DOI: 10.1177/0967033517719383

Discrimination of lactic acid bacteria *Enterococcus* and *Lactococcus* by infrared spectroscopy and multivariate techniques

Cecile Levasseur-Garcia[1], Christel Couderc[2] and Helene Tormo[2]

Abstract

Raw milk is often described as a major source of lactic acid bacteria for indigenous lactic starter. These indigenous starters contribute to the sensorial quality of cheese. Raw milk, rich in *Lactoccocus lactis* may therefore be very interesting for the cheese making. Currently, the most commonly used methods to differentiate lactic acid bacteria, and particularly the closely related phenotypes *Lactococcus* and *Enterococcus*, are based on DNA sequencing, but the cost and time required for these analytical methods hinder their use for rapid screening of raw material. The present study therefore proposes a simple alternative method to identify and discriminate against *Lactococcus* and *Enterococcus*, at the genus, but also at the species level, that is based on collecting near infrared spectra directly from bacterial colonies in Petri dishes. The infrared spectra of 280 strains of *Lactococcus* and *Enterococcus* cultured on solid media were collected by using a spectrometer with a wavelength range of 908 to 1684 nm and a remote probe. The best Classification And Regression Trees models for genus and species discrimination gave an excellent classification rate of 87% on an external validation set (30 strains). Loading line plots, with prominent bands at 900–960 and 1270–1390 nm, confirmed that the source of variation was due to changes in the polysaccharides.

Keywords

Enterococcus, *Lactococcus*, lactic acid bacteria, discrimination, genera, species, infrared spectroscopy

Introduction

Lactoccocus lactis is a bacterium present in raw milk and is often used as a lactic acid starter to make cheese.[1,5] Due to their acidifying capacity, lactic acid bacteria play a key role in the acidification of the curd essential for cheese making, but they also contribute to cheese aroma and texture as they possess endo and exopeptidases which are involved in the production of sapid molecules; they generate precursors of aromatic compounds.[6,7] So, to elaborate indigenous starter from the milk, it is necessary to differentiate between *Lactococcus* and its closely related phenotype *Enterococcus*. Currently, the most commonly used methods for bacterial typing are based on phenotyping, molecular biology or analysis of ribosomal proteins (MALDI-TOF MS).[8,11] The cost and time required for these analytical methods limit their use for rapid analysis of raw material. Thus, an alternative method is needed. Spectroscopy which is based on measuring absorbance of infrared radiation is a fast and nondestructive tool that is commonly used in the food industry. When combined with clustering[12,13] or artificial neural networks (ANNs), mid-infrared spectroscopy can be used to identify and classify bacteria.[14,15] Microorganisms such as lactic acid bacteria have been studied with Fourier transform mid-infrared spectroscopy (FT-mid-IR). For example, Amiel et al.[16] classified *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Weissella*, and *Streptococcus* isolates at the species level, whereas Savic et al.[17] focused on *Lactobacillus* strains. Samelis et al.[18] discriminated between *Lactobacillus casei/paracasei*, *Lactobacillus plantarum*, *Streptococcus thermophilus*, *Enterococcus faecium*, and *L. lactis* by using cluster analysis. In another study, different strains of *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Pediococcus*, and *Streptococcus* were discriminated by using FT-mid-IR and cluster analysis[19] and ANNs.[20] Two other studies[21,22] investigated several strains of *Enterococcus* by both FT-mid-IR and cluster analysis, which required rather long sample preparation. In all

[1]Laboratoire de Chimie Agro-industrielle (LCA), Université de Toulouse, INRA, INPT, INP-Purpan, Toulouse, France
[2]Equipe Agroalimentaire et Nutrition, Université de Toulouse, INPT, INP-Purpan, Toulouse, France

Corresponding author:
Cecile Levasseur-Garcia, INP-Ecole d'Ingénieurs de Purpan, 75 voie du T.O.E.C., F-31076 Toulouse, Cedex 03, France.
Email: cecile.levasseur@purpan.fr

those studies, bacteria cells first had to be put in suspension and then, before drying, placed on an optical plate or other infrared-transparent material such as ZnSe. However, to the best of our knowledge, near infrared (NIR) spectroscopy has yet to be used to discriminate between *Lactococcus* and *Enterococcus*. Indeed, the novelty of the study is that this infrared tool requires no preparation and enables nondestructive analysis of the samples.

The present study thus focuses on a database of NIR spectra rapidly collected with a remote probe directly from Petri dishes, without preparing the samples. Furthermore, the strains did not come from a collection, but are representative of those found in the natural ecosystem of the farms involved in the study. In this paper, we evaluated the use of Classification And Regression Trees (CART) to build a model to distinguish between genera and between species of bacteria.

After preprocessing the spectra, 280 strains were discriminated with CART.

Materials and methods

Milk samples

In a single week, 12 raw-milk samples were collected from five breeds of goats in different areas of France. After the milking, the raw samples were collected in sterilized 500 mL bottles, cooled immediately to 4 °C, and stored at −25°C until analysis.

Bacterial culture

After inoculation on the surface of serially diluted milk in sterile buffered peptone water (Biomérieux, France) and incubation for 72 h at 20°C, approximately 10 acidifying bacterial colonies (colonies with yellow halo) were isolated from the suitable dilution (between 30 and 100 colonies in the Petri dish) and incubated in Elliker broth overnight at 30°C. A loopful of the culture was taken from the Petri dish and transferred to another Petri dish containing a modified Elliker agar. Each Petri dish was maintained in the dark for 72 h at 30 °C before collecting any infrared spectra.

Identification of bacteria species by partial 16S rDNA sequencing

A loopful of the culture was taken from the plates and suspended in 5 mL of Elliker broth, then genomic DNA was isolated with NucleoSpin® Tissue (Macherey-Nagel, Hoerd, France). The polymerase chain reaction (PCR) amplification was done in a thermal cycler Bio-Rad MyCycler (Bio-Rad Laboratories, Hercules, CA).

