spectroscopy journal

MDPI

Article

Selective Enumeration and Identification of a Multi-Strain Probiotic Consortium Using Fourier Transform Infrared Spectroscopy Paired with Plate Count: A Proof-of-Concept Study

Francesca Deidda [1], Miriam Cordovana [2], Carlotta Morazzoni [1], Serena Allesina [1], Matteo Calgaro [3], Nicola Vitulo [3], Martina Bausani [1] and Marco Pane [1,*]

[1] Probiotical Research Srl, 28100 Novara, Italy; f.deidda@probiotical.com (F.D.); m.bausani@probiotical.com (M.B.)
[2] Bruker Daltonics GmbH, 28359 Bremen, Germany; miriam.cordovana@bruker.com
[3] Department of Biotechnology, University of Verona, 37129 Verona, Italy
* Correspondence: m.pane@probiotical.com

Abstract

Background: The accurate enumeration and identification of probiotic strains are essential for product quality. The plate count (PC) gold standard enumerates viable, culturable cells but does not by itself resolve individual strains within multi-strain consortia, and molecular methods (qPCR, ddPCR) are costly and face recognised challenges in quantifying relative strain abundance. Fourier transform infrared (FTIR) spectroscopy is a promising phenotypic alternative. Methods: We developed an FTIR-based artificial neural network classifier to identify and quantify a four-strain probiotic blend comprising *Lactobacillus acidophilus* LA02, *Lacticaseibacillus rhamnosus* LR04, *Limosilactobacillus fermentum* LF08, and *Bifidobacterium animalis* subsp. *lactis* BS01 compared against selective plate counting and species-specific PCR. Results: The classifier correctly identified all 36 test colonies (100%; 95% Clopper–Pearson CI: 90.3–100%); descriptive cluster analysis indicated spectral distinctiveness (silhouette = 0.908; Davies–Bouldin = 0.127; cophenetic correlation = 0.955). Enumeration agreement with selective plate counting was assessed descriptively (Pearson r = 0.78, 95% CI [−0.72, 1.00], $n = 4$ strain means; mean difference −0.028 $\log_{10}$ CFU/mL; all strain-level differences < 0.1 $\log_{10}$ CFU/mL). PCR confirmed all FTIR classifications. Conclusions: This proof-of-concept study demonstrates the feasibility of coupling cultivation with spectroscopic identification in a hybrid PC + FTIR workflow for multi-strain probiotic quality control. Because the findings derive from a single blend preparation analysed in technical replicates, they characterise this dataset and require confirmation on independently prepared batches before the approach can be regarded as a validated method.

Keywords: probiotics; enumeration; identification; fourier transform infrared spectroscopy; IR Biotyper; plate count; flow cytometry; live biotherapeutic products; machine learning; quality control

Check for updates

Academic Editor: Vincent Boudon

Received: 16 June 2026
Revised: 21 July 2026
Accepted: 24 July 2026
Published: 30 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Probiotics are defined as "live microorganisms that, when administered in adequate amounts, confer a health benefit on the host", as established by a joint FAO/WHO working group and reaffirmed by an expert International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus panel [1,2]. This definition underscores two critical requirements for probiotic products: the microorganisms must be viable, and they must

be present in sufficient quantities to exert beneficial effects. Consequently, the accurate enumeration of viable probiotic cells is fundamental to product quality assurance and regulatory compliance.

The plate count (PC) method has long served as the gold standard for enumerating viable probiotic bacteria and is endorsed by regulatory agencies worldwide [3]. This technique involves counting colonies grown on agar after inoculating serially diluted samples, with results expressed as colony-forming units (CFUs) that estimate the number of viable, culturable organisms in the original sample. It is important to note that PC therefore reports the total plate counts rather than the total viable counts in the strict sense: cells in a viable but non-culturable (VBNC) state, or cells injured by processing, do not form colonies on synthetic media and are not captured. VBNC cells are characterised by their loss of culturability, their ability to maintain a nonculturable state, and their potential to resuscitate [4]; the resuscitation potential is central to the definition, since it distinguishes VBNC cells from dead cells and counters the view that VBNC populations are simply non-viable [5]. Furthermore, PC is an enumeration technique, and on a non-selective medium, it does not by itself assign colonies to individual strains within a multi-strain consortium. Strain-level resolution therefore requires an additional discriminating step—selective media exploiting differential antibiotic resistance or metabolic traits, or a downstream identification method applied to isolated colonies. This constitutes a practical limitation for quality control given the increasing market prevalence of products containing multiple probiotic strains [6].

