Available online at:
www.dairy-journal.org

Original article

Multivariate statistical methods for discrimination of lactobacilli based on their FTIR spectra

Dragiša SAVIĆ[1]*, Nataša JOKOVIĆ[1], Ljubiša TOPISIROVIĆ[2]

[1] Laboratory for Food and Biotechnology, Faculty of Technology, Bulevar oslobodjenja 124, 16000 Leskovac, Serbia
[2] Laboratory for Molecular Genetics of Industrial Microorganisms, Institute of Molecular Genetics and Genetic Engineering, Vojvode Stepe 444a, P.O. Box 23, 11010 Belgrade, Serbia

Abstract – In this study, 41 strains belonging to 6 *Lactobacillus* species originating from culture collections, and isolates from homemade cheese that had previously been genetically classified, were discriminated based on their FTIR spectra. The lactobacilli species *L. brevis, L. diolivorans, L. plantarum* and *L. rhamnosus* could be differentiated on the basis of their FTIR spectra by applying multivariate statistical techniques (Hierarchical Cluster Analysis, Principal Component Analysis and Multidimensional Scaling). The strains belonging to *L. paracasei* and *L. casei* could not be separated from each other, and formed one group that was distanced from the other investigated species. The combination of common multivariate statistical methods used for the analysis of FTIR spectra could provide a powerful method for discrimination and even the identification of lactobacilli on a species level. The strains and subspecies within the lactobacilli species could not be differentiated by applying multivariate exploratory statistical methods. The FTIR spectra of lactobacilli were robust against the variation in the consistency of the growth medium (broth or agar medium of the same composition), applied spectroscopic techniques (KRS-5 or tablets) and the recorded biomass quantities of 2 and 3 mg. Furthermore, the distances between the two spectra of the same selected strain, which were sampled ca. 12 months apart, were insignificant.

lactobacilli / FTIR spectroscopy / multivariate statistical methods

摘要 – 基于傅立叶变换红外光谱和多元统计分析方法鉴定乳杆菌。本文采用傅立叶变换红外光谱 (FTIR) 法对源于手工制作干酪中的 6 个种 41 株乳酸菌进行了鉴定。运用多元统计分析方法 (聚类分析, 主成分分析和多维尺度分析) 对短乳杆菌 (*L. brevis*), *L. diolivorans,* 植物乳杆菌 (*L. plantarum*) 和鼠李糖乳杆菌 (*L. rhamnosus*) 的 FTIR 光谱数据进行了分析。但对其中副干酪乳杆菌 (*L. paracasei*) 和干酪乳杆菌 (*L. casei*) 不能有效地识别。FTIR 光谱技术结合多元统计分析方法可以将乳杆菌鉴定到种, 但是应用这种探索性的多变量统计分析方法不能鉴定到株或亚种。通过研究发现乳杆菌的 FTIR 光谱与生长条件 (在相同成分的肉汤或琼脂培养基) 不同、使用的光谱制样技术 (KRS-5 光滑面或压片式) 以及用来记录的生物量 (2 mg 和 3 mg) 的不同而有显著的差异。此外, 保存期相差约 12 个月的同菌株之间的光谱没有显著差异。

乳杆菌 / 傅立叶变换红外光谱 (FTIR) / 多元统计方法

* Corresponding author (通讯作者): savic@ni.ac.yu

Résumé – Méthodes statistiques multidimensionnelles pour la discrimination des lactobacilles basée sur leur spectre IRTF. Dans cette étude, 41 souches appartenant à six espèces de *Lactobacillus* provenant de collections de cultures ou isolées de fromages faits maison, précédemment classées génétiquement, sont discriminées selon leur spectre infrarouge à transformée de Fourier (IRTF). Les espèces *Lactobacillus brevis*, *L. diolivorans*, *L. plantarum* et *L. rhamnosus* peuvent être différenciées à partir de leur spectre IRTF en appliquant des techniques statistiques multidimensionnelles (classification ascendante hiérarchique, analyse en composantes principales et positionnement multidimensionnel). Les souches de *L. paracasei* et *L. casei* ne peuvent pas être séparées entre elles et forment un groupe distant des autres espèces étudiées. La combinaison de méthodes statistiques multidimensionnelles, couramment utilisées pour l'analyse de spectres IRTF, peut fournir une méthode puissante pour la discrimination et même l'identification de lactobacilles à l'échelle de l'espèce. Les souches et sous-espèces parmi les espèces de lactobacilles ne peuvent pas être différenciées en appliquant des méthodes statistiques multidimensionnelles exploratoires. Les spectres IRTF des lactobacilles sont robustes quels que soient les changements de consistance des milieux de culture (milieu liquide ou gélosé de la même composition), les méthodes spectroscopiques appliquées (KRS-5 ou comprimé) ou la quantité de biomasse analysée (2 ou 3 mg). De plus, la distance entre deux spectres d'une même souche sélectionnée, échantillonnée avec 12 mois d'écart, est insignifiante.

lactobacilles / spectroscopie IRTF / méthodes statistiques multidimensionnelles

1. INTRODUCTION

The great potential of Fourier Transform Infrared (FTIR) spectroscopy, being a rapid, easy to use and cost-effective technique, has been recognized in discrimination, classification and identification of bacteria, yeast and other microorganisms (for a review, see [14]). The cells are measured in a non-destructive manner and the infrared spectra of the microorganism studied are considered as its chemical fingerprint. The FTIR spectrum is given by the vibrational properties of all the cell constituents, including fatty acids, proteins, polysaccharides and nucleic acids [2]. A database with the reference microorganisms' spectra has to be established, and the unknown microorganisms can be identified by comparing their FTIR spectra with the spectra in the database.

The most commonly used FTIR technique applied so far for recording spectra in studies of microorganism identification is a transmission by placing the sample on a ZnSe optical plate [1, 2, 4, 16–18, 22]. Besides this, spectroscopic studies using the samples in the form of tablets have also been used for discrimination of species

within the genera *Listeria* [12] and *Brucella* [5]. The samples were prepared by collecting the biomass from agar plates [6, 8, 16–18, 22] or a centrifuged broth culture [1, 5, 10, 12, 13, 20].

Since the amount of the data obtained is large, the differences between the spectra of the microorganisms are described by using statistical methods. Many statistical techniques have been extensively applied in the analysis of the infrared spectra in microorganism discrimination: Partial Least Squares Regression [16–18], Soft Independent Modeling of Class Analogy [10, 17], K-Nearest Neighbor [10, 17] and Artificial Neural Networks [6]. However, multivariate exploratory techniques, such as Hierarchical Cluster Analysis [2, 4, 8, 17], Principal Component Analysis [5, 10, 17], Discriminant Analysis [1, 22] and Canonical Variate Analysis [12] have been the most common and most often applied statistical methods. So far, there have been no studies on the use of Multidimensional Scaling (MDS) analysis for differentiation of microorganisms based on their FTIR spectra. In general, the goal of MDS analysis is to detect meaningful underlying dimensions that allow the researcher

to explain the observed similarities or dissimilarities (distances) between the objects investigated. Recently, MDS was used to obtain a bidimensional representation of the position and the diversity of bacterial communities based on their phenotypic properties [19].

