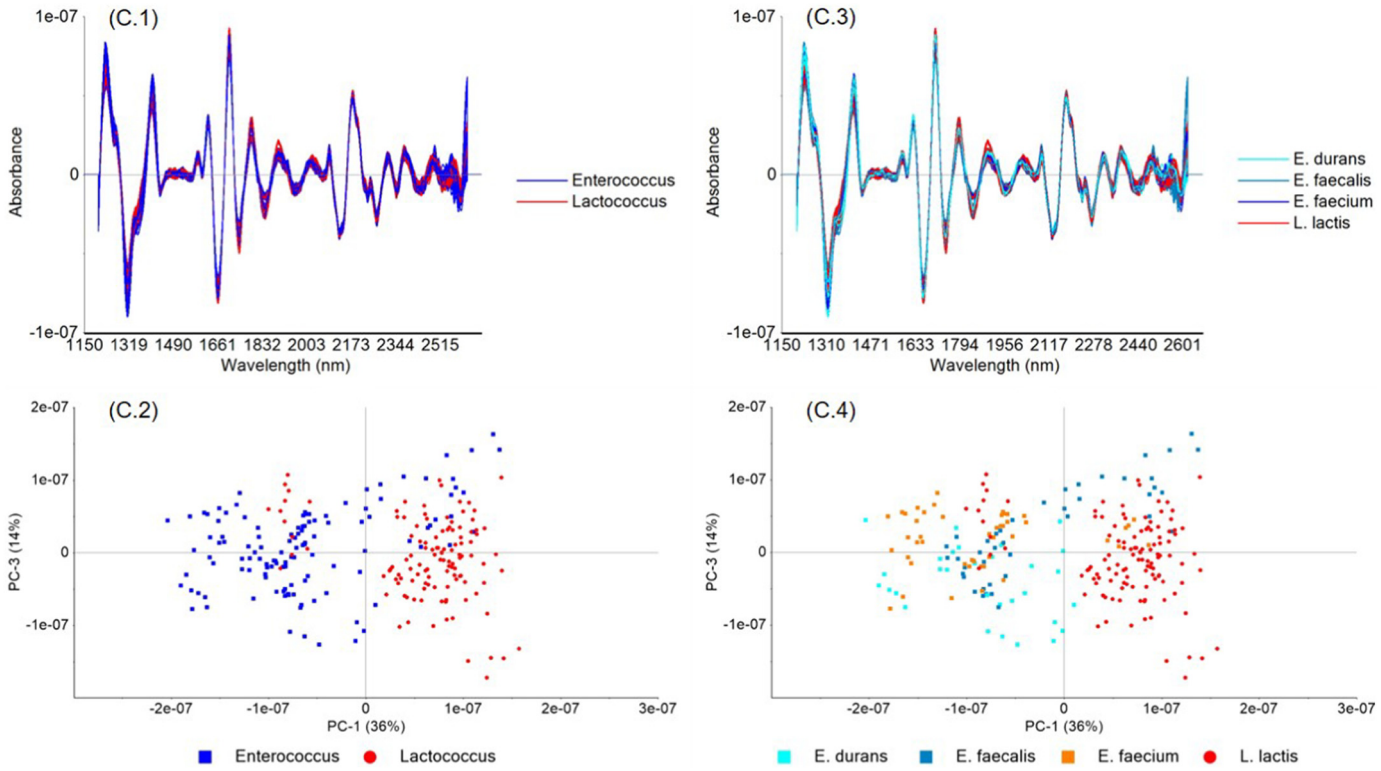

Fig. C. Third-derivative spectra from samples with different genera (C.1) and species (C.3); PCA of third-derivative spectra from samples with different genera (C.2) and species (C.4). [2 column].