Several innovative approaches have been explored to overcome this limitation. Molecular methods such as quantitative PCR (qPCR) and digital droplet PCR (ddPCR), along with flow cytometry with viability staining, have advanced microbial detection capabilities [3]. However, these genome-based techniques present notable drawbacks: they are costly, time-consuming, and often unreliable for accurately quantifying the relative abundance of multiple strains in blended formulations. Furthermore, both product quality and efficacy depend on phenotypic traits—including cell viability, metabolic activity, and functional properties—that DNA-based methods cannot assess [7].

Within this context, phenotypic identification technologies have gained attention as alternatives for selective strain enumeration. Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) has been applied to characterise multi-strain live biotherapeutic products with encouraging results [8]. Fourier transform infrared (FTIR) spectroscopy represents another promising approach, demonstrating the ability to discriminate microorganisms at various taxonomic levels—from species and subspecies down to strain level—across clinical [9–12], food [13–15], veterinary [16–18], and environmental applications [19,20].

FTIR spectroscopy functions by measuring the absorption of infrared light by cellular biomolecules at specific wavelength regions [21]. Polysaccharides absorb strongly in the 1200–900 cm^{-1} range, C–H bond vibrations appear around 3000–2800 cm^{-1}, and protein amide groups absorb in the 1800–1500 cm^{-1} region [22,23]. Consequently, each bacterium produces a unique infrared absorption spectrum—a molecular "fingerprint" reflecting its overall biomolecular composition—that enables precise identification and differentiation, even at the strain level.

The present study introduces a novel proof-of-concept approach for enumerating and identifying members of a four-strain probiotic consortium based entirely on cultivable, phenotypic characteristics. The approach integrates FTIR spectroscopy with traditional plate counting in a tandem workflow. We developed an automated classifier using an artificial neural network (ANN) trained on FTIR spectral data to distinguish the four probiotic strains following growth on agar. The performance of this FTIR-based method was

then evaluated by comparison with the standard selective plate count method. Figure 1 illustrates the overall workflow. Our primary aim was to assess whether FTIR spectroscopy, paired with machine learning, can achieve selective strain enumeration comparable to conventional techniques, thereby providing a rapid and cost-effective tool for probiotic product analysis.

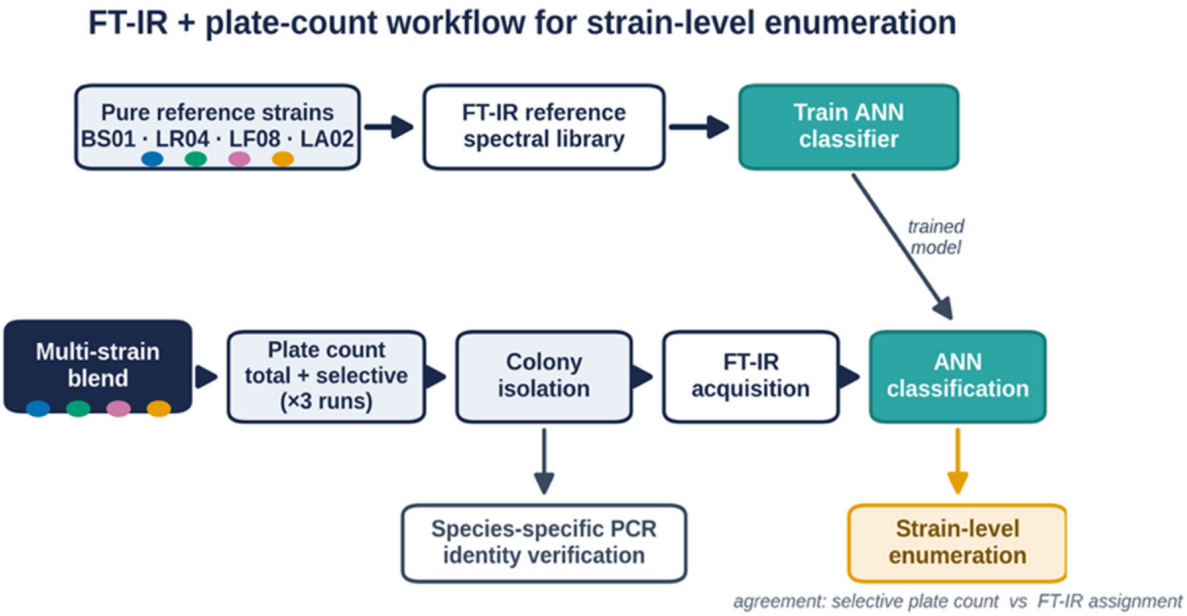

Figure 1. Schematic workflow integrating plate count and FTIR analysis for selective strain enumeration. The procedure involves: (1) Selection of individual probiotic strains and preparation of pure cultures; (2) Machine learning model training (ANN classifier) on FTIR spectra of pure strains; (3) Blending of strains into a multi-strain consortium; (4) Total plate count and selective plate counts; (5) Isolation of single colonies and propagation in broth culture; (6) FTIR spectral acquisition and automated classification; (7) Determination of each strain's count based on FTIR assignments; (8) PCR verification; (9) Comparison of selective plate count versus FTIR-based methods. Strain identity is colour-coded throughout using the Okabe–Ito colourblind-safe palette applied in all figures: *B. animalis* subsp. *lactis* BS01, blue; *L. rhamnosus* LR04, green; *L. fermentum* LF08, reddish-purple; *L. acidophilus* LA02, orange.