Since lactic acid bacteria (LAB) are widely used in the food industry, especially as starters in fermentation processes, the rapid and reliable identification and characterization of these microorganisms are important. There are few reports on the analysis of the relationships and the identification of LAB using FTIR spectroscopy. Oust et al. [17] analyzed 56 strains from four *Lactobacillus* species (*L. sakei*, *L. plantarum*, *L. curvatus* and *L. paracasei*). By applying FTIR analysis, the authors succeeded in recognizing the strains that were incorrectly identified by conventional methods prior to the FTIR analysis. Likewise, FTIR spectroscopy correctly identified 91% of the facultatively heterofermentative lactobacilli previously identified with SDS-PAGE, with high scores for *L. plantarum* (100% identification) and *L. rhamnosus* (91%), and a relatively low score for *L. casei* (87%) [22]. The results of Lefier et al. [13] demonstrated that FTIR was a rapid and robust method for a qualitative analysis of the density of lactococcal flora during soft cheese maturation. The identification of 100% at the genus level and 69% at the species level for LAB isolates from the cheese industry was established by using FTIR [1]. Moreover, 53 strains from 14 *Lactobacillus* species found in breweries were analyzed by FTIR spectroscopy and identification at the species level was correct for 94% of the strains, and at the strain level for 91% of the strains [2].

The main purpose of this study was to discriminate and to find the relations among *Lactobacillus* isolates from homemade cheeses that had previously been genetically classified, and species originating from culture collections, on the basis of their FTIR spectra. The quality of discrimination was evaluated by comparing the identification results from commonly used multivariate exploratory statistical methods and, for the first time, a Multidimensional Scaling Analysis. The influence of the growth conditions (the agar plate and broth of the same composition), the spectroscopic sampling technique applied (a transmission through tablets or optical plate), recorded biomass quantity and storage by freezing on lactobacilli discrimination was also examined.

2. MATERIALS AND METHODS

2.1. Strains and growth conditions

Forty-one *Lactobacillus* strains of different origin (Tab. I) were analyzed with FTIR spectroscopy. The strains under investigation were reference strains from culture collections and isolates from homemade cheeses that had previously been genetically classified [9]. Four strains were *L. brevis* (the same strain type obtained from two collections, LMG 6906 and NRRL B-4527, was the subject of the study), 5 strains were *L. diolivorans,* 2 strains were *L. casei,* 7 were *L. paracasei* (2 subspecies), 20 strains were *L. plantarum* (2 of which belonged to 2 subspecies) and 3 were *L. rhamnosus.* The nomenclatures of the strains used are cited from the culture collections, and the taxonomic status of the species and the strains at issue is not discussed.

The strains were cultivated on MRS agar and MRS broth (Torlak, Belgrade) at 30 °C for 48 ± 1 h. A platinum loop was used to inoculate the agar plate (streaking out) and the broth. Three sampling replications were done for each culture.

Table I. Lactobacilli strains used in this study.

Species	Strain	Isolation environment
L. brevis	NRRL[1] B-4527[T] (LMG 6906[T])	Feces, human
	LMG 7761	Green fermenting olives
	LMG 11401	Beer
	LMG 18022	Zabady (yogurt) Burundi
L. diolivorans	BGGJ 7-18	Cheese
	BGGJ 7-24	
	BGGJ 7-27	
	BGGJ 8-25	
	BGGJ 8-13	
L. casei	NRRL B-1922[T]	Cheese
	NRLL B-441	Cheese, Emmental
L. paracasei ssp. paracasei	NRRL B-4560[T]	Milking machine
	NRRL B-4564[T]	Saliva, child
	LMG 11459	Dental caries
	BGBUK2-16/K2	Cheese
	CI 5-25*	20-day ripened cow's cheese
	CI 2-21*	Ewe's milk
L. paracasei ssp. tolerans	LMG 9191[T]	Pasteurized milk
L. plantarum	NRRL B-4496[T]	Pickled cabbage
	LMG 18024	Buffalo milk, Egypt
	LMG 18023	Mish cheese, Egypt
	LMG 9212	Italian human, saliva
	BG A112	Cheese
	CI 4-3*	10-day ripened cow's cheese
	CI 6-8*	30-day ripened cow's cheese
	CI 6-9*	
	CI 6-11*	
	CI 6-16*	
	CI 6-28*	
	CI 7-12*	60-day ripened cow's cheese
	CI 7-16*	
	CI 7-20*	
	CI 7-21*	
	CI 7-22*	
	CI 7-31*	
	CI 7-32*	
L. plantarum ssp. argentoratensis	LMG 9205[T]	Fermented corn product
L. plantarum ssp. rudensis	LMG 11475	Cheese showing rusty spot
L. rhamnosus	NRRL B-442[T]	–
	BGZL-19	Cheese
	BGHV-747	Vaginal isolate

[1] **NRRL**: Agricultural Research Service Culture Collection, National Center for Agricultural Utilization Research, Peoria, Illinois, USA.
LMG: Bacteria Collection, Laboratorium voor Microbiologie, Universiteit Gent, Belgium.
BG: Bacteria Collection, Institute of Molecular Genetics and Genetic Enginering, Belgrade, Serbia.
[T] Marked as Type strain in the Culture Collection.
* CI: Isolates from cheese determined by Rep-PCR [9].

2.2. Sample preparation for FTIR spectroscopy

After the cultivation on agar plates, a metal spatula was used to remove the bacterial biomass. The cells grown in broth were separated by centrifugation (5000 min^{-1} for 15 min), washed twice with distilled water and dried at 60 °C for 60 min under a moderate vacuum.

Optical plate measurement

Preliminary trials were performed to determine the density of the cell suspensions, necessary to produce spectra with a good signal to noise ratio without band saturation. The required densities were obtained by transferring 5 mg of collected biomass from an agar plate to a microtube (1.5 cm^3) and suspending it in 0.250 cm^3 of distilled water. 0.1 cm^3 (containing ca. 2 mg of biomass) of the aliquot was placed on a KRS-5 optical plate (25 × 4 mm, DSC pol. Crystal Labs, NY), and dried under a moderate vacuum in a desiccator to form film suitable for FTIR measurement. Since previous studies showed that drying of between 15 and 45 min had no influence on a FTIR spectrum (data not shown), the samples were desiccated for ca. 20 min.

Tablet preparation

The biomass collected from dried sediments or the agar plate (ca. 2 or 3 mg) was mixed and ground with 150 mg KBr (Merck, Germany). This powder was then compressed into a thin KBr disc (∅ 13 mm) under a pressure of 200 MPa for about 5 min under a moderate vacuum.