3.2. Classification of the spectra regarding genera or species

3.2.1. Classification at the genus level

The raw and preprocessed spectra gave very consistent classification results in calibration and validation, both with PLS-DA (Table B) and ANN (Table C). The most efficient model was obtained by combining an ANN with the Savitzky-Golay smoothing plus an EMSC preprocess. The network architecture consisted of 3086 input neurons, each corresponding to a wavelength of the spectrum, as well as 19 neurons in the hidden layer and 1 output neuron corresponding to the decision between *Enterococcus* and *Lactococcus*. The architecture was trained with 50 iterations, the best of which were selected based on external validation. All the samples were correctly classified in calibration and internal validation, whereas 98.8% of the samples were correctly classified in external validation. The confusion matrices of the three sets are presented in Tables D–F, respectively.

3.2.2. Classification at the species level

The ANN (Table G) leads to significantly better discrimination at the species and subspecies level than PLS-DA (Table H), regardless of the dataset used. Moreover, with both methods, raw and smoothed spectra give better external validation than spectra corrected by MSC and EMSC, which themselves perform better than derivative spectra. The model with the highest classification results in external validation and the closest between the three sets is the ANN model performed with Savitzky-Golay smoothed spectra. The network architecture consists of 3086 input neurons, each corresponding to a given wavelength within the spectrum, in addition to 22 neurons in the hidden layer and 4 output neurons corresponding to the decision between *E. durans*, *E. faecalis*, *E. faecium*, and L. *lactis*. The architecture was trained with 50

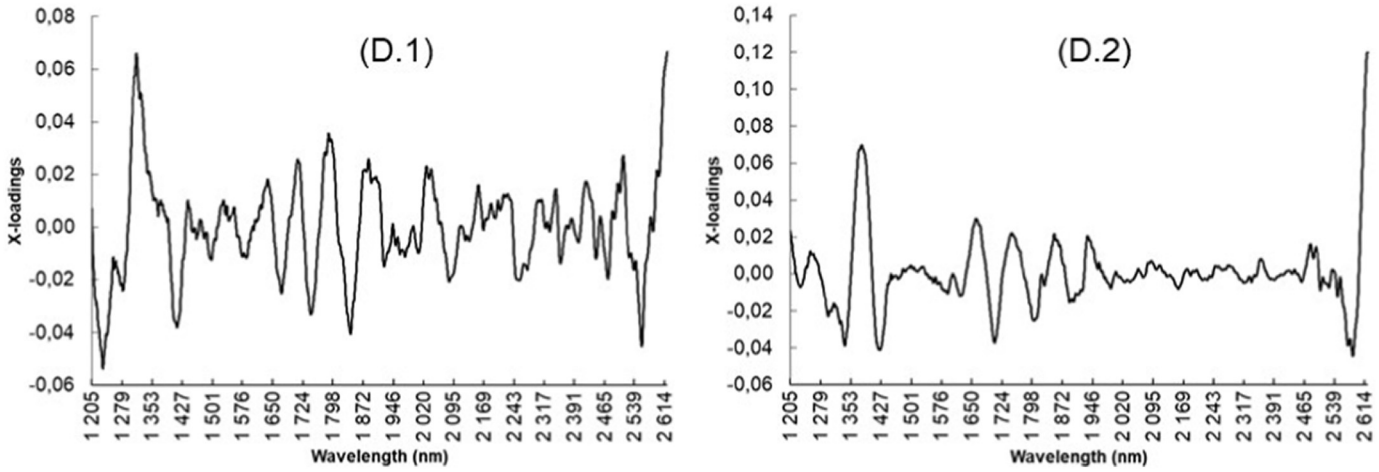

Fig. D. PC1 (D.1) and PC3 (D.2) loadings from PCA of third-derivative spectra. [2 column].

Table B
Performance of discrimination models at the genus level by PLS-DA.