The 16S rDNA gene (V1–V4) was amplified by PCR using primers E8F and E807R.[23]

The reaction mixture (50 µL) contained 5 µM primer E8F (Invitrogen, Carlsbad, CA), 5 µM Primer E807R, 5 mM for each deoxynucleotide (dNTPS), 1.25 U/mL TaqDNA polymerase, a 10x Thermopol buffer (Brolab, NewEngland, UK), 50 ng of bacterial DNA, and was completed with ultrapure water.

DNA was sequenced by the Sanger method (GATC Biotech, Constance, Germany). Sequences were identified by using GenBank from the National Center for Biotechnology Information, USA.

Instrumentation

Spectra were acquired by using a NIR356 1.7 GetSpec Spectrometer (getAMO, Dresden, Germany), equipped with a tungsten-halogen light source Sentrolight HalS and a reflecting probe with sapphire windows mounted at 45° (Figure 1).

The NIR spectra were obtained over the range 908–1684 nm (a bandwidth of 777 nm) with a spectral resolution of 1 nm. Each spectrum was the average of 100 scans. Absorbance spectra of bacterial samples were obtained on top of the background spectrum of Spectralon® (Labsphere Inc., Bures Sur Yvette, France). The reflecting probe was positioned 3 cm above the Petri dish. Spectra were recorded in random order. The spectra were recorded as absorbance. The laboratory temperature was maintained at 22 ± 2 °C during the NIR absorbance measurements. Three spectra were recorded for each strain and were used separately in the model. A total of 280 spectra (134 isolates of *Enterococcus* and 146 of *Lactococcus*) were collected. The spectra were collected by using CDI Spec32 data acquisition software (v1737, Control Development Inc., South Bend, IN) on a Windows XP laptop computer.

Spectral classification

Prior to spectral analysis, NIR spectra were pre-treated with multiplicative scattering correction (MSC). Data preprocessing was done by using The Unscrambler® Multivariate Data Analysis (v. X; CAMO A/S, Oslo, Norway).

Prior to developing calibration, a principal components analysis (PCA) was performed to get an overview

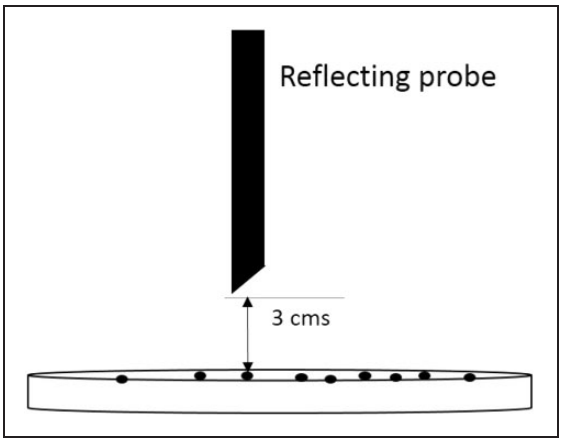

Figure 1. Positioning of the infrared probe above the Petri dish.

Table 1. Identification of bacteria in 280 samples: 250 samples in the training set and 30 samples in the external validation set.

	Training (*n* 250)		External validation (*n* 30)		
	No. of samples	Percentage	No. of samples	Percentage	Total
Enterococcus	120	48	14	47	134
Lactococcus	130	52	16	53	146
Enterococcus faecalis	91	36	9	30	100
Enterococcus faecium	19	8	2	7	21
Enterococcus hirae	2	1	1	3	3
Enterococcus italicus	6	2	1	3	7
Enterococcus munditi	2	1	1	3	3
Lactococcus garvieae	5	2	1	3	6
Lactococcus lactis	125	50	15	50	140

of the data, to highlight sample clustering, to detect outlier samples, and to reduce the dimensionality of spectral data.[24] Finally, we applied CART to discriminate between lactic acid bacteria genera (*Enterococcus* and *Lactococcus*), and species: *Enterococcus faecalis*, *E. faecium*, other *Enterococcus* (*Enterococcus hirae*, *Enterococcus italicus* and *Enterococcus munditi*), *Lactococcus garvieae* and *L. lactis*. CART is a non-parametric statistical technique which is able to solve classification problems.[25,26] The model was a decision tree of binary recursive partitioning, with node-splitting rules. The split selection criterion was Gini Index, as defined by Breiman.[25 27] The quality of the model was evaluated by calculating prediction accuracy and misclassification rates from confusion matrices.[28] The *true positive* rate (TP) was the proportion of positive cases that were correctly identified; the *false positive* rate (FP) was the proportion of negatives cases that were incorrectly classified as positive; the *true negative* rate (TN) was defined as the proportion of negatives cases that were classified correctly and the *false negative* rate (FN) was the proportion of positives cases that were incorrectly classified as negative. The prediction accuracy and the classification error could be calculated from these rates. The prediction accuracy was obtained as