2. Materials and Methods

2.1. Probiotic Strain Selection and Preparation

Four probiotic bacterial strains were selected for this proof-of-concept study, each representing a different genus: *Lactobacillus acidophilus* LA02, *Lacticaseibacillus rhamnosus* LR04, *Limosilactobacillus fermentum* LF08, and *Bifidobacterium animalis* subsp. *lactis* BS01. Strains were obtained in lyophilised form from Probiotical SpA (Novara, Italy) and were selected based on phenotypic traits expected to facilitate selective discrimination using different media and antibiotics [6].

Each strain was initially enumerated by standard plate count to establish its culturable titre. Total plate counts for the lyophilised stocks were on the order of 10^{11} CFU/g (LA02: 1.47×10^{11}; LR04: 1.08×10^{11}; LF08: 1.05×10^{11}; BS01: 1.63×10^{11} CFU/g). Cell counts were also determined by fluorescence flow cytometry yielding approximately 2.5–3.8 $\times$ 10^{11} cells/g. Cells were dual-stained with thiazole orange (TO) and propidium iodide (PI) (BD Cell Viability Kit, BD Biosciences, Cat. no. 349483) in accordance with ISO 19344 [24]. Gating was performed on a TO-positive primary gate to discriminate bacterial events from instrument noise and medium debris; the values reported here correspond to total fluorescent units (TFUs), defined as all TO-positive events irrespective of PI signal,

and therefore represent total (intact plus permeabilised) cells rather than active fluorescent units (AFU, TO-positive/PI-negative events). TFU values served a single purpose in this study: they were used to formulate the test blend on an equal-total-cell basis, as described in Section 2.2. They are a formulation input and do not enter the enumeration comparison reported in Section 3, both arms of which are plate counts.

The four strains were combined to create a multi-strain probiotic consortium (hereafter referred to as the "test blend"). The blend was formulated on an equal-total-cell basis: the four lyophilised stocks were combined so as to contribute equal numbers of total cells, quantified as TFUs by flow cytometry (Section 2.1), rather than equal numbers of culturable cells. This choice was made for conceptual simplicity, so that no strain was deliberately privileged or under-represented at the point of formulation. Because culturability—the ratio of culturable to total cells—is strain-specific, an input balanced on total cells does not yield an equal distribution of colony-forming units. The composition of the resulting blend, determined by plate count, was: *L. acidophilus* LA02 31 $\times$ 10^9 CFU/g (29% of total), *L. rhamnosus* LR04 17 $\times$ 10^9 CFU/g (16%), *L. fermentum* LF08 24 $\times$ 10^9 CFU/g (22%), and *B. animalis* subsp. *lactis* BS01 35 $\times$ 10^9 CFU/g (33%), totalling 107 $\times$ 10^9 CFU/g. The spread of these proportions is therefore an outcome of strain-specific culturability rather than a designed feature. It is in this sense representative of commercial multi-strain products, in which strain ratios expressed in CFU likewise reflect both the formulation and the differential culturability of the constituent strains. The blend underwent three independent experimental runs (R&S 1, 2, 3) according to the experimental plan summarised in Supplementary Figure S4, with normalisation of strain proportions across runs implemented in line with the MALDI-TOF MS-based enumeration approach previously developed for defined multi-strain bacterial consortia [8].

2.2. Enumeration Methods

2.2.1. Total Plate Count Enumeration

Total plate count determination of the four-strain blend was performed in three technical replicates (Figure 1, step 4). Serial tenfold dilutions were prepared in sterile phosphate-buffered saline (PBS, pH 7.2) (Gibco™ PBS, Thermo Fisher Scientific, Grand Island, NY, USA). From appropriate dilutions (10^{-9}–10^{-10}), 1.0 mL samples were plated on LAPTg agar supplemented with 0.1% (*w*/*v*) L-cysteine hydrochloride (hereafter LAPTg + CIST). LAPTg + CIST was prepared per litre of distilled water from: yeast extract 10.0 g, tryptone 10.0 g, peptone 15.0 g, glucose 10.0 g, Tween 80 1.0 mL, and agar 15.0 g; the medium was adjusted to pH 6.5 $\pm$ 0.2 and sterilised at 121 °C for 15 min. L-cysteine hydrochloride was prepared as a filter-sterilised 10% (*w*/*v*) stock and added aseptically to the molten medium cooled to approximately 50 °C to a final concentration of 0.1% (*w*/*v*) immediately before pouring. Plates were incubated at 37 °C for 72 h under anaerobic conditions, then refrigerated at 5 °C until further processing. Colonies used for downstream isolation and FTIR analysis were picked from the highest countable dilution yielding well-separated colonies within the 30–300 CFU/plate range (10^{-10} for total plate counts and 10^{-9} for selective plate counts), so that colonies originated from the most dilute plate on which discrete colonies could be reliably resolved.