2.3. FTIR spectroscopy

All the spectra were recorded in the region of between 4000 and 400 cm^{-1} using a BOMEM MB-100 FTIR Spectrometer (Hartmann & Braun, Quebec, Canada) equipped with a DTGS detector model D31B. For data acquisition, the software Win-Bomem Easy 3.1 was used. For each sample recording (through tablets or optical plate), 10 scans were averaged with a spectral resolution of 4 cm^{-1}.

2.4. Preprocessing of the spectra

The IR spectra obtained by recording the samples on the KRS-5 optical plate or in the form of tablets were in accordance with the quality control test based on the criteria of a signal-noise ratio and signal water relations [4]. In order to remove variations due to baseline shifts, the first derivative of every spectrum was calculated using a Savitzky-Golay algorithm with nine smoothing points [8, 17].

2.5. Data analysis

All statistical analyses were performed using the STATISTICA 7.0 program for Windows (StatSoft Inc., USA). The processed spectra of the strains were taken as variables (columns of the input matrix) and wavenumbers as statistical cases (rows of the matrix). Multivariate exploratory techniques, Hierarchical Cluster Analysis, Principal Component Analysis and Multidimensional Scaling Analysis, were performed to compare the results of lactobacilli discrimination. As it was found to be best suited for lactobacilli identification [1, 2, 17, 18], the region between 1400 and 700 cm^{-1} was used for discrimination of the species.

Hierarchical Cluster Analysis (HCA) was used to measure the similarities and to classify the lactobacilli spectra into groups. The distances were calculated using Pearson's product moment correlation coefficient, and clustering was carried out using Ward's algorithm. The Statistica software calculates and

transforms Pearson's correlations into dissimilarities (1 – Pearson *r*), meaning that a value of 1.00 represents a lack of correlation, whereas a value of 0.00 represents a perfect correlation.

Principal Component Analysis (PCA) was performed by using the algorithm known as NIPALS (Non-linear Iterative Partial Least Squares). The NIPALS algorithm is an iterative procedure used for constructing PCA models for the representation of multivariate variables using a number of principal components that are less in dimension than a number of the original variables. It is the most commonly used method for calculating the principal components, and it provides numerically more accurate results as compared with other methods [21].

Multidimensional Scaling (MDS) is typically used as an exploratory method. To discriminate the spectra of the lactobacilli used in this study, the matrix of the Pearson's product moment correlation coefficients was computed and submitted to MDS.

2.6. Experimental setup

Experiments were performed in two parts. The first part of the experiment involved the assessment of the importance of the recording technique. The spectra of eight strains, representing the species studied (*L. plantarum* NRRL B-4496 and CI 7-12, *L. plantarum* ssp. *argentoratensis* LMG 9205, *L. brevis* NRRL B-4527, *L. diolivorans* BGGJ 8-25, *L. casei* NRRL B-441, *L. paracasei* ssp. *paracasei* LMG 11459 and *L. rhamnosus* NRRL B-442) were obtained by using two transmission techniques, placing the samples both as tablets and on the KRS-5 optical plate. To provide approximately the same quantity of biomass for FTIR spectra recording,

ca. 2 mg of the biomass collected from the agar plate was used both for tablet preparation and for placing on the KRS-5 plate (0.1 cm^3 suspension of ca. 5 mg biomass from the agar plate in 0.250 cm^3 of distilled water). In order to determine the effects of the biomass concentration on FTIR spectra, the tablets were also prepared with ca. 3 mg of the collected biomass. Additionally, spectra were recorded by using tablets prepared with ca. 3 mg of the biomass collected from the cells grown in MRS broth.

In the second part, the spectra of all the rest of the reference strains and the isolates from cheese (Tab. I) were recorded using the tablets that were prepared with ca. 3 mg of the biomass collected from the agar plate. Finally, the spectra of all the strains were recorded ca. 12 months apart after storing bacteria by freezing (–18 °C) in a mixture of glycerin (15%) and MRS broth. After storing, the spectra of the bacteria were recorded by applying the tablet technique (with ca. 3 mg biomass), after reviving them by subculturing twice on the MRS agar (48 h, 30 °C).

3. RESULTS

During the FTIR measurement, a typical spectrum of lactobacilli with various spectral regions was obtained (data not shown) as described earlier [17]. In data analysis, only the parts of the spectra between 1400 and 700 cm^{-1}, which cover the mixed region, the polysaccharide region and the "true" [17] fingerprint region, were applied for statistical analysis, since that region was found to be best suited for differentiation and the identification of the *Lactobacillus* species [1, 2, 17, 18]. The absorption bands at ∼1080 cm^{-1} and ∼1242 cm^{-1} assigned to symmetric and asymmetric stretching vibration of phosphate functional groups of nucleic acids, respectively [10], dominated the FTIR spectra region of between 1400 and 700 cm^{-1} of
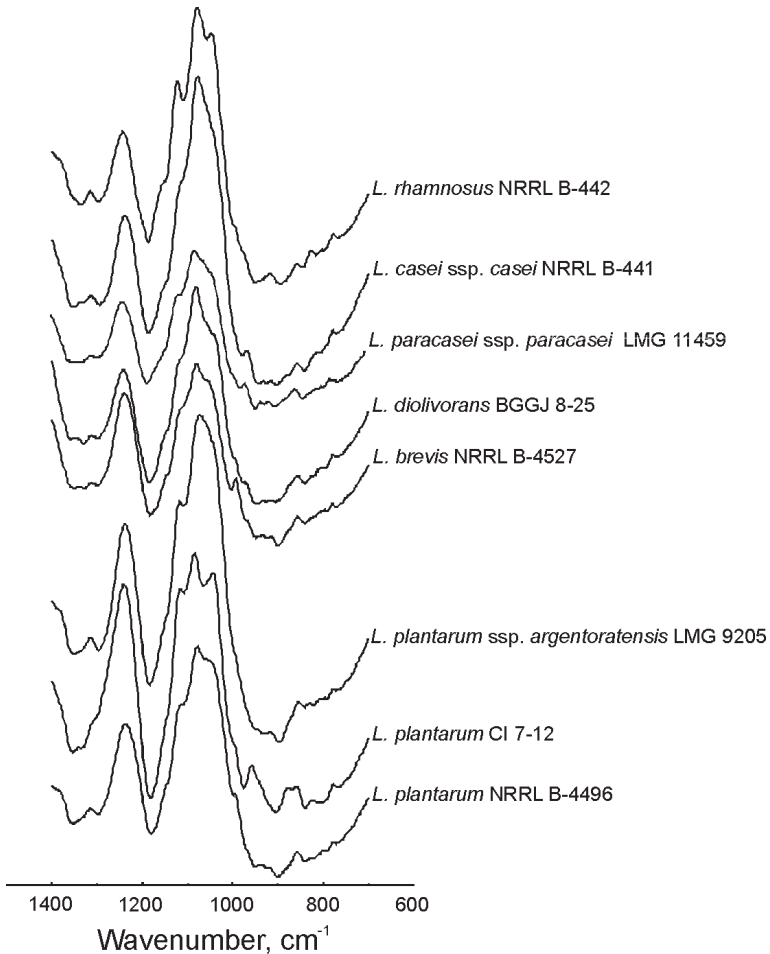
Figure 1. The 1400–700 cm^{-1} region of FTIR spectra (KBr tablets, 3 mg of biomass) of selected *Lactobacillus* strains cultivated on MRS agar (30 °C for 48 ± 1 h).

all the selected lactobacilli (Fig. 1). The significant differences in the raw spectra of the selected lactobacilli could be observed in the range of 1200–950 cm^{-1}.