Preprocessing	Principal components	Calibration accuracy	External validation accuracy
Raw spectra	6	93.1%	91.3%
Savitzky-Golay smoothing	6	93.1%	91.3%
Savitzky-Golay smoothing + MSC	6	95.6%	91.3%
Savitzky-Golay smoothing + EMSC	6	95.6%	91.3%
Savitzky-Golay 1st derivative	6	93.1%	91.3%
Savitzky-Golay 2nd derivative	3	93.1%	91.3%
Savitzky-Golay 3rd derivative	2	93.1%	91.3%

Table C
Performance of discrimination models at the genus level by ANN.

Preprocessing	Network architecture	Iterations	Calibration accuracy	Internal validation accuracy	External validation accuracy
Raw spectra	3086-15-1	50	96.1%	93.8%	96.3%
Savitzky-Golay smoothing	3086–23-1	50	96.1%	93.8%	96.3%
Savitzky-Golay smoothing + MSC	3086-9-1	50	100%	93.8%	98.8%
Savitzky-Golay smoothing + EMSC	**3086-19-1**	**50**	**100%**	**100%**	**98.8%**
Savitzky-Golay 1st derivative	3086-2-1	50	95.3%	93.8%	92.5%
Savitzky-Golay 2nd derivative	3443-8-1	50	100%	96.9%	92.5%
Savitzky-Golay 3rd dérivative	2886-11-1	50	100%	96.9%	93.8%

The bold represents the model with the highest classification results.

Table D
Confusion matrix of selected model in calibration.

	Enterococcus (output)	*Lactococcus* (output)
Enterococcus (target)	**60**	0
Lactococcus (target)	0	**68**

The bold are the number of correctly classified samples.

Table E
Confusion matrix of selected model in internal validation.

	Enterococcus (output)	*Lactococcus* (output)
Enterococcus (target)	**14**	0
Lactococcus (target)	0	**18**

The bold are the number of correctly classified samples.

iterations, the best of which were selected based on the performance in external validation and the smallest difference between calibration, internal, and external validation error. A total of 92.2% of the samples were classified correctly in calibration, 87.5% in internal validation, and 86.3% in external validation. The confusion matrices of the three

Table F
Confusion matrix of selected model in external validation.

	Enterococcus (output)	*Lactococcus* (output)
Enterococcus (target)	**45**	1
Lactococcus (target)	0	**34**

The bold are the number of correctly classified samples.

sets are presented in Tables I–K, respectively.

4. Discussion

Overall, nonlinear models almost systematically outperformed linear models, which is like the results of Feng et al. (2015) and Mu et al. (2018). Although the validation was done on an external set of 80 samples, correct classification rates remain comparable to those obtained by both Feng et al. (2015) and Mu et al. (2018), as well as by Alexandrakis et al. (2008), de Sousa Marques et al. (2013), and Levasseur-Garcia et al. (2017), for which the correct classification rates all ranged from 80% to 100% at the genus and species level.

The combination of Savitzky-Golay smoothing plus EMSC worked

Table G
Performance of discrimination models at the species level by PLS-DA.

Preprocessing	Principal components	Calibration accuracy	External validation accuracy
Raw spectra	10	98.1%	73.8%
Savitzky-Golay smoothing	9	90.6%	71.3%
Savitzky-Golay smoothing + MSC	7	86.9%	70.0%
Savitzky-Golay smoothing + EMSC	7	86.9%	70.0%
Savitzky-Golay 1st derivative	3	78.8%	68.8%
Savitzky-Golay 2nd derivative	3	78.1%	70.0%
Savitzky-Golay 3rd derivative	3	75.6%	72.5%

Table H
Performance of discrimination models at the species level by ANN.

Preprocessing	Network architecture	Iterations	Calibration accuracy	Internal validation accuracy	External validation accuracy
Raw spectra	3086–29-4	50	95.3%	90.6%	86.3%
Savitzky-Golay smoothing	**3086–22-4**	**50**	**92.2%**	**87.5%**	**86.3%**
Savitzky-Golay smoothing + MSC	3086–27-4	50	100%	90.6%	86.3%
Savitzky-Golay smoothing + EMSC	3086–22-4	50	100%	90.6%	86.3%
Savitzky-Golay 1st derivative	3026-7-4	50	99.2%	84.4%	72.5%
Savitzky-Golay 2nd derivative	3443-24-4	15	98.4%	84.4%	71.3%
Savitzky-Golay 3rd derivative	2866-6-4	50	93.0%	81.3%	71.3%

The bold represents the model with the highest classification results.