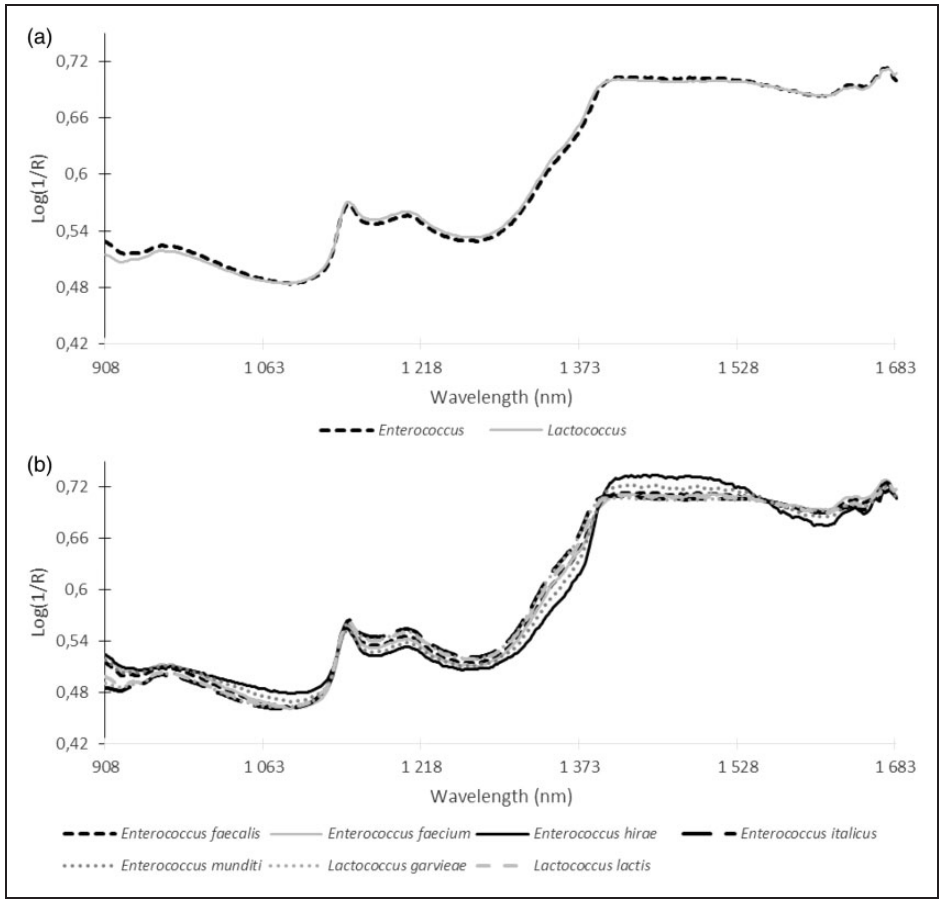

Figure 2. NIR spectra treated by multiplicative scattering correction of bacteria from (a) different genera and (b) different species. NIR: near infrared.

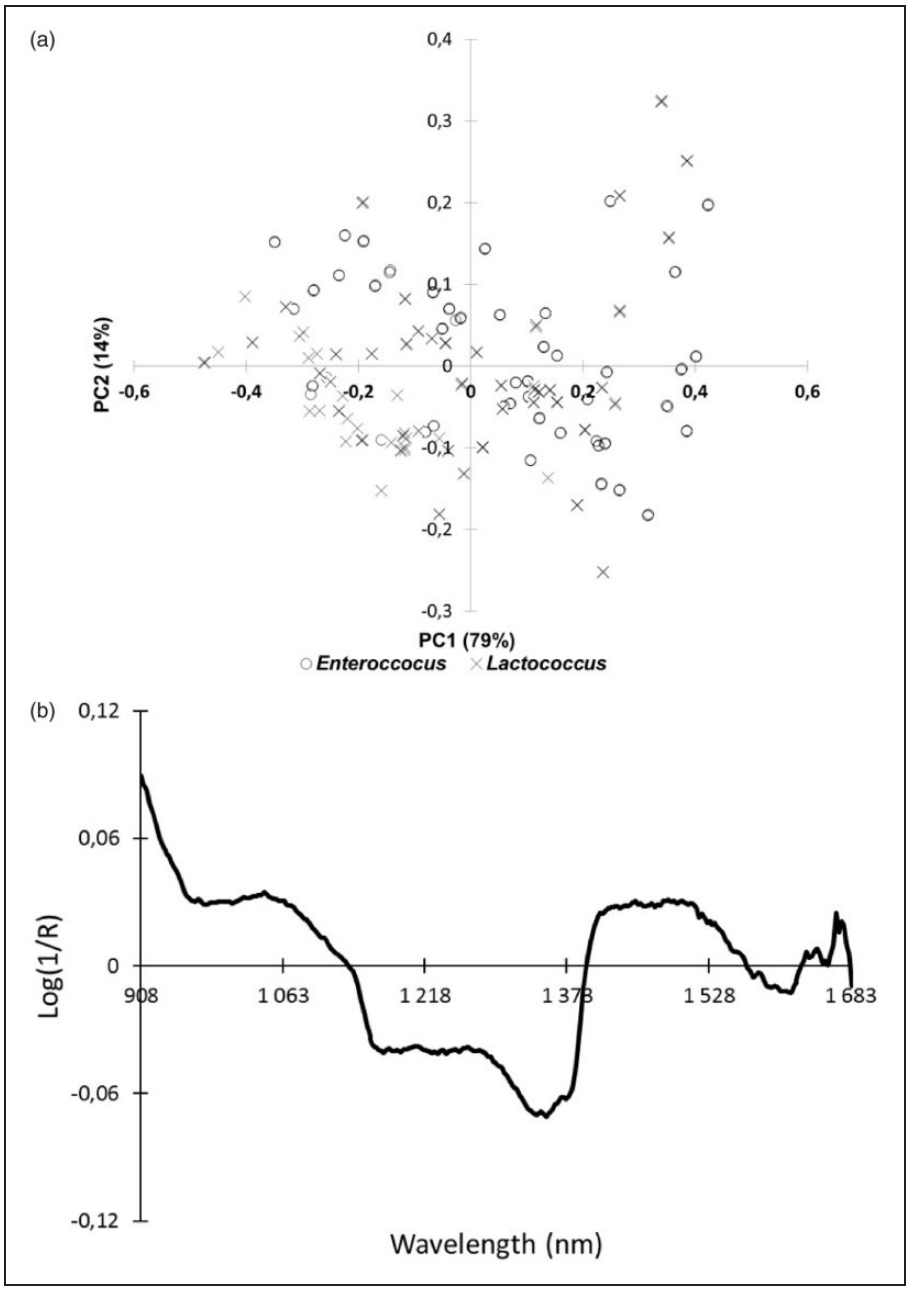

Figure 3. (a) Scatter plot of first two principal components (PC1 and PC2) for all bacteria spectra. (b) Loading vector plot of first principal component corresponding to the bacteria dataset.

follows (equation (1))

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (1)$$

Description of the two sets

For the classification analysis, 30 samples were set for testing (randomly selected) and remained 250 samples for training. The accuracy of the classification models was assessed based on the percent of correct classification. The best model was the one with the highest accuracy. The test set was used to estimate how well the models would perform on new data. In the training set, the bacterial species were identified as *E. faecalis* (36%), *E. faecium* (8%), *E. hirae* (1%), *E. italicus* (2%), *E. munditi* (1%), *L. garvieae* (2%), and *L. lactis* (50%). Table 1 summarizes the two sets.