To facilitate downstream identification, 15–30 representative colonies were randomly isolated from each count plate. Each colony was transferred into de Man, Rogosa, and Sharpe (MRS) broth (BD™ Difco™ MRS Broth, Becton, Dickinson and Company, Sparks, MD, USA) supplemented with 0.1% L-cysteine HCl and incubated at 37 °C for 18 h. A 2% (*v*/*v*) aliquot of each overnight culture was subsequently inoculated into fresh MRS broth and grown for an additional 18 h under identical conditions. These pure cultures were used for FTIR analysis and confirmatory PCR.

2.2.2. Selective Plate Count Enumeration

Selective plate count methods were employed to enumerate each strain individually from the four-strain mixture (Figure 1, step 4b). Three technical replicates were performed for each selective count following established ISO protocols: ISO 29981:2010/IDF 220:2010 [25] for bifidobacteria, ISO 7889:2003/IDF 117:2003 [26] for heterofermentative lactobacilli, and ISO 20128:2006/IDF 192:2006 [27] for homofermentative lactobacilli. Appropriate dilutions were plated on strain-specific selective agar media and incubated at 37 °C for 72 h.

B. animalis subsp. *lactis* (strain BS01) was enumerated by the pour plate technique on transgalactosylated oligosaccharide (TOS) propionate agar (Sharlau, Scharlab S.L., Sentmenat, Barcelona, Spain) supplemented with 0.05% L-cysteine HCl and mupirocin (50 µg/mL) (Scharlab S.L., Sentmenat, Barcelona, Spain) to selectively inhibit non-bifidobacteria. Plates were incubated anaerobically, and black, glossy colonies characteristic of bifidobacteria were counted.

L. rhamnosus (strain LR04) and *L. fermentum* (strain LF08) were enumerated on selective media targeting heterofermentative lactobacilli. Pour plates were performed on MRS agar containing vancomycin (4 µg/mL) (Sigma-Aldrich, Merck, Darmstadt, Germany). Additionally, an overlay technique on homofermentative–heterofermentative differential (HHD) agar supplemented with vancomycin was used to distinguish LF08 from LR04 by colony appearance: *L. fermentum* LF08 forms non-acidifying, opaque white colonies, whereas *L. rhamnosus* LR04 produces acidifying colonies that turn the medium yellow/green.

L. acidophilus (strain LA02) was enumerated by the pour plating on MRS agar (BD™ Difco™ MRS Broth, Becton, Dickinson and Company, Sparks, MD, USA) supplemented with clindamycin (0.1 µg/mL) and ciprofloxacin (10 µg/mL) (Sigma-Aldrich, Merck, Darmstadt, Germany), which inhibit other bacteria including other lactobacilli, allowing for the specific counting of *L. acidophilus* colonies.

2.2.3. Flow Cytometry Enumeration

Fluorescence-based flow cytometry was performed in parallel with plate counts to enumerate total cells and assess viability. A BD FACSLyric™ cytometer (Becton Dickinson Biosciences, San Jose, CA, USA) equipped with a 488 nm laser was used with volumetric counting beads as internal standards. The assay followed ISO 19344/IDF 232 (2015) [24] for differentiated live/dead bacterial counting.

Briefly, samples were diluted in buffered peptone water to approximately 10^5–10^6 cells/mL. Then, 835 µL of dilution was mixed with 50 µL fluorescent counting bead suspension (BD Trucount beads), 10 µL propidium iodide (PI, 0.2 mM), and 5 µL thiazole orange (TO, 42 µM). The mixture was incubated at 37 °C for 15 min. PI enters only membrane-compromised (dead) cells, while TO penetrates all cells but exhibits stronger fluorescence in live cells. Bacterial cells were gated on forward scatter versus side scatter plots, then differentiated on FL1 (TO) versus FL3 (PI) plots to distinguish live from dead populations.