3.1. Effects of growth conditions, spectroscopic technique and biomass concentration

In preliminary studies (experiment part I) eight strains, representing the species studied (*L. plantarum* NRRL B-4496 and CI 7-12, *L. plantarum* ssp. *argentoratensis* LMG 9205, *L. brevis* NRRL B-4527, *L. diolivorans* BGGJ 8–25, *L. casei* NRRL B-441, *L. paracasei* ssp. *paracasei* LMG 11459 and *L. rhamnosus* NRRL B-442) were grown on MRS agar plates. The biomasses were collected from the agar plates and the spectra were recorded by applying two spectroscopic techniques (KRS-5 plate and tablets). To compare data, approximately the same

quantity of the biomass was put both on the KRS-5 plate and in tablets (ca. 2 mg). Since it was very difficult to measure small quantities, such as 2 mg, precisely, the tablets were prepared with ca. 2 or 3 mg of the biomass collected from the plate in order to determine the influence of the biomass concentration in that range on the spectra differentiation. In order to examine the possibility of including the lactobacilli spectra that grew poorly on the MRS agar (could not yield 2–3 mg of biomass per plate) in discrimination studies, the effects of the spectra of the cells grown in MRS broth of the selected strains were also included in this study.

For all the eight strains, it was shown that the effects on the spectra caused by growth in either broth or agar plate, applied spectroscopic technique and the biomass quantity, were so small that the separation of the strains was nearly unaffected. This was illustrated in the distances (expressed as 1 – Pearson's correlation coefficients) between the strains and within the strain (Tab. II). However, the average spectra distances within the same strain caused by variation in experimental conditions varied in the range of 0.01 to 0.04. The average distances between the spectra of the strains belonging to different species were significantly larger than spectral distances between the FTIR spectra of the same species. An exception was noticed in the case of *L. casei* NRRL B-441 and *L. paracasei* ssp. *paracasei* LMG 11459, where the interspecies distances were equal to intraspecies distances (0.03). Also, the spectra of the reference strains *L. plantarum* NRRL B-4496 and *L. plantarum* ssp. *argentoratensis* LMG 9205 showed a high level of interspecies similarity (spectra distance 0.03), but the spectrum of the isolate from cheese, *L. plantarum* CI7-12, was significantly distanced from them (spectra distances 0.08 and 0.10) (Tab. II).

For further studies, a technique of preparing tablets with ca. 3 mg of the biomass was selected despite the opinion [6] that preparing a sample as tablets was a slow and rather tedious method and replacement with the application of liquid samples was recommended. Preparing tablets was selected because it was faster and it allowed continuous laboratory work.

3.2. Discrimination of species

In order to determine the distances between the FTIR spectra of the species commonly found in cheeses, and to test a statistical approach, HCA was first performed to differentiate the average spectra of 26 reference strains from the culture collections (LMG, NRRL and BG). The dendrogram from the HCA of the FTIR spectra of the 7 reference strains (experiment part I) and the spectra of all the rest of the reference strains (recorded before and after storing as tablets with ca. 3 mg biomass from agar plates) was constructed by applying Pearson's product moment correlation coefficient for a distance measurement and Ward's algorithm as aggregation algorithm (Fig. 2A). Pearson's correlation coefficient was chosen since it is a common parameter used to quantify the strength of the linear association between variables [2, 8, 20], as well as Ward's algorithm, which is widely used and well suited for the FTIR spectra of microorganisms [14].

The reference strains *L. brevis* (interspecies linkage distance ca. 0.2), *L. diolivorans* (interspecies linkage distance ca. 0.1), *L. rhamnosus* (interspecies linkage distance ca. 0.1) and *L. plantarum* (interspecies linkage distance ca. 0.3) clustered according to the species description regardless of their origin. Even *L. plantarum* ssp. *argentoratensis* (LMG 9205) and *L. plantarum* ssp. *rudensis* (LMG 11475) were not separated from other *L. plantarum* species (Fig. 2A). The HCA technique applied could not separate *L. paracasei* ssp. *paracasei, L. paracasei* ssp.

Table II. Distance measurement average (1 – Pearson's r[1]) and standard deviation (in parenthesis) between FTIR spectra[2] of the strains representing the lactobacilli studied.

Strain	L. plantarum			L. brevis	L. diolivorans	L. casei	L. paracasei	L. rhamnosus
	NRRL B-4496	CI 7-12	ssp. argen. LMG 9205	NRRL B-4527	BGGJ 8-25	NRRL B-441	LMG 11459	NRRL B-442
L. plantarum NRRL B-4496	**0.01** (0.006)							
CI 7-12	0.08 (0.03)	**0.04** (0.01)						
ssp. argenoratensis LMG 9205	0.03 (0.01)	0.10 (0.03)	**0.02** (0.01)					
L. brevis NRRL B-4527	0.10 (0.01)	0.10 (0.03)	0.11 (0.01)	**0.02** (0.006)				
L. diolivorans BGGJ 8-25	0.11 (0.02)	0.14 (0.01)	0.11 (0.02)	0.08 (0.02)	**0.02** (0.01)			
L. casei NRRL B-441	0.08 (0.01)	0.13 (0.02)	0.05 (0.01)	0.14 (0.02)	0.08 (0.02)	**0.03** (0.01)		
L. paracasei ssp. paracasei LMG 11459	0.09 (0.02)	0.11 (0.01)	0.06 (0.01)	0.10 (0.02)	0.05 (0.01)	0.03 (0.01)	**0.01** (0.007)	
L. rhamnosus NRRL B-442	0.12 (0.02)	0.17 (0.02)	0.09 (0.02)	0.19 (0.02)	0.12 (0.02)	0.08 (0.01)	0.08 (0.01)	**0.02** (0.01)

[1] The Statistica software calculates and transforms correlations into dissimilarities (1 – Pearson r) [21].