Results and discussion

NIR spectral features

Figure 2(a) shows the mean NIR spectra of the two bacterial genera (Figure 2(a)) and the seven species (Figure 2(b)).

Slight differences are observed between the different pretreated spectral profiles. However, a detailed analysis of the spectra requires a PCA.

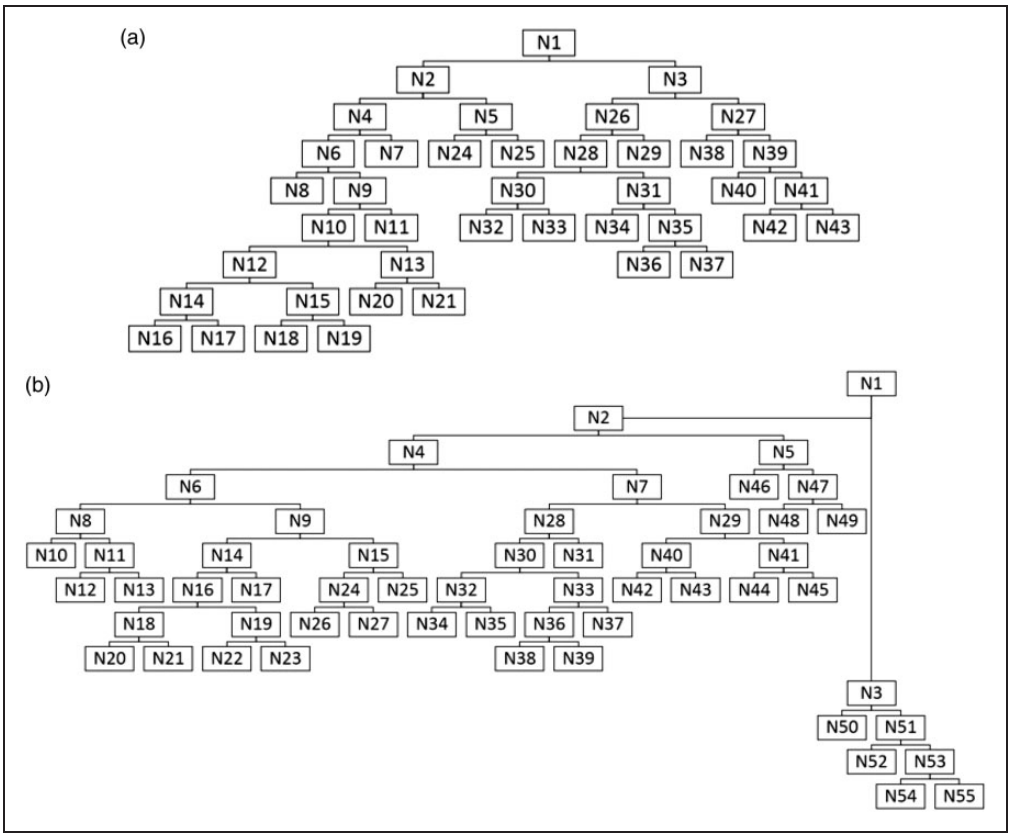

Figure 4. Binary decision tree obtained to predict (a) genera and (b) species from infrared spectra.

Table 2. Intermediate node division conditions for discriminating genera.

		%		Wavelengths (nm)																			
Nodes	Total no.	*Enterococcus*	*Lactococcus*	915	930	931	951	1047	1063	1131	1152	1219	1386	1388	1391	1399	1554	1562	1563	1576	1589	1613	1629
Node 1	250	48%	52%																				
Node 2	177	60%	40%											x									
Node 3	73	19%	81%											x									
Node 4	143	52%	48%											x		x							
Node 5	34	94%	6%											x		x							
Node 6	125	45%	55%											x		x						x	
Node 7	18	100%	0%											x		x						x	
Node 8	16	0%	100%											x		x		x				x	
Node 9	109	51%	49%											x		x		x				x	
Node 10	90	59%	41%											x		x		x	x			x	
Node 11	19	16%	84%											x		x		x	x			x	
Node 12	68	71%	29%						x					x		x		x	x			x	
Node 13	22	23%	77%						x					x		x		x	x			x	
Node 14	8	13%	88%	x					x					x		x		x	x			x	
Node 15	60	78%	22%	x					x					x		x		x	x			x	
Node 16	7	0%	100%	x					x		x			x		x		x	x			x	
Node 17	1	100%	0%	x					x		x			x		x		x	x			x	
Node 18	41	93%	7%	x					x					x		x	x	x	x			x	
Node 19	19	47%	53%	x					x					x		x	x	x	x			x	
Node 20	5	100%	0%			x			x					x		x		x	x			x	
Node 21	17	0%	100%			x			x					x		x		x	x			x	
Node 22	16	0%	100%				x							x		x		x	x			x	

(continued)