2.3. FTIR Spectroscopy for Colony Identification

2.3.1. Sample Preparation

Isolated colony cultures were prepared for infrared spectral analysis using the IR Biotyper® system (IRBT; Bruker Daltonics GmbH & Co. KG, Bremen, Germany) following a modified protocol from Deidda et al. [7]. For each colony culture, a 5 mL aliquot of broth was centrifuged at 5000 rpm for 5 min. The supernatant was discarded, and the cell pellet was washed twice by resuspension in 1 mL PBS followed by centrifugation to remove media components. The final washed pellet was resuspended in 500 µL PBS.

Then, 15 µL of cell suspension was spotted onto a 96-well IRBT silicon sample plate in four technical replicate spots. Spots were air-dried at room temperature for 30–40 min until an even transparent film formed. Remaining cell pellets were stored at 5 °C for subsequent PCR verification.

2.3.2. Spectral Acquisition and Processing

FTIR spectra were acquired using an IR Biotyper in transmission mode over the wavenumber range 4000–500 cm^{-1}. The IRBT software (version 4.0) was used for instrument control and data processing. Each spectrum represented the average of 32 scans at 6 cm^{-1} resolution. Spectra were preprocessed by applying a smoothing filter, followed by calculation of the second derivative using the Savitzky–Golay algorithm (nine-point window). Spectra were then truncated to the fingerprint region (1300–800 cm^{-1}) and vector-normalised. Quality control standards (Bruker IRTS1 and IRTS2) were measured in each run to verify instrument performance. Only spectra passing the built-in quality criteria were used for analysis.

For full reproducibility, the instrument and assay configuration are reported as follows: IR Biotyper system (Bruker Daltonics, catalogue #1860001) operated with OPUS software (v8.5) and IR Biotyper analysis software (v4.0); strain reference library version 3.2.1, internally curated and validated against the four target strains. Across the three independent R&S runs, intra-run technical reproducibility (assessed on the four replicate spots per colony) showed coefficients of variation (CV) in spectral intensity ranging from 1.8% to 3.5% across strains; inter-run reproducibility (between R&S sessions, on QC standards IRTS1 and IRTS2) yielded CV values of 4.1% to 5.3%. The limit of detection (LoD) for reliable colony classification is constrained by the minimum biomass required for spectral acquisition (approximately 10^8 CFU equivalent per spot), with no upper limit of quantification (LLOQ/ULOQ) being relevant in this colony-counting application. All raw spectra (4000–500 cm^{-1}) and pre-processed datasets (1300–800 cm^{-1} fingerprint region) are provided in the Zenodo data deposit (see Data Availability Statement).

2.3.3. Exploratory Data Analysis

Initial exploratory analysis of the FTIR spectral data was performed to evaluate natural clustering of the four probiotic strains. The IR Biotyper software (v4.0) generated principal component analysis (PCA) plots combined with linear discriminant analysis (LDA) for dimensionality reduction and visualisation. Hierarchical cluster analysis (HCA) was also performed to assess spectral similarity. These unsupervised analyses confirmed spectral distinctiveness of the four strains prior to supervised classifier development.

2.3.4. Classifier Development

A supervised machine learning model was developed to automate strain identification from the FTIR spectra (Figure 1, step 2). The IR Biotyper software's artificial neural network (ANN) algorithm was employed to build a multi-class classifier. Input features consisted of the second-derivative, vector-normalised spectra described in Section 2.3.2, sampled at 501 points across the 1300–800 cm^{-1} fingerprint region, giving an input feature vector of 501 variables per spectrum. A training library of 128 spectra was compiled, comprising 32 spectra per strain. For each strain, this comprised 4 independent colony isolates × 2 culture matrices (MRS agar and MRS broth) × 4 replicate measurement spots per isolate-matrix combination (4 × 2 × 4 = 32). This training set captured both agar-grown and broth-grown phenotypes to ensure classifier robustness. All cultures were harvested at the same incubation time point (18 h) to standardise growth phase, as FTIR spectra are known to be sensitive to metabolic state. The ANN was trained over 100 iterative learning cycles, and the final model underwent internal validation via fourfold cross-validation. Ap-

plied to the 36 test colonies recovered from the blend, the classifier assigned all 36 correctly (100%; 95% Clopper–Pearson exact binomial confidence interval 90.3–100%). The ANN implementation is embedded within the proprietary IR Biotyper software (version 4.0). We report here every user-configurable parameter of the model (input feature count, spectral region, derivative and normalisation preprocessing, number of training cycles, and cross-validation scheme); however, the manufacturer does not disclose the number of hidden layers, the number of neurons per layer, the activation function or the training algorithm, and these parameters are neither visible nor modifiable by the user. Independent re-implementation of the identical classifier is therefore not currently possible, and we acknowledge this as a reproducibility limitation inherent to the use of a closed commercial platform. To mitigate this, the complete spectral distance matrix and the derived analysis tables are provided in the data deposit, so that the underlying spectral separation on which the classifier operates can be independently verified with open-source tools.