Table I
Confusion matrix of selected model in calibration.

	E. durans (output)	*E. faecalis* (output)	*E. faecium* (output)	*L. lactis* (output)
E. durans (target)	**10**	3	1	0
E. faecalis (target)	3	**18**	1	0
E. faecium (target)	0	2	**22**	0
L. lactis (target)	0	0	0	**68**

The bold are the number of correctly classified samples.

Table J
Confusion matrix of selected model in internal validation.

	E. durans (output)	*E. faecalis* (output)	*E. faecium* (output)	*L. lactis* (output)
E. durans (target)	**4**	0	0	0
E. faecalis (target)	1	**2**	1	0
E. faecium (target)	0	1	**4**	1
L. lactis (target)	0	0	0	**18**

The bold are the number of correctly classified samples.

Table K
Confusion matrix of selected model in external validation.

	E. durans (output)	*E. faecalis* (output)	*E. faecium* (output)	*L. lactis* (output)
E. durans (target)	**10**	2	0	0
E. faecalis (target)	2	**11**	3	0
E. faecium (target)	0	2	**14**	2
L. lactis (target)	0	0	0	**34**

The bold are the number of correctly classified samples.

best for discrimination at the genus level whereas the use of Savitzky-Golay alone provided the optimal pretreatment at the species level. This result may be due to differences at the genus level being highlighted when eliminating environmental effects with EMSC, whereas species-related information might have been removed when subtracting the agar spectrum.

In both cases, this information is mainly attributed to differences in polysaccharides, even though several other cell-wall constituents such as proteins, phospholipids, and nucleic acid are responsible for differences at various levels of the bacterial phyla (Rodriguez-Saona et al., 2001).

On the PCA of raw spectra, highlighting a spectral region characteristic of inter-species differences is difficult for PC1 loadings (not shown). However, separation on PC2 may be attributed to the absorbance around 1160 and 1260 nm, which may reflect the different polysaccharide composition of Enterococcus and Lactococcus species (Amiel et al., 2000; Dziuba et al., 2007; Dziuba and Nalepa, 2012; Osborne, 2006; Savić et al., 2008; van Heijenoort, 2001).

For the PCA realized on third-derivative spectra, the highest PC1 loadings correspond to several peaks, for which the differences were accentuated by preprocessing (Fig. D.1). The highest peak is around 1315 nm, in the region of the second combination of C−H stretching and deformation; this peak may again be attributed to cell-wall polysaccharides (Levasseur-Garcia et al., 2017). Many other high loadings may be associated with polysaccharides; for example, the peaks around 2510, 2560, and 2620 nm may also be assigned to the C−H bonds of aromatic rings; the peak at 1840 nm might correspond to the combination of O−H stretching and C−O bonds of cellulose; those at 1895 nm and 2025 nm were tentatively attributed to carbonyl groups of polysaccharides (Workman Jr. and Weyer, 2012). Apart from these assignments, the peak at 1230 nm has been associated with the second overtone of C−H stretching from CH2 groups, which could originate from the aliphatic chains of lipid compounds in the cell wall (Dubois et al., 2005). Another peak around 1410 nm was previously attributed in the work of Kammies et al. (2016) to O−H stretching of teichoic acid, which is a compound characteristic of the cell wall of Gram-positive bacteria. Finally, peaks around 1665, 1705, and 1745 nm have been assigned to the first overtone of C−H stretching and may be linked to the production of exopolysaccharides by lactic acid bacteria during the fermentation process (Macedo et al., 2002). PC3 allowed effective discrimination of Enterococcus and Lactococcus species by using a peak around 1380 nm with very high loadings. This peak corresponds to stretching or deformation of C−H bonds, which may again be associated with cell-wall polysaccharides (Levasseur-Garcia et al., 2017). Significant loadings were found at wavelengths common to those of PC1, such as the peak at 1410 nm, which was attributed to teichoic acid, along with various peaks associated with polysaccharides around 1840, 2560, and 2620 nm, and with exopolysaccharide production around 1665, 1705, and 1745 nm (Fig. D.2).