Table 2. Continued

Nodes	Total no.	% *Enterococcus*	% *Lactococcus*	Wavelengths (nm) 915	930	931	951	1047	1063	1131	1152	1219	1386	1388	1391	1399	1554	1562	1563	1576	1589	1613	1629
Node 23	3	100%	0%				x							x		x		x	x			x	
Node 24	2	0%	100%		x									x		x							
Node 25	32	100%	0%		x									x		x							
Node 26	60	8%	92%										x	x									
Node 27	13	69%	31%										x	x									
Node 28	58	5%	95%										x	x									x
Node 29	2	100%	0%										x	x									x
Node 30	3	67%	33%										x	x	x								x
Node 31	55	2%	98%										x	x	x								x
Node 32	1	0%	100%							x			x	x	x								x
Node 33	2	100%	0%							x			x	x	x								x
Node 34	47	0%	100%										x	x	x						x		x
Node 35	8	13%	88%										x	x	x						x		x
Node 36	7	0%	100%									x	x	x	x						x		x
Node 37	1	100%	0%									x	x	x	x						x		x
Node 38	2	0%	100%										x	x						x			
Node 39	11	82%	18%										x	x	x					x			
Node 40	6	100%	0%										x	x	x					x			
Node 41	5	60%	40%					x					x	x	x					x			
Node 42	2	0%	100%					x					x	x	x					x			
Node 43	3	100%	0%					x					x	x	x					x			

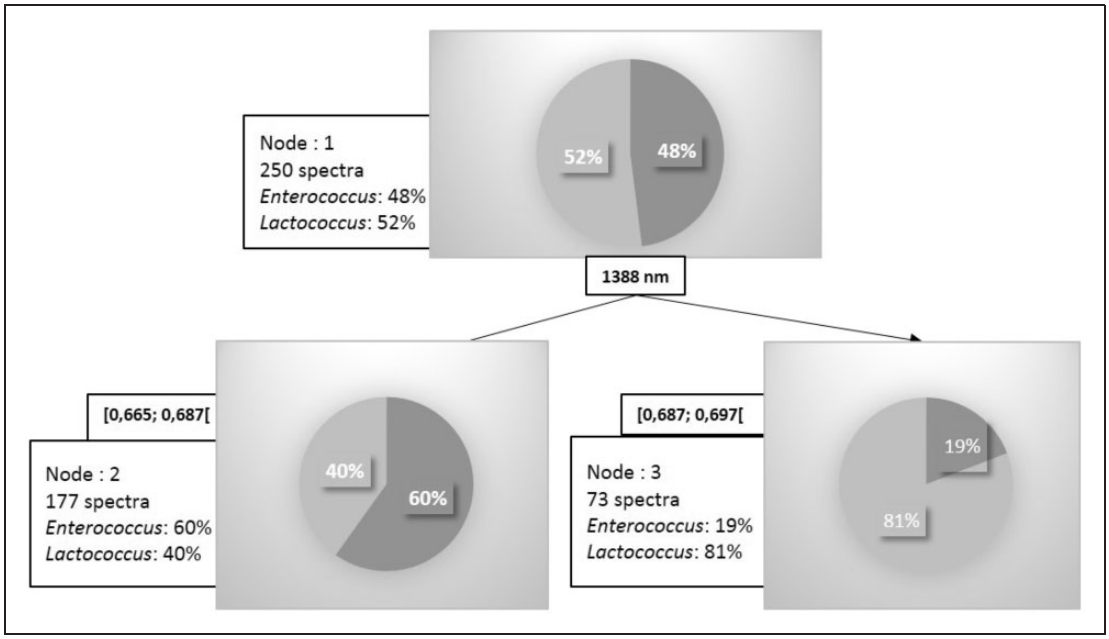

Figure 5. Procedure for classifying bacterial spectra regarding genera.

Principal component analysis

After processing, the entire spectrum was submitted to PCA analysis. The PCA model was built from seven principal components (PC1 to PC7). No distinct clusters appeared in the score plots of all pairwise combinations of PC1 to PC7. Figure 3 shows a bidimensional representation of PC1 and PC2, which together explained 93% of the variance in the data. All samples are labeled according to their genus.

In Figure 3(a), no clear separation appears between the two genera, and no obvious outliers are detected. The loadings plot (Figure 3(b)) allows the variables most closely related to the first principal

component to be identified. In this figure, variables in the ranges 900–960 and 1270–1390 nm are more relevant for characterizing the genera and species of bacteria. Bacterial cell walls are known to have a high taxonomic value. Indeed, the peptidoglycans of the bacterial cell wall are made up of linear glycan chains interlinked by short peptides. Those chemical species vary by both microorganism and culture substrate.[29,30] The specific infrared bands identified by the loading plot (900–960 and 1270–1390 nm) might be relevant for indicating differences in peptidoglycans' patterns or in the amount of given hydrocarbons.[31] Those bands were tentatively assigned to specific chemical functional groups; the range 1270–1390 nm could be related to CH stretch and CH deformation or to CH second combination region; whereas the range 900–960 nm could be related to third overtone methyl peak.[32,33]

Discrimination of Enterococcus and Lactococcus

We applied CART with 777 variables to 250 learning samples for bacterial classification. Figure 4(a) shows the decision tree built to classify bacteria as *Enterococcus* or *Lactococcus*.

In Table 2, intermediate node-division conditions are given next to each node, corresponding to the nodes in Figure 4(a). Figure 5 shows the example of node 1-division condition. Node 1 is called the decisive node and nodes 2 and 3 are final nodes, called leaves. For each node, a diagram gives an overview of the distribution of the decision variable (i.e. the number of bacterial spectra belonging to *Enterococcus* or *Lactococcus* genera). The splitting attribute, the absorbance at 1388 nm in this case, is also indicated, as well as the corresponding absorbance ranges (0.665–0.687 and 0.687–0.697). Table 2 gives intermediate node-division conditions next to each node (wavelengths) and the partitioning results (percent of *Enterococcus* or *Lactococcus*).

The following wavelengths were used as node-splitting rules: 915, 930, 931, 951, 1047, 1063, 1131, 1152, 1219, 1386, 1388, 1391, 1399, 1554, 1562, 1563, 1576, 1589, 1613, and 1629 nm (Table 2).

Based on the results given in Table 2, several recurrent wavelengths are seen to discriminate between genera. These are 1386, 1388, and 1399 nm, corresponding to CH stretch and CH deformation, and 1613 nm, corresponding to the CONHR H-bonded band.[31]

The results of applying this model are presented in Table 3.