2.3.5. Colony Identification by the FTIR Classifier

FTIR spectra from unknown colonies were identified by applying the trained classifier (Figure 1, steps 6–7). The classifier outputs an identification along with a reliability score conveyed as a colour-coded indicator: green signifies high-confidence identification, yellow indicates lower confidence, and red indicates unreliable classification. Results were compared with expected identities from selective plate count and with PCR confirmation.

2.4. PCR Confirmation of Strain Identity

Species-specific PCR was employed to confirm FTIR classifications (Figure 1, step 8). Genomic DNA was extracted from stored cultures using Lysis Buffer FS (Applied Biosystems, Warrington, UK) according to the manufacturer's instructions. Four primer sets were used, each targeting unique genomic sequences: ACI/16SII for *L. acidophilus* (750 bp product; [28]), W3/Y2 for *L. rhamnosus* (290 bp; [29]), LF1/LF2 for *L. fermentum* (337 bp; [30]), and BifLac2/BifLac5 for *B. animalis* subsp. *lactis* (740 bp; [31]). PCR products were analysed by electrophoresis on 2% agarose gels stained with ethidium bromide.

2.5. Statistical Analysis

2.5.1. Enumeration Data Analysis

Statistical analyses were performed using R software (v4.3.3) for enumeration data and Python (v3.11; scipy v1.11, scikit-learn v1.3) for spectral cluster analysis. Because the four-strain blend was prepared as a single batch, the three experimental runs represent technical replicates that cannot be treated as independent biological observations. Accordingly, no inferential (hypothesis-testing) statistics are reported anywhere in this study, and all metrics below are presented as descriptive characterisations of the dataset obtained. Furthermore, individual technical replicates from selective plate counting and FTIR classification could not be paired across methods, as each method processed colonies from independent plates. Therefore, for each strain, the mean of three technical replicates was computed for both the observed (selective plate count) and estimated (FTIR-based) counts, expressed as $\log_{10}$ CFU/mL. Method agreement was assessed descriptively by computing the mean difference (bias) and standard deviation of differences between paired strain means (n = 4 strains). Pearson's correlation coefficient was computed to describe the linear association between methods, with a 95% confidence interval via Fisher z-transformation; with n = 4 strain means, this interval is necessarily wide, and the coefficient is reported as a descriptive index only. Ordinary least-squares linear regression was fitted (FTIR estimated vs. plate count observed) for visualisation. Bland–Altman analysis was used to visualise method agreement by plotting differences against averages of the two methods. We emphasise that with n = 4 paired strain means, the conventional 95% limits of agreement

(mean difference ± 1.96 × SD) are statistically unstable, because both the bias and the SD of differences are estimated from four values; the limits are therefore presented as an illustrative exercise to visualise the magnitude and direction of the observed differences, and are not interpreted as establishing method equivalence. A post hoc power analysis was performed to estimate the sample size required for a future, adequately powered agreement study. The 95% Clopper–Pearson exact binomial confidence interval was computed for the classifier accuracy.

For each strain, the observed (selective plate count) value reported in Table 1 was computed as $\log_{10}$ [mean (strain colony count/total selective colony count) × mean (LAPTg + CIST total count) × 10^9], where means were taken across the three R&S runs of the test blend. The estimated (FTIR) value was computed as the mean across runs of $\log_{10}$ [(IR-assigned strain count/total IR-classified colonies in run) × LAPTg + CIST total count in run × 10^9]. This normalisation scheme parallels the per-strain CFU back-calculation framework introduced by Coryell et al. [8] for defined multi-strain consortia analysed by MALDI-TOF MS, adapted here to FTIR-based colony classification. A reproducibility script (Python) implementing both calculations from the raw colony count data is provided in the data repository (see Data Availability Statement).

Table 1. Summary of the observed (selective plate count) versus FTIR estimated counts per strain (mean $\log_{10}$ CFU/mL ± standard deviation, n = 3 technical replicates per strain per method). Difference = Observed − Estimated.

Strain	Observed	Estimated	Difference
LR04	10.31 ± 0.04	10.34 ± 0.41	−0.03
LF08	10.41 ± 0.11	10.49 ± 0.37	−0.07
BS01	10.43 ± 0.07	10.42 ± 0.04	+0.01
LA02	10.40 ± 0.02	10.41 ± 0.15	−0.01

Values are mean (±SD) of $\log_{10}$ CFU/mL. SD computed from three technical replicates within a single blend preparation. As replicates are technical (not biological), no inferential statistics are reported; see Bland–Altman analysis (Supplementary Figure S1) for method agreement.