[2] For each strain, four averaged spectra of three independent samples were obtained: 1 on a KRS-5 optical plate (grown on MRS agar) and 3 in the form of tablets (1 with ca. 2 mg and 1 with ca. 3 mg of cells grown on MRS agar, and 1 with ca. 3 mg biomass collected from cells grown in MRS broth). The values in the table were calculated as average distances between each of the strains' 4 spectra. The bold values represent average distances between the spectra of the same strain obtained by 4 different methods. Zero values between the same spectra were excluded from calculations.
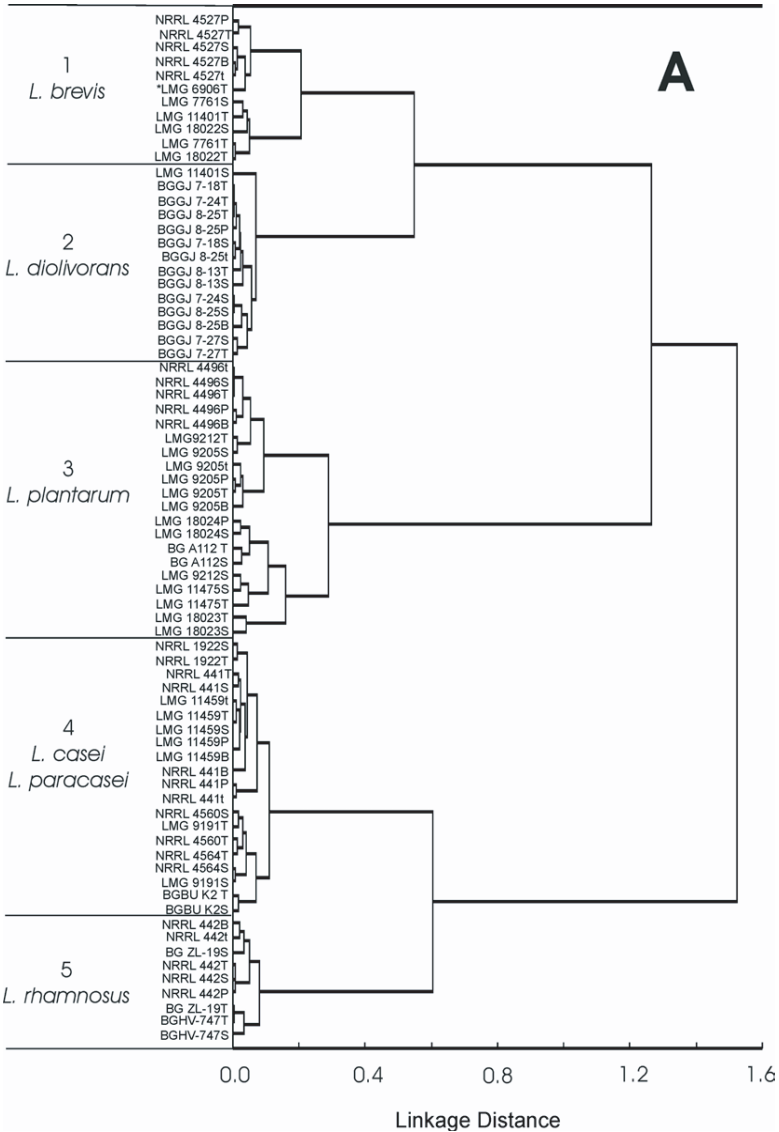
Figure 2. Dendrograms from HCA of the average spectra of lactobacilli investigated sampled ca. 12 months apart. The samples were recorded as tablets with ca. 3 mg of biomass collected from an agar plate before (suffix T) and after storing the bacteria by freezing (suffix S). The distances were calculated using Pearson's product moment correlation coefficient and grouping was performed by Ward's algorithm. (A) Dendrogram of strains from the culture collections including spectra of 7 reference strains recorded under different experimental conditions: on a KRS optical plate (suffix P), as tablets with ca. 2 mg of biomass collected from an agar plate (suffix t) and as tablets with biomass collected from broth by centrifugation (suffix B).

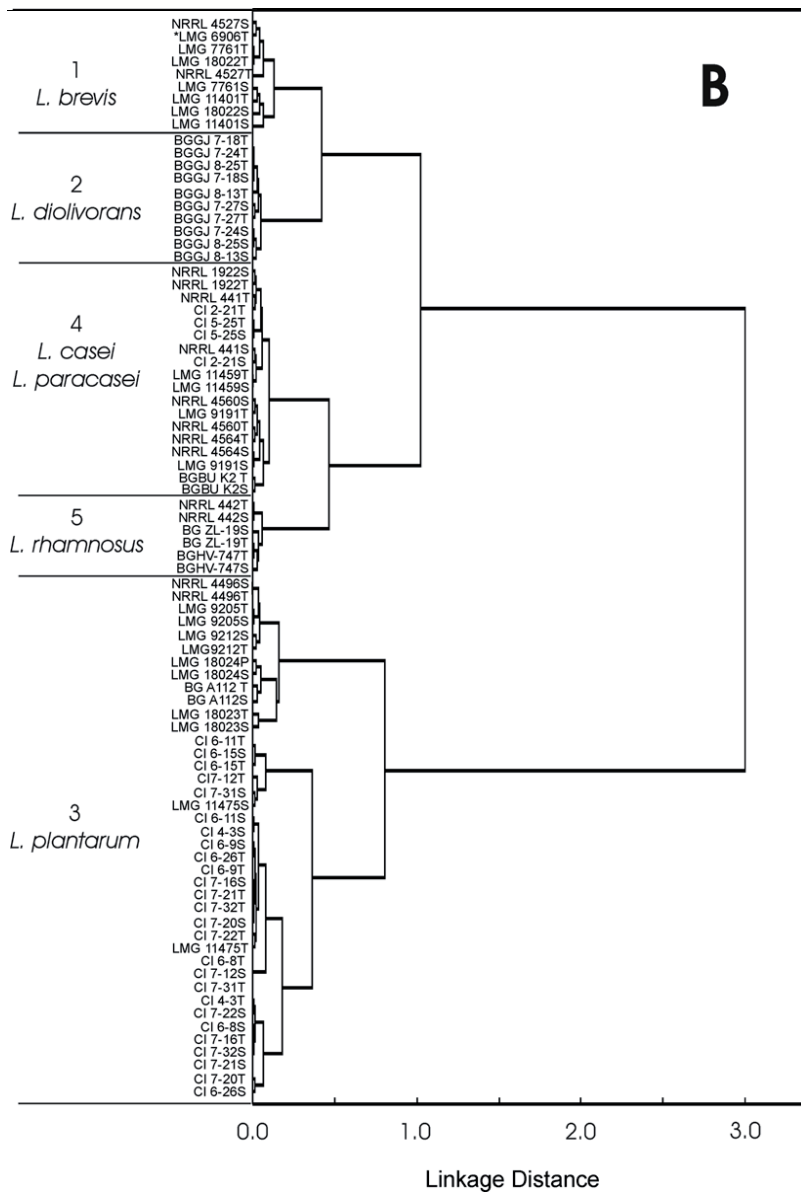* Strain *L. brevis* LMG 6906 was not recorded after storing since it was the same strain as NRRL B-4527.

Figure 2. Continued. (B) Dendrogram of spectra of all lactobacilli investigated.

tolerans and *L. casei*, since those strains formed one cluster corresponding to the linkage distance of 0.1 (Fig. 2A). So, with the exception of *L. paracasei* and *L. casei* strains, the reference strains were satisfactorily discriminated on the basis of their FTIR spectra regardless of the spectroscopic technique applied (KRS-5 or tablets), the growth medium consistency (agar or broth) and the quantity of the recorded biomass (2 or 3 mg per sample). Furthermore, the strains sampled ca. 12 months apart (before and after storing the bacteria by freezing) were assigned to their species. The only exception was *L. brevis* LMG 11401 which, after being stored, was shown to be closer to the *L. diolivorans* strains (Fig. 2A).