Both capsular polysaccharides and rhamnose-containing cell-wall

polysaccharides may have contributed to these spectral differences. *L. lactis* contains rhamnose-containing cell-wall polysaccharides with a specific structure based on hexasaccharides or pentasaccharides linked by phosphodiester bonds (Mistou et al., 2016). In addition, *Enterococci* have a genus-specific cellular polysaccharide called Enterococcal polysaccharide antigen and whose locus differs in organization and content between species (Palmer et al., 2012). *E. faecium* and *E. faecalis* share a common capsular polysaccharide, which may explain the constant misclassification between the two species in the confusion matrices (Huebner et al., 1999). Those assignments were corroborated by findings from Amiel et al. (2000), Dziuba et al. (2007, 2012), Savić et al. (2008) in the mid-infrared region, and by Levasseur-Garcia et al. (2017) in the NIR region. However, in other works, critical NIR bands for bacterial discrimination were assigned to lipids (Alexandrakis et al., 2008; Feng et al., 2015; Mu et al., 2018).

These initial results could be improved upon by adding more samples in the database to increase the robustness of the model, which is especially important for ANNs to avoid overfitting the data. Expanding the number of genus and species in the database would allow us to further evaluate the capacity of NIRS to identify lactic acid bacteria at the genus and species level. Other possibilities to improve classification rates include selecting key wavelengths associated with differences in genus and species and building multilayered models to assess first the genus and then the species. Developing a panel for an inclusivity and exclusivity study would help estimating the sensitivity and specificity of the method. Parallel analysis of the panel strains with a gold standard diagnostic tool would also help assessing the potential of this method as an alternative to more conventional techniques.

5. Conclusion

This study evaluates the capacity of NIR spectroscopy to assess the genus and species of lactic acid bacteria grown on agar. After acquiring 240 absorbance spectra, PLS-DA and ANN models were built by using 80 spectra as an external validation set. The ANN models provide good results both at the genus and the species level, with 98.8% correct classification as *Enterococcus* and *Lactococcus*, and 86.3% correct classification as *Enterococcus* spp. and *Lactococcus lactis*. These results indicate that, once the developed models are implemented in routine acquisitions, NIRS can provide a fast, easy and nondestructive method to analyze bacteria directly on agar medium, avoiding the stages of resuspension, centrifugation and drying of supernatant on a crystal plate.

Funding

This work was supported by the Languedoc Roussillon Midi Pyrenees Region under Grant number 15065249.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

References

Alexandrakis, D., Downey, G., Scannell, A.G., 2008. Detection and identification of bacteria in an isolated system with near-infrared spectroscopy and multivariate analysis. J. Agric. Food Chem. 56 (10), 3431–3437. https://doi.org/10.1021/jf073407x.

Amiel, C., Mariey, L., Curk-Daubié, M.C., Pichon, P., Travert, J., 2000. Potentiality of Fourier transform infrared spectroscopy (FTIR) for discrimination and identification of dairy lactic acid bacteria. Lait 80 (4), 445–459. Retrieved from. https://doi.org/10.1051/lait:2000137.

Bachmann, H.P., McNulty, D.A., McSweeney, P.L.H., Rüegg, M., 1996. Experimental designs for studying the influence of the raw milk flora on cheese characteristics: a review. Int. J. Dairy Tech. 49 (2), 53–56. https://doi.org/10.1111/j.1471-0307.1996.tb02489.x.