2.5.2. Cluster Validation Analysis

To characterise the spectral separation between the four strains, the IR Biotyper Euclidean distance matrix (160 spectra) was subjected to multiple complementary cluster analyses. The silhouette coefficient [32], Davies–Bouldin index [33] and cophenetic correlation were computed to quantify within- and between-cluster separation. Cluster stability was assessed by bootstrap resampling (200 iterations; [34]), reporting the maximum Jaccard similarity between the original and bootstrap clusters. The optimal number of clusters was independently determined by silhouette maximisation across k = 2–10. All of these metrics are descriptive indices computed on the fixed spectral dataset; because all spectra derive from a single blend preparation, permutation-based hypothesis tests of group separation are not applicable and are therefore not reported.

2.5.3. Figure Preparation and Reporting Standards

All main and supplementary figures were generated programmatically using matplotlib (v3.7) in Python (v3.11) and exported as 600 DPI TIFF files with LZW compression, alongside vector PDF and PNG companions, in compliance with the MDPI figure submission requirements (minimum 300 DPI for raster formats). Strain identity is encoded consistently throughout the manuscript by both colour and marker shape, drawn from the Okabe–Ito colourblind-safe palette: BS01 (blue #0072B2, circle), LR04 (green #009E73, square), LF08 (reddish-purple #CC79A7, triangle) and LA02 (orange #E69F00, diamond). This dual encoding ensures that the strains remain distinguishable for readers with colour

vision deficiencies, including under simulated deuteranopia and protanopia. Error bars represent ± 1 SD across the three R&S runs; sample sizes (N) and statistical conventions are indicated in each figure legend. All figures use sans-serif typography (Arial 9–11 pt) with consistent spacing, axis tick formatting, and legend placement to maintain a unified visual language across the manuscript.

Use of generative AI. Generative AI (Claude; Anthropic, San Francisco, CA, USA) was used as an assistive tool during manuscript preparation and was not involved in the generation, collection, or interpretation of the primary experimental data. Specifically, it was used to (i) draft and language-edit selected passages; (ii) assist in writing the Python code for the statistical re-analysis and reproducibility checks (cluster-validation metrics and Bland–Altman analysis), whose numerical outputs were independently verified by the authors against the source data; (iii) write the matplotlib code used to programmatically render figures from author-defined schematic content and author-provided experimental data—no diffusion- or GAN-based image-generation models were used; and (iv) format references and figure captions. All AI-assisted outputs, including figures and the outputs of analysis code, were reviewed, verified against the primary data, and edited by the authors, who take full responsibility for the integrity and content of the work.

3. Results

3.1. FTIR Exploratory Analysis of Strain Spectra

Visual inspection of the acquired spectra confirmed that the four strains presented distinct infrared fingerprints within the 1300–800 cm^{-1} region used for classification. Representative second-derivative spectra of the four reference strains are presented in Figure 2, each shown as four replicate acquisitions. Replicate spectra of the same strain were closely superimposable across the whole fingerprint region, whereas the four strains diverged at multiple wavenumbers. The most pronounced differences were located between 1200 and 900 cm^{-1}, the region dominated by C–O–C and C–O–P vibrations of cell-wall and cell-surface carbohydrates, consistent with the polysaccharide-composition differences discussed below.

Unsupervised analysis of the FTIR spectral data confirmed that the four probiotic strains possessed distinct infrared fingerprints. Principal coordinates analysis (PCoA) of the Euclidean distance matrix revealed four well-separated clusters in multivariate space, with PCoA 1 and PCoA 2 explaining 68.3% and 18.6% of total variance, respectively (cumulative: 86.9%; Figure 3A). Spectra from colonies isolated on plate count media aligned with the corresponding reference strain clusters used to train the classifier. The 95% confidence ellipses for the four strains did not overlap, demonstrating complete spectral separation. Hierarchical cluster analysis (HCA) confirmed these findings, with spectra grouping strictly by strain identity at the calculated cut-off (4.67) and a high cophenetic correlation coefficient (0.955; Figure 4). This exploratory result demonstrated that FTIR spectroscopy could readily differentiate the four strains based on their phenotypic spectral signatures.

Spectral discrimination among the four strains arises predominantly from differences in the carbohydrate fingerprint region (1200–900 cm^{-1}), where exopolysaccharides, teichoic acids, and cell-wall polysaccharides absorb strongly [22,23]. Secondary contributions from amide I and II vibrations (1700–1500 cm^{-1}), reflecting protein secondary structure, and from C–H stretching of membrane lipids (3000–2800 cm^{-1}) further enhance strain separation in linear discriminant space. The four strains used in this study span four different genera, and their distinct cell-wall architectures and surface polysaccharide profiles plausibly underlie the strong spectral separation observed.