In order to find out if multivariate statistical methods could be used to discriminate and to identify lactobacilli of different origin at the species level based on FTIR spectroscopy, the isolates from homemade cheese that had previously been genetically classified were analyzed together with the reference strains. The average spectra of all the investigated lactobacilli obtained and recorded under the same conditions (recorded before and after storing as tablets with ca. 3 mg biomass from agar plates) were analyzed with HCA, PCA and MDS.

The isolates from homemade cheese fell into already defined clusters by introducing them into HCA (Fig. 2B). When the FTIR spectra of 13 identified *L. plantarum* isolates from cheese were analyzed together with 7 reference strains, they increased the interspecies distances from ca. 0.3 to ca. 0.9. This is probably due to differences among the isolates from cheese and the reference *L. plantarum* strains already noticed in part I of the experiment, where spectra of the strain CI7-12 were distanced from the reference strains *L. plantarum* NRRL B-4496 and *L. plantarum* ssp. *argenoratensis* LMG 9205 (Tab. II). However, all *L. plantarum* strains were clustered into

one group that was differentiated from the other investigated lactobacilli.

On the other hand, the introduction of 2 isolates, which were previously genetically classified as *L. paracasei* ssp. *paracasei*, did not significantly influence the interspecies distances in the *L. paracasei* and *L. casei* group. All the *L. plantarum* isolates seemed to be most related to *L. plantarum* ssp. *rudensis* LMG 11475, which they clustered with on a linkage distance of 0.4 (Fig. 2B). Furthermore, a dendrogram showed that the spectra of all the strains investigated, which were recorded ca. 12 months apart (before and after storing by freezing), clustered into the group corresponding to the species. When the number of strains was increased in HCA, the previous misgrouping of *L. brevis* LMG 11401 (Fig. 2A) was corrected and the strain sampled after storing was assigned to the *L. brevis* cluster (Fig. 2B). In this case, introducing more strains of the same species in a statistical analysis helped to achieve more appropriate differentiation among the lactobacilli.

The clusters formed by HCA analysis (Fig. 2) could be clearly observed on the plots from the PCA and the MDS (Figs. 3 and 4). Each cluster of the strains was represented by an ellipse for its samples. This type of ellipse was determined in such a way that the length of its horizontal and vertical projection onto the x- and y-axis (respectively) was equal to the mean ± (range 0.95), where the mean and the range refer to the X or Y variable [21].

PCA discriminated all lactobacilli, yielding 4 Principal Components (PC) that explained 96.9% of the total variance in the data (data not shown). The first Principal Component (PC1 accounted for 89.8% of the total variance) correlated well (correlation coefficients between 0.88 and 0.98, data not shown), with spectra of all the strains investigated. So, a physical meaning of PC1 could be attributed to the spectra of all lactobacilli. PC2 (accounting
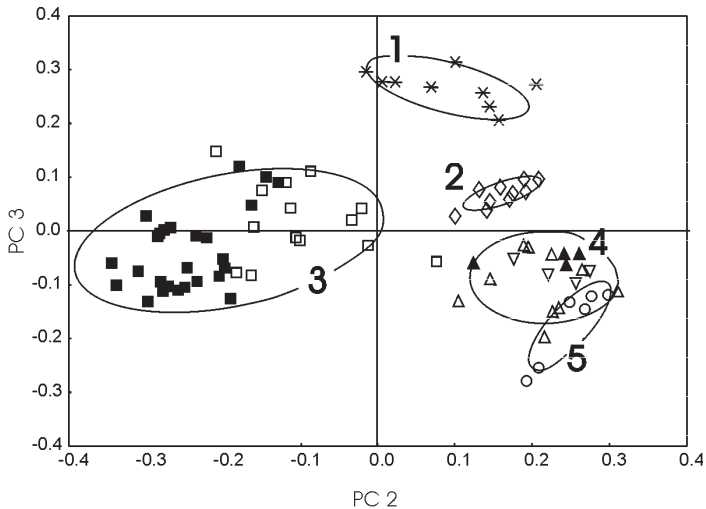
Figure 3. Principal component loadings from the PCA analysis of the reference strains' FTIR spectra (open symbols) and isolates that had previously been genetically classified (closed symbols). The explained variance by PC2 and PC3 were 4.2% and 1.6%, respectively. Numbers represent clusters from Figure 2.
L. brevis (✶), *L. plantarum* (□), *L. casei* (∇), *L. paracasei* (∆), *L. diolivorans* (◇), *L. rhamnosus* (○).

for 4.2% of the total variance) separated *L. plantarum* from all other investigated lactobacilli, while PC3 (1.6% of the total variance) separated *L. brevis* and *L. diolivorans* from *L. rhamnosus* and the group of *L. paracasei* and *L. casei* strains (Fig. 3).

To evaluate if MDS was able to provide a representation of the similarity relationships among the strains, which was coherent with that obtained by HCA and PCA, the matrix of the data obtained from the spectra was used for the calculation of the Pearson's product moment correlation coefficient for the distance measurement among the strains and then for MDS in two dimensions. The plot obtained by MDS (Fig. 4) was related to the dendrogram from HCA (Fig. 2). As shown in Figure 4, distances between strains belonging to the same cluster were not necessarily lower than distances between strains belonging to different clusters, despite the fact that the same distance metrics (Pearson) were

used for both statistical procedures, meaning that HCA and MDS arranged intercluster distances in a different manner.

The relationship between the positions of the ellipses obtained in plots from PCA and MDS (Figs. 3 and 4) and the position of the clusters identified with HCA (Fig. 2) was evident. The distances and the degree of overlapping among the ellipses indicated the relations among the strains investigated and the efficiency of a statistical technique at separating the clusters. It was obvious that lactobacilli belonging to *L. casei* and *L. paracasei* species could not be separated by applying multivariate statistical techniques to their FTIR spectra. As shown in Figure 4, MDS separated all the clusters without overlapping of the ellipses. The PCA could not clearly separate the *L. rhamnosus*, *L. paracasei* and *L. casei* clusters since there was a significant overlapping between ellipses representing these two clusters (Fig. 3).
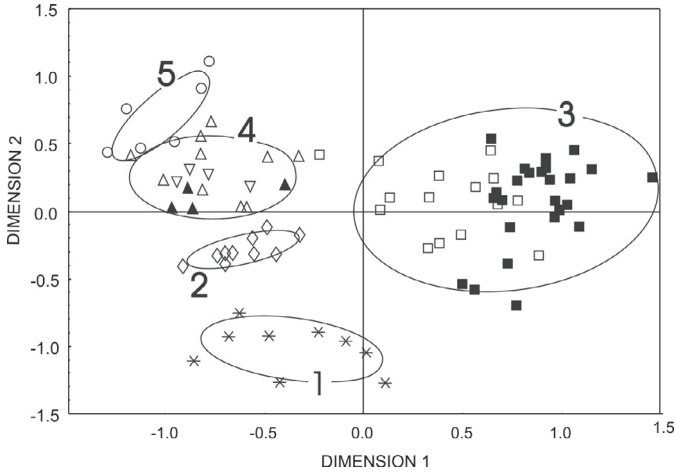
Figure 4. Bidimensional representation of the reference strains' FTIR spectra (open symbols) and isolates that had previously been genetically classified (closed symbols) obtained by MDS of the matrix of Pearson's product moment correlation coefficients. Numbers represent clusters from Figure 2.
L. brevis (✳), *L. plantarum* (□), *L. casei* (▽), *L. paracasei* (△), *L. diolivorans* (◇), *L. rhamnosus* (○).