Badis, A., Guetarni, D., Moussa Boudjema, B., Henni, D.E., Kihal, M., 2004a. Identification and technological properties of lactic acid bacteria isolated from raw goat milk of four Algerian races. Food Microbiol. 21 (5), 579–588. https://doi.org/10.1016/j.fm.2003.11.006.

Badis, A., Guetarni, D., Moussa-Boudjemâa, B., Henni, D.E., Tornadijo, M.E., Kihal, M., 2004b. Identification of cultivable lactic acid bacteria isolated from Algerian raw goat's milk and evaluation of their technological properties. Food Microbiol. 21 (3), 343–349. https://doi.org/10.1016/s0740-0020(03)00072-8.

Barreiro, J.R., Braga, P.A., Ferreira, C.R., Kostrzewa, M., Maier, T., Wegemann, B., ... dos Santos, M.V., 2012. Nonculture-based identification of bacteria in milk by protein fingerprinting. Proteomics 12 (17), 2739–2745. https://doi.org/10.1002/pmic.201200053.

Callon, C., Duthoit, F., Delbes, C., Ferrand, M., Le Frileux, Y., De Cremoux, R., Montel, M.C., 2007. Stability of microbial communities in goat milk during a lactation year: molecular approaches. Syst. Appl. Microbiol. 30 (7), 547–560. https://doi.org/10.1016/j.syapm.2007.05.004.

Cámara-Martos, F., Zurera-Cosano, G., Moreno-Rojas, R., García-Gimeno, R.M., Pérez-Rodríguez, F., 2011. Identification and quantification of lactic acid bacteria in a water-based matrix with near-infrared spectroscopy and multivariate regression modeling. Food Anal. Methods 5 (1), 19–28.

Centeno, J.A., Menendez, S., Rodriguez-Otero, J.L., 1996. Main microbial flora present as natural starters in Cebreiro raw cow's-milk cheese (Northwest Spain). Int. J. Food Microbiol. 33 (2–3), 307–313. https://doi.org/10.1016/0168-1605(96)01165-8.

Cheriguene, A., Chougrani, F., Bekada, A.M.A., El Soda, M., Bensoltane, A., 2007. Enumeration and identification of lactic microflora in Algerian goats' milk. Afr. J. Biotechnol. 6 (15). https://doi.org/10.5897/AJB2007.000-2275.

Davis, C., 2014. Enumeration of probiotic strains: review of culture-dependent and alternative techniques to quantify viable bacteria. J. Microbiol. Methods 103, 9–17. https://doi.org/10.1016/j.mimet.2014.04.012.

de Sousa Marques, A., Nicacio, J.T., Cidral, T.A., de Melo, M.C., de Lima, K.M., 2013. The use of near infrared spectroscopy and multivariate techniques to differentiate Escherichia coli and Salmonella Enteritidis inoculated into pulp juice. J. Microbiol. Methods 93 (2), 90–94. https://doi.org/10.1016/j.mimet.2013.02.003.

Demarigny, Y., Sabatier, C., Laurent, N., Prestoz, S., Rigobello, V., Blachier, M.J., 2006. Microbial diversity in natural whey starters used to make traditional Rocamadour goat cheese and possible relationships with its bitterness. Ital. J. Food Sci. 18, 261–276.

Dubois, J., Neil Lewis, E., Fry, F.S.J., Calvey, E.M., 2005. Bacterial identification by near-infrared chemical imaging of food-specific cards. Food Microbiol. 22, 577–583.

Dufour, É., 2009. Chapter 1 - principles of infrared spectroscopy A2 - Sun, Da-Wen. In: Infrared Spectroscopy for Food Quality Analysis and Control. Academic Press, San Diego, pp. 1–27.

Dziuba, B., Nalepa, B., 2012. Identification of lactic acid bacteria and propionic acid bacteria using FTIR spectroscopy and artificial neural networks. Food Technol. Biotechnol. 50 (4), 399–405.