In training, 93% and 98% of the spectra corresponding to *Enterococcus* and *Lactococcus* were correctly classified, respectively. Applying this detection method to the test set containing 30 bacteria spectra resulted in 93% and 81% proper classification for *Enterococcus* and *Lactococcus*, respectively.

Table 3. Confusion matrix for genera.

To				
From	*Enterococcus*	*Lactococcus*	Total	Well classified samples
Training				
Enterococcus	111	9	120	93%
Lactococcus	3	127	130	98%
Total	114	136	250	95%
Validation				
Enterococcus	13	1	14	93%
Lactococcus	3	13	16	81%
Total	16	14	30	87%

Bacterial spectra for discriminating between species

Figure 4(b) shows the decision-tree built to classify the species of bacteria.

The following wavelengths were used for node-splitting rules: 908, 915, 930, 939, 979, 1017, 1033, 1047, 1119, 1131, 1227, 1275, 1381, 1386, 1391, 1415, 1539, 1557, 1558, 1563, 1576, 1595, 1597, 1605, 1615, and 1629 nm, as detailed in Table 4.

The following recurrent wavelengths can be highlighted: 1227, 1386, 1415, 1539, and 1597 nm. Chemical assignments may be made corresponding to ROH and CH_2 groups (1227 nm for second overtone of CH stretch, 1386 nm for CH stretch and CH deformation, 1415 nm for two CH stretches + CH deformation, 1539 nm for first overtone of CH or OH stretch).[31,33,34]

The wavelengths 915, 930, 1047, 1131, 1386, 1391, 1563, 1576, and 1629 nm discriminated not only between genera but also between species. These results support those of the loading-vector plot of the first principal component corresponding to the bacteria dataset, which highlights the two spectral ranges 900–960 and 1270–1390 nm.

The results of applying this model are presented in Table 5.

In training, 95% of the spectra were correctly classified: 100% for *E. faecalis*, *E. faecium*, *E. hirae*, *E. munditi*, and *L. garvieae*, 93% for *L. lactis*, and 50% for *E. italicus*. Applying this method to classify the bacteria species in the test set containing 30 bacteria spectra gave 87% correct classification: 100 % for *E. faecium*, *E. italicus*, and *E. munditi*, 93% for *L. lactis*, 89% for *E. faecalis*, and 0% for *E. hirae* and *L. garvieae*. The performance of these models approach that of the Udelhoven et al.[15] or Dziuba and Nalepa's neural-network models,[35] which gave 89% to 95% correct classification at the genus level, and surpasses that of the Guibet model,[21] which gave 78% correct classification for *E. faecalis* (7 of 9 samples) and, in the best case, 75% correct classification for *E. faecium*.

Table 4. Intermediate node-division conditions for discrimination between species.

		Outcomes (%)							Wave engths (nm)																												
Node	tota no.	*E faecalis*	*E faecium*	*E hirae*	*E italicus*	*E munditi*	*L garvieae*	*L lactis*	908	915	930	939	979	1017	1033	1047	1119	1131	1227	1275	1381	1386	1391	1415	1539	1557	1558	1563	1576	1595	1597	1605	1615	1629
1	250	36	8	1	2	1	2	50																											
2	214	29	9	0	3	0	2	57																x										
3	36	78	0	6	0	6	0	11																x										
4	189	23	9	0	3	0	3	62																x							x			
5	25	80	8	0	0	0	0	12																x							x			
6	100	34	15	0	0	0	2	49											x					x							x			
7	89	10	2	0	7	0	3	78											x					x							x			
8	21	76	14	0	0	0	0	10											x					x	x						x			
9	79	23	15	0	0	0	3	59											x					x	x						x			
10	16	100	0	0	0	0	0	0	x										x					x	x						x			
11	5	0	60	0	0	0	0	40	x										x					x	x						x			
12	3	0	100	0	0	0	0	0	x										x					x	x						x			
13	2	0	0	0	0	0	0	100	x										x					x	x						x			
14	47	32	26	0	0	0	0	43										x	x					x	x						x			
15	32	9	0	0	0	0	6	84										x	x					x	x						x			
16	41	37	15	0	0	0	0	49										x	x					x	x						x			
17	6	0	100	0	0	0	0	0										x	x					x	x		x				x			
18	9	22	67	0	0	0	0	11										x	x					x	x	x	x				x			
19	32	41	0	0	0	0	0	59										x	x					x	x	x	x				x			
20	3	67	0	0	0	0	0	33							x			x	x					x	x	x	x				x			
21	6	0	100	0	0	0	0	0							x			x	x					x	x	x	x				x			
22	11	0	0	0	0	0	0	100										x	x					x	x	x	x	x			x			
23	21	62	0	0	0	0	0	38										x	x					x	x	x	x	x			x			
24	29	0	0	0	0	0	7	93				x						x	x					x	x						x			
25	3	100	0	0	0	0	0	0				x						x	x					x	x						x			
26	2	0	0	0	0	0	100	0				x	x					x	x					x	x						x			
27	27	0	0	0	0	0	0	100				x	x					x	x					x	x						x			
28	76	4	3	0	4	0	4	86											x			x		x							x			
29	13	46	0	0	23	0	0	31											x			x		x							x			
30	74	4	0	0	4	0	4	88											x			x		x						x	x			
31	2	0	100	0	0	0	0	0											x			x		x						x	x			
32	47	2	0	0	0	0	0	98											x		x	x		x						x	x			
33	27	7	0	0	11	0	11	70											x		x	x		x						x	x			

(cont nued)