The *L. plantarum* cluster ellipse, which represented 20 strains included in the study, was the largest (Figs. 3 and 4). However, the size of an ellipse was not necessarily proportional to the number of different strains indicated. In fact, the ellipse of 3 strains of *L. rhamnosus* species was disproportionately larger compared with the ellipse of 5 *L. diolivorans* strains (Figs. 3 and 4). This might reflect not only the interspecies diversity (all *L. diolivorans* strains were isolated from a single sample), but also the statistical approach used to evaluate the diversity.

From the plots obtained from PCA and MDS (Figs. 3 and 4), it can be seen that one point representing the *L. plantarum* strain is significantly distanced from the *L. plantarum* ellipse and shown to be closer to the cluster representing *L. paracasei* and *L. casei* strains. This point represented *L. plantarum* ssp. *argentoratensis* LMG 9205 recorded as tablets with 3 mg biomass collected from the agar plate. The same strain sampled after ca. 12 months

of storage was assigned into the correct species (*L. plantarum*) ellipse. The separation of this sample was not observed by applying HCA (Fig. 2).

4. DISCUSSION

According to a well-defined fact of building libraries with the reference spectra and standardization of the cultivation conditions, in order to ensure reproducible results as a prerequisite for using FTIR spectroscopy as a means for identification of lactobacilli, more strains and species have to be included in the library. In this paper, 41 lactobacilli strains belonging to six species commonly found in cheese were examined in order to help establish standard conditions for cultivating and standardizing a spectroscopic technique for recording FTIR spectra and finding a suitable statistical method for discrimination of lactobacilli. The results of applying statistical multivariate techniques

showed that FTIR spectra of lactobacilli were robust against a variation in the consistency of the growth medium (broth or agar medium of the same composition), the applied spectroscopic techniques (KRS-5 or tablets) and the biomass concentrations of ca. 2 and ca. 3 mg per sample. Furthermore, all spectra of the investigated strains, which were sampled ca. 12 months apart (before and after storing by freezing), clustered into groups corresponding to the species.

The applied multivariate statistical methods appeared to be very effective for discrimination of the four species *L. brevis, L. diolivorans, L. plantarum* and *L. rhamnosus* based on their FTIR spectra. In fact, the clusters obtained by HCA, and the position and the shape of the ellipses obtained by PCA and MDS were related to the species, and plotting the position of the clusters of the species on the same graph helped identify this relationship.

In previous studies of lactobacilli identification by applying HCA to their FTIR spectra, the data sets consisted of few strains from each of many species. With the exception of *L. diolivorans*, lactobacilli of all other species used in the current study were among 14 studied species isolated from beer in the research of Curk et al. [2]. Despite some differences, good discrimination could be observed between the lactobacilli species, as well as a high degree of relatedness between *L. casei* and *L. paracasei* strains. On the other hand, Oust et al. [17] showed that HCA was not suited for the analysis of FTIR spectra of microorganisms with respect to species differentiation since it could not separate lactobacilli on the species level when the number of strains was increased.

PCA modeling is used to replace a complex multi-dimensional data set, such as infrared spectra, by a simplified version involving fewer dimensions which facilitates an easier analysis of the variance in the data set [10]. In addition, in or-

der to reduce the dimensions of the original space of variables, PCA can also be used as a classification technique, so that the relations among the variables and cases can be highlighted [21]. To discriminate closely related lactobacilli, Oust et al. [17] used Partial Least Squared Regression (PLSR), a statistical method that extracts PLS components from a data matrix. It was shown that PC1 separated *L. plantarum* from three other studied species (*L. sakei, L. curvatus* and *L. paracasei*), PC2 separated *L. paracasei*, while PC3 separated *L. curvatus* from the rest of the species. The results were not directly comparable with the results in this paper because of the application of a different statistical technique for extracting components and a different model of data preprocessing. In the work of Oust et al. [17], the first derivation, vector normalization and weighting spectral data were used prior to statistical analysis. In the present study, the normalization of the spectra was not done, although it is a very often recommended and used method for vector or integral spectra normalization [10,14,17,18]. The method of normalization was not applied since the first derivative was the measurement of a curve slope, so normalization, i.e. moving spectra up or down, did not influence the first derivative. Furthermore, attempts at the additional weighting of spectral data by dividing them by the standard deviation of the data for each wavenumber, as recommended earlier [17], did not give satisfactory discrimination (data not shown). So, on the basis of this study and the results of Oust et al. [17], it can be concluded that statistical techniques which extract principal components from FTIR spectra make it possible for *L. plantarum* to be clearly separated from other lactobacilli.

According to the distances between the ellipses that represent clusters of the investigated strains, MDS showed better grouping (Fig. 4) and discrimination of the

investigated species than PCA (Fig. 3). In terms of resultant differences, PCA tends to extract more components (dimensions) than MDS; as a result, MDS often yields more readily interpretable solutions [21]. Therefore, the lactobacilli were better discriminated in the first two dimensions (Fig. 4) by using MDS analysis, but not by implementation of 3 or 4 components from PCA. MDS is not so much an exact procedure as rather a way to "rearrange" the objects in an efficient manner, so as to arrive at a configuration that best approximates the observed distances. The program actually moves the objects around in the space defined by the requested number of dimensions, and checks how well the distances between the objects can be reproduced by the new configuration. In more technical terms, the program uses a function minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit (or minimizing "lack of fit") [21].

By introducing the previously genetically classified isolates into the study, FTIR spectroscopic analysis came to the same result as a genomic identification analysis. In this way, genetic determinations of the isolates from cheese were confirmed by comparing their FTIR spectra with the spectra of the reference strains. This indicates that the reproducibility of the methods is good, which is important if FTIR spectroscopy is to be used for the identification of *Lactobacillus* at the species level. However, the average spectra distances of the same strain caused by variation in experimental conditions were not always lower compared with distances between the strains belonging to the same species, regardless of the statistical method applied. So, the strains and subspecies within the species cannot be differentiated by applying multivariate exploratory statistical methods to their FTIR spectra.