Dziuba, B., Babuchowski, A., Nałęcz, D., Niklewicz, M., 2007. Identification of lactic acid bacteria using FTIR spectroscopy and cluster analysis. Int. Dairy J. 17 (3), 183–189. https://doi.org/10.1016/j.idairyj.2006.02.013.

Edalatian Dovom, M.R., Habibi Najafi, M.B., Ali Mortazavi, S., Alegría, Á., Nassiri, M., Bassami, M.R., Mayo, B., 2012. Microbial diversity of the traditional Iranian cheeses Lighvan & Koozeh, as revealed by polyphasic culturing and culture-independent approaches. Dairy Sci. Technol. 92 (1), 75–90. https://doi.org/10.1007/s13594-011-0045-2.

Feng, Y.-Z., Downey, G., Sun, D.-W., Walsh, D., Xu, J.-L., 2015. 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. J. Food Eng. 149, 87–96. https://doi.org/10.1016/j.jfoodeng.2014.09.016.

Grassi, S., Alamprese, C., Bono, V., Picozzi, C., Foschino, R., Casiraghi, E., 2013. Monitoring of lactic acid fermentation process using Fourier transform near infrared spectroscopy. J. Near Infrared Spectrosc. 21 (5), 417–425. https://doi.org/10.1255/jnirs.1058.

Guessas, B., Kihal, M., 2004. Characterization of lactic acid bacteria isolated from Algerian arid zone raw goats milk. Afr. J. Biotechnol. 3 (6), 339–342. https://doi.org/10.5897/AJB2004.000-2062.

Helland, I., Næs, T., Isaksson, T., 1995. Related versions of multiplicative scatter correction method for preprocessing spectroscopic data. Chemom. Intell. Lab. Syst. 29, 233–241. https://doi.org/10.1016/0169-7439(95)80098-T.

Huebner, J., Wang, Y., Krueger, W.A., Madoff, L.C., Martirosian, G., Boisot, S., ... Pier, G.B., 1999. Isolation and chemical characterization of a capsular polysaccharide antigen shared by clinical isolates of Enterococcus faecalis and vancomycin-resistant Enterococcus faecium. Infect. Immun. 67 (3), 1213–1219.

Kammies, T.L., Manley, M., Gouws, P.A., Williams, P.J., 2016. Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. Appl. Microbiol. Biotechnol. 100 (21), 9305–9320. https://doi.org/10.1007/s00253-016-7801-4.

Levasseur-Garcia, C., Couderc, C., Tormo, H., 2017. Discrimination of lactic acid bacteria Enterococcus and Lactococcus by infrared spectroscopy and multivariate techniques. J. Near Infrared Spectrosc. 25 (4), 231–241. https://doi.org/10.1177/0967033517719383.

Liu, Y., Lu, C., Meng, Q., Lu, J., Fu, Y., Liu, B., ... Teng, L., 2016. Near infrared spectroscopy coupled with radial basis function neural network for at-line monitoring of Lactococcus lactis subsp. fermentation. Saudi J. Biol. Sci. 23 (1), S106–S112. https://doi.org/10.1016/j.sjbs.2015.06.023.

Macedo, M.G., Laporte, M.F., Lacroix, C., 2002. Quantification of exopolysaccharide, lactic acid, and lactose concentrations in culture broth by near-infrared spectroscopy. J. Agric. Food Chem. 50 (7), 1774–1779.

Martens, H., Stark, E., 1991. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. J. Pharm. Biomed. Anal. 9 (8), 625–635. https://doi.org/10.1016/0731-7085(91)