Table 4. Continued

Node	total no.	Outcomes (%)							Wave engths (nm)																																	
		E faecalis	*E faecium*	*E hirae*	*E italicus*	*E munditi*	*L garvieae*	*L lactis*	908	915	930	939	979	1017	1033	1047	1119	1131	1227	1275	1381	1386	1391	1415	1539	1557	1558	1563	1576	1595	1597	1605	1615	1629							
34	1	100	0	0	0	0	0	0											x	x	x	x		x					x	x											
35	46	0	0	0	0	0	0	100											x	x	x	x		x					x	x											
36	24	8	0	0	13	0	0	79											x		x	x		x					x	x		x									
37	3	0	0	0	0	0	100	0											x		x	x		x					x	x		x									
38	22	0	0	0	14	0	0	86											x		x	x		x					x	x		x	x								
39	2	100	0	0	0	0	0	0											x		x	x		x					x	x		x	x								
40	7	86	0	0	0	0	0	14											x			x	x	x							x										
41	6	0	0	0	50	0	0	50											x			x	x	x							x										
42	1	0	0	0	0	0	0	100											x			x	x	x					x		x										
43	6	100	0	0	0	0	0	0											x			x	x	x					x		x										
44	3	0	0	0	0	0	0	100								x			x			x	x	x							x										
45	3	0	0	0	100	0	0	0								x			x			x	x	x							x										
46	20	100	0	0	0	0	0	0																x							x	x									
47	5	0	40	0	0	0	0	60																x							x	x									
48	3	0	0	0	0	0	0	100		x														x							x	x									
49	2	0	100	0	0	0	0	0		x														x							x	x									
50	4	0	0	0	0	0	0	100			x													x																	
51	32	88	0	6	0	6	0	0			x													x																	
52	2	0	0	0	0	100	0	0			x			x										x																	
53	30	93	0	7	0	0	0	0			x			x										x																	
54	28	100	0	0	0	0	0	0			x			x			x							x																	
55	2	0	0	100	0	0	0	0			x			x			x							x																	

Table 5. Confusion matrix for species.

From	To							Total	Well classified samples
	E. faecalis	*E. faecium*	*E. hirae*	*E. italicus*	*E. munditi*	*L. garvieae*	*L. lactis*		
Training									
E. faecalis	91	0	0	0	0	0	0	91	100%
E. faecium	0	19	0	0	0	0	0	19	100%
E. hirae	0	0	2	0	0	0	0	2	100%
E. italicus	0	0	0	3	0	0	3	6	50%
E. munditi	0	0	0	0	2	0	0	2	100%
L. garvieae	0	0	0	0	0	5	0	5	100%
L. lactis	9	0	0	0	0	0	116	125	93%
Total	100	19	2	3	2	5	119	250	95%
Independent test									
E. faecalis	8	0	0	0	0	0	1	9	89%
E. faecium	0	2	0	0	0	0	0	2	100%
E. hirae	1	0	0	0	0	0	0	1	0%
E. italicus	0	0	0	1	0	0	0	1	100%
E. munditi	0	0	0	0	1	0	0	1	100%
L. garvieae	0	0	0	0	0	0	1	1	0%
L. lactis	1	0	0	0	0	0	14	15	93%
Total	10	2	0	1	1	0	16	30	87%

The misclassification rate was 13% for the validation set. By classifying a random spectrum using this model, we had 13 out of 100 risk of bad assignment, whereas without using the model, if we randomly attribute a label "specie" to a random spectrum, taking into account the initial repartition, we have a risk of 86% (6/7) of bad assignment.

Thus, based on infrared absorption spectra collected from bacterium cultures, the two CART models provided conclusive classification. Although *Enterococcus* and *Lactococcus* have the same number of strains, this is not the case for the species. *E. faecalis*, *E. faecium*, and *L. lactis* were over-represented, although this reflects the natural distribution. The CART model is thus better suited to classify these species—in fact, the rate of correct classification for under-represented species was inferior (50% for *E. italicus* in training and 0% for *E. hiae* and *L. garvieae* in tests). To improve the model, new samples of these species could be added to the database.

We thus identify the wavelengths of interest. A disadvantage of CART is that it selects only one wavelength to define each split. Nevertheless, the selected wavelengths seem to correspond to the ROH and CH_2 groups. The differences are attributed to the composition of the cell walls, which differs for the polysaccharides between genus and species of *Enterococcus* et *Lactococcus*.[36] The polysaccharide region is evoked not only by Dziuba and Nalepa [35] but also by Savic et al.,[17] Amiel et al.,[16] and Naumann et al.[13] to explain the spectral discrimination between lactic acid bacteria and propionic acid bacteria.

Conclusion

This work shows that NIR spectroscopy with a remote probe can be used for robust screening of *Enterococcus* and *Lactococcus* in raw colonies of lactic acid bacteria cultured in Petri dishes, at genus and species levels. The best CART model based on infrared data identified 95% and 87% of genera, in training and independent testing, respectively (these genera were initially identified by DNA isolation and PCR amplification); and 95% and 87% of the species, in training and independent test, respectively. The bacteria preparation required for this NIR technique is easier than that required for the mid-IR technique, which makes the NIR-based analytical technique promising for everyday use, as it is also rapid (<1 min), and nondestructive.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Bachmann HP, McNulty DA, McSweeney PLH, et al. Experimental designs for studying the influence of the raw milk flora on cheese characteristics: a review. *J Soc Dairy Technol* 1996; 49: 53.