FTIR spectroscopy is not a reliable parameter for establishing taxonomic rela-

tionships between microorganisms [15], and the results of this study confirm difficulties in discrimination of phenotypically related species, such as *L. rhamnosus*, *L. casei* and *L. paracasei* [7]. By applying multivariate statistical methods to FTIR spectra, highly related strains belonging to *L. paracasei* ssp. *paracasei*, *L. paracasei* ssp. *tolerans* and *L. casei* cannot be separated from each other and form one group that is distanced from the other investigated species. The systematics of the *L. casei* and *L. paracasei* group have been problematic [3, 11] and adjustments have been introduced based on new taxonomic methods that have gradually become available. In this paper, the taxonomic status of the species and the strains used was accepted as suggested in the respective culture collections. In addition, it seems that the *L. rhamnosus* strain can be separated from the related group of *L. casei* and *L. paracasei* by applying common multivariate statistical methods, especially HCA and MDS, but it is obvious that more strains of these species have to be included in making libraries of FTIR spectra. Generally, the identification results for all lactobacilli species will probably be improved if more strains of different origin are included in the study.

Using HCA, PCA and MDS, the discrimination of all species was satisfactory, but there was at least one exception in each statistical approach. HCA misgrouped a *L. brevis* strain sampled after storage (Fig. 2A), while PCA and MDS distanced one *L. plantarum* strain and brought it closer to the group of *L. casei* and *L. paracasei* (Figs. 3 and 4). So, discrimination that is more complete and even the identification of lactobacilli can be provided by applying a combination of common multivariate statistical methods, such as HCA, PCA and MDS. By representation in the plot, MDS allows the measurement of dispersion, which supports some relationships with interspecies diversity and may help to

extend the differentiation capacity of FTIR towards an elegant identification technique of lactobacilli.

REFERENCES

[1] Amiel C., Mariey L., Curk-Daubie M.C., Pichon P., Travert J., Potentiality of Fourier transform infrared spectroscoy (FT-IR) for discrimination and identification of dairy lactic acid bacteria, Lait 80 (2000) 445–459.

[2] Curk M.C., Peladan F., Hubert J.C., Fourier-transform infrared (FT-IR) spectroscopy for identifying *Lactobacillus* species, FEMS Microbiol. Lett. 123 (1994) 241–248.

[3] Dellaglio F., Felis G.E., Torriani S., The status of the species, *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins et al., 1989. Request for an Opinion, Int. J. Syst. Evol. Microbiol. 52 (2002) 285–287.

[4] Fischer G., Braun S., Thissen R., Dott W., FT-IR spectroscopy as a tool for rapid identification and intra-species characterization of airborne filamentous fungi, J. Microbiol. Methods. 64 (2006) 63–77.

[5] Gomez M., Perez B., Gil M., Diez D., Rodriguez M., Rodriguez G., Domingo O., Torres R., Identification of species of *Brucella* using Fourier transform infrared spectroscopy, J. Microbiol. Methods. 55 (2003) 121–131.

[6] Goodacre R., Timmins M., Rooney P., Rowland J., Kell D., Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks, FEMS Microbiol. Lett. 140 (1996) 233–239.

[7] Hammas W., Hertel C., The genera *Lactobacillus* and *Carnobacterium*, in: Dworkin M., Falkow S., Rosenberg E., Schleifer K., Stackebrandt E. (Eds.), The Procaryotes, Vol. IV, Bacteria: Firmicutes, Cyanobacteria, Springer, New York, 2006, pp. 320–403.

[8] Helm D., Labischinski H., Schallehn G., Naumann D., Classification and identification of bacteria by Fourier-transform infrared spectroscopy, J. Gen. Microbiol. 137 (1991) 69–79.

[9] Joković N., Isolation and characterization of lactic acid bacteria isolated from cheese from Radan Mountain (in Serbian), M.Sc. thesis, Faculty of Biology, Belgrade, Serbia, 2004, pp. 71–75.

[10] Kansiz M., Heraud P., Wood B., Burden F., Beardall J., McNaughton D., Fourier transform infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains, Phytochemistry 52 (1999) 407–417.

[11] Klein G., International Committee on Systematic Bacteriology, Subcommittee on the taxonomy of *Bifidobacterium*, *Lactobacillus* and related organisms, Minutes of the meetings, 22–23 September 1999, Veldhoven, The Netherlands, Int. J. Syst. Evol. Microbiol. 51 (2001) 259–261.

[12] Lefier D., Hirst D., Holt C., Williams A., Effect of sampling procedure and strain variation in *Listeria monocytogenes* on the discrimination of species in the genus *Listeria* by Fourier transform infrared spectroscopy and canonical variates analysis, FEMS Microbiol. Lett. 147 (1997) 45–50.

[13] Lefier D., Lamprell H., Mazero G., Evolution of *Lactococcus* strains during ripening in Brie cheese using Fourier transform infrared spectroscopy, Lait 80 (2000) 247–254.

[14] Mariey L., Signolle J.P., Amiel C., Travert J., Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics, Vibrational Spectrosc. 26 (2001) 151–159.

[15] Oberreuter H., Charzinski J., Scherer S., Intraspecific diversity of *Brevibacterium linens*, *Corynebacterium glutamicum* and *Rhodococcus erythropolis* based on partial 16S rDNA sequence analysis and Fourier-

transform infrared (FT-IR) spectroscopy, Microbiology 148 (2002) 1523–1532.

[16] Oberreuter H., Mertens F., Seiler H., Scherer S., Quantification of microorganisms in binary mixed populations by Fourier transform infrared (FT-IR) spectroscopy, Lett. Appl. Microbiol. 30 (2000) 85–89.

[17] Oust A., Moretro T., Kirschner C., Narvhus J., Kohler A., FT-IR spectroscopy for identification of closely related lactobacilli, J. Microbiol. Methods. 59 (2004) 149–162.

[18] Oust A., Moretro T., Kirschner C., Narvhus J., Kohler A., Evaluation of the robustness of FT-IR spectra of lactobacilli towards changes in the bacterial growth conditions, FEMS Microbiol. Lett. 239 (2004) 111–116.

[19] Parente E., Ricciardi A., A statistical procedure for the analysis of microbial communities based on phenotypic properties of isolates, J. Microbiol. Methods. 49 (2002) 121–134.

[20] Schuster K.C., Mertens F., Gapes J.R., FTIR spectroscopy applied to bacterial cells as a novel method for monitoring complex biotechnological processes, Vibrational Spectrosc. 19 (1999) 467–477.

[21] StatSoft, Inc. Electronic Statistics Textbook, Tulsa, OK: 2006, StatSoft, WEB: http://www.statsoft.com/textbook/stathome.html.

[22] Weinrichter B., Luginbühl W., Rohm H., Jimeno J., Differentiation of facultatively heterofermentative Lactobacilli from plants, milk, and hard type cheeses by SDS-PAGE, RAPD, FTIR, energy source utilization and autolysis type, Lebensm.-Wiss. Technol. 34 (2001) 556–566.