80188-F.
Martens, H., Karstang, T., Næs, T., 1987. Improved selectivity in spectroscopy by multivariate calibration. J. Chemom. 1 (4), 201–219. https://doi.org/10.1002/cem.1180010403.
Mas, M., Tabla, R., Moriche, J., Roa, I., Gonzalez, J., Rebollo, J.E., Cáceres, P., 2002. Ibores goat's milk cheese: microbiological and physicochemical changes throughout ripening. Lait 82 (5), 579–587. https://doi.org/10.1051/lait:2002034.
Mistou, M.Y., Sutcliffe, I.C., van Sorge, N.M., 2016. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in gram-positive bacteria. FEMS Microbiol. Rev. 40 (4), 464–479. https://doi.org/10.1093/femsre/fuw006.
Mu, K.-X., Feng, Y.-Z., Chen, W., Yu, W., 2018. Near infrared spectroscopy for classification of bacterial pathogen strains based on spectral transforms and machine learning. Chemom. Intell. Lab. Syst. 179, 46–53. https://doi.org/10.1016/j.chemolab.2018.06.003.
Onda, T., Tsuji, M., Yanagida, F., Shinohara, T., Ogino, S., 2001. Determination of fermentation type of lactic acid Bacteria by near infrared spectroscopy. Food Preserv. Sci. 27 (4), 189–195. https://doi.org/10.5891/jafps.27.189.
Osborne, B.G., 2006. Near-infrared spectroscopy in food analysis. Encycl. Anal. Chem. https://doi.org/10.1002/9780470027318.a1018.
Palmer, K.L., Godfrey, P., Griggs, A., Kos, V.N., Zucker, J., Desjardins, C., ... Gilmore, M.S., 2012. Comparative genomics of enterococci: variation in Enterococcus faecalis, clade structure in E. faecium, and defining characteristics of E. gallinarum and E. casseliflavus. mBio 3 (1). https://doi.org/10.1128/mBio.00318-11. e00318-00311.
Rodriguez-Saona, L.E., Khambaty, F.M., Fry, F.S., Calvey, E.M., 2001. Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy. J. Agric. Food Chem. 49 (2), 574–579.
Sandrin, T.R., Goldstein, J.E., Schumaker, S., 2013. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom. Rev. 32 (3), 188–217. https://doi.org/10.1002/mas.21359.
Savić, D., Joković, N., Topisirović, L.J.D.S., Technology, 2008. Multivariate statistical methods for discrimination of lactobacilli based on their FTIR spectra. Dairy Sci. Technol. 88 (3), 273–290. https://doi.org/10.1051/dst:2008003.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36 (8), 1627–1639. https://doi.org/10.1021/ac60214a047.
Soomro, A.H., Masud, T., 2007. Protein pattern and plasmid profile of lactic acid bacteria isolated from dahi, a traditional fermented milk product of Pakistan. Food Technol. Biotechnol. 45 (4), 447–453. Retrieved from. https://hrcak.srce.hr/24037.
Svendsen, C., Cieplak, T., van den Berg, F.W.J., 2017. Exploring process dynamics by near infrared spectroscopy in lactic fermentations. J. Near Infrared Spectrosc. 24 (5), 443–451. https://doi.org/10.1255/jnirs.1244.
Tormo, H., Ali Haimoud Lekhal, D., Roques, C., 2015. Phenotypic and genotypic characterization of lactic acid bacteria isolated from raw goat milk and effect of farming practices on the dominant species of lactic acid bacteria. Int. J. Food Microbiol. 210, 9–15. https://doi.org/10.1016/j.ijfoodmicro.2015.02.002.
van Heijenoort, J., 2001. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. Glycobiology 11 (3), 25r–36r.
Visa, S., Ramsay, B., Ralescu, A., Knaap, E., 2011. Confusion matrix-based feature selection. In: Paper presented at the Proceedings of the 22nd Midwest Artificial Intelligence and Cognitive Science Conference 2011, Cincinnati, Ohio, USA.
Workman Jr., J., Weyer, L., 2012. Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, 2nd edition. CRC Press, Boca Raton.
Yang, J., Cao, Y., Cai, Y., Terada, F., 2010. Natural populations of lactic acid bacteria isolated from vegetable residues and silage fermentation. J. Dairy Sci. 93 (7), 3136–3145. https://doi.org/10.3168/jds.2009-2898.