2. Centeno JA, Menendez S and Rodriguez Otero JL. Main microbial flora present as natural starters in cebreiro raw cow's milk cheese (Northwest Spain). *Int J Food Microbiol* 1996; 33: 307.
3. Manolopoulou E, Sarantinopoulos P, Zoidou E, et al. Evolution of microbial populations during traditional feta cheese manufacture and ripening. *Int J Food Microbiol* 2003; 82: 153.
4. Duthoit F, Tessier L and Montel MC. Diversity, dynam ics and activity of bacterial populations in 'registered des ignation of origin' salers cheese by single strand conformation polymorphism analysis of 16s rRNA genes. *J Appl Microbiol* 2005; 98: 1198.
5. Demarigny Y, Sabatier C, Laurent N, et al. Microbial diversity in natural whey starters used to make traditional rocamadour goat cheese and possible relationships with its bitterness. *Italian J Food Sci* 2006; 3: 251.
6. Mauriello G, Moio L, Moschetti G, et al. Characterization of lactic acid bacteria strains on the basis of neutral volatile compounds produced in whey. *J Appl Microbiol* 2001; 90: 928.
7. Herreros MA, Fresno JM, Prieto MJG, et al. Technological characterization of lactic acid bacteria iso lated from armada cheese (a Spanish goats' milk cheese). *Int Dairy J* 2003; 13: 469.
8. Davis C. Enumeration of probiotic strains: review of cul ture dependent and alternative techniques to quantify viable bacteria. *J Microbiol Methods* 2014; 103: 9.
9. Wenning M, Breitenwieser F, Konrad R, et al. Identification and differentiation of food related bac teria: a comparison of FTIR spectroscopy and MALDI TOF mass spectrometry. *J Microbiol Methods* 2014; 103: 44.
10. Sandrin TR, Goldstein JE and Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: a review. *Mass Spectrom Rev* 2013; 32: 188.
11. Griffiths A, Gelbart W, Lewontin R, et al. *An introduc tion to genetic analysis*. New York, NY: W.H. Freeman, 2004.
12. Helm D, Labischinski H, Schallehn G, et al. Classification and identification of bacteria by Fourier transform infrared spectroscopy. *J Gen Microbiol* 1991; 137: 69.
13. Naumann D, Fijala V, Labischinski H, et al. The rapid differentiation and identification of pathogenic bacteria using Fourier transform infrared spectroscopic and multivariate statistical analysis. *J Mol Struct* 1988; 174: 165.
14. Goodacre R, Timmins EM, Burton R, et al. Rapid iden tification of urinary tract infection bacteria using hyper spectral whole organism fingerprinting and artificial neural networks. *Microbiology* 1998; 1157: 1157 1170.
15. Udelhoven T, Naumann D and Schmitt J. Development of a hierarchical classification system with artificial neural networks and FT IR spectra for the identification of bacteria. *Appl Spectrosc* 2000; 54: 1471.
16. Amiel C, Mariey L, Curk Daubie MC, et al. Potentiality of Fourier transform infrared spectroscopy (Ftir) for dis crimination and identification of dairy lactic acid bac teria. *Lait* 2000; 80: 445.
17. Savic D, Jokovic N and Topisirovic L. Multivariate stat istical methods for discrimination of lactobacilli based on their FTIR spectra. *Dairy Sci Technol* 2008; 88: 273.
18. Samelis J, Bleicher A, Delbes Paus C, et al. FTIR based polyphasic identification of lactic acid bacteria isolated from traditional Greek Graviera cheese. *Food Microbiol* 2011; 28: 76.
19. Dziuba B, Babuchowski A, Nałecz D, et al. Identification of lactic acid bacteria using FTIR spectroscopy and clus ter analysis. *Int Dairy J* 2007; 17: 183.
20. Dziuba B. Identification of selected species and subspe cies of Lactococcus genus with the use of FUR spectros copy and artificial neural networks. *Zywn Nauk Technol Ja* 2012; 19: 103.
21. Guibet F, Amiel C, Cadot P, et al. Discrimination and classification of enterococci by Fourier transform infra red (FT IR) spectroscopy. *Vib Spectrosc* 2003; 33: 133.
22. Kirschner C, Maquelin K, Pina P, et al. Classification and identification of Enterococci: a comparative pheno typic, genotypic, and vibrational spectroscopic study. *J Clin Microbiol* 2001; 39: 1763.
23. Baker GC, Smith JJ and Cowan DA. Review and re analysis of domain specific 16s primers. *J Microbiol Methods* 2003; 55: 541.
24. Martens H and Naes T. *Multivariate calibration*. New York: Wiley, 1989.
25. Breiman L, Friedman J, Stone CJ, et al. *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/ Cole, 1984.
26. Breiman L, Friedman JH, Olshen RA, et al. *Classification and regression trees*. New York, NY: Chapman & Hall, 1984.
27. Strobl C, Boulesteix AL and Augustin T. Unbiased split selection for classification trees based on the Gini Index. *Comput Stat Data Anal* 2007; 52: 483.
28. Visa S, Ramsay B, Ralescu A, et al. Confusion matrix based feature selection. *22nd midwest artificial intelligence and cognitive science conference 2011*, Cincinnati, OH, 16 17 April 2011, pp.120.
29. Wenning M and Scherer S. Identification of microorgan isms by FTIR spectroscopy: perspectives and limitations of the method. *Appl Microbiol Biotechnol* 2013; 97: 7111.
30. Heijenoort Jv. Formation of the glycan chains in the syn thesis of bacterial peptidoglycan. *Glycobiology* 2001; 11: 25R.
31. Osborne B, Fearn T and Hindle PH. *Practical NIR spec troscopy with applications in food and beverage analysis*. 2nd ed. Addison Wesley Longman Ltd: Harlow UK, 1993.
32. Osborne BG and Fearn T. *Near infrared spectroscopy in food analysis*. Harlow UK: Longman Scientific & Technical, 1986.
33. Workman J and Weyer L. *Practical guide to interpretive near infrared spectroscopy*. Boca Raton, FL: CRC Press, 2007.
34. McGoverin C and Manley M. Classification of maize kernel hardness using near infrared hyperspectral ima ging. *J Near Infrared Spectrosc* 2012; 20: 529.
35. Dziuba B and Nalepa B. Identification of lactic acid bac teria and propionic acid bacteria using FTIR spectros copy and artificial neural networks. *Food Technol Biotechnol* 2012; 50: 399.
36. Mistou MY, Sutcliffe IC and van Sorge NM. Bacterial glycobiology: rhamnose containing cell wall polysacchar ides in gram positive bacteria. *FEMS Microbiol Rev* 2016; 40: 464 479.