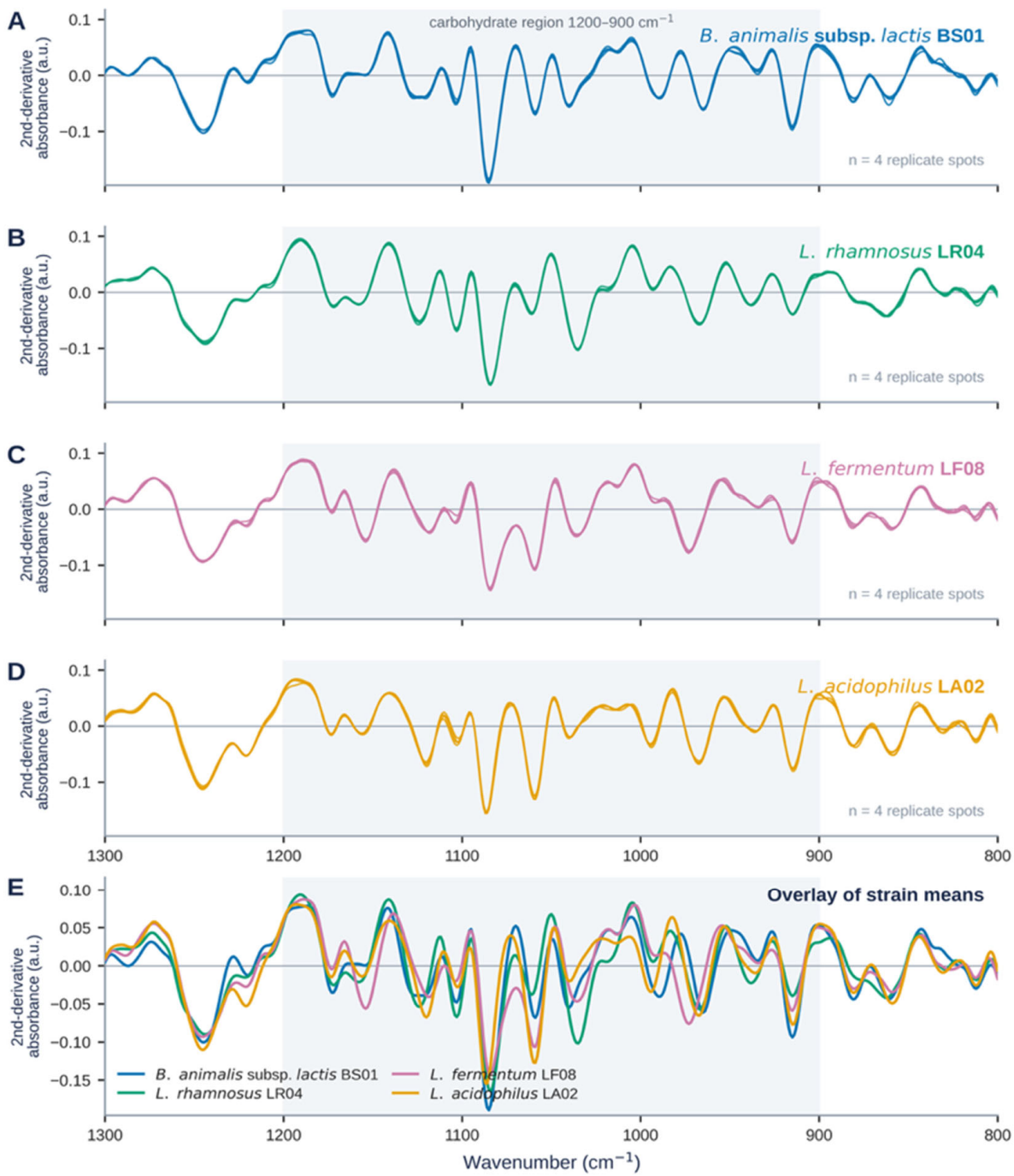

Figure 2. Representative second-derivative FTIR spectra of the four reference strains across the 1300–800 cm^{-1} fingerprint region. Panels are, from top to bottom: (**A**) *Bifidobacterium animalis* subsp. *lactis* BS01; (**B**) *Lacticaseibacillus rhamnosus* LR04; (**C**) *Limosilactobacillus fermentum* LF08; (**D**) *Lactobacillus acidophilus* LA02. In each of panels (**A**–**D**), the four replicate spot acquisitions of the reference isolate are overlaid. (**E**) Overlay of the four strain means. Spectra are second derivatives following vector normalisation, i.e., the same representation used as classifier input. The shaded band marks the 1200–900 cm^{-1} carbohydrate region. All spectra shown passed the IR Biotyper quality-control gate. Reference spectra were acquired in three independent sessions (LA02, 3 November 2023; LR04 and BS01, 9 November 2023; LF08, 15 November 2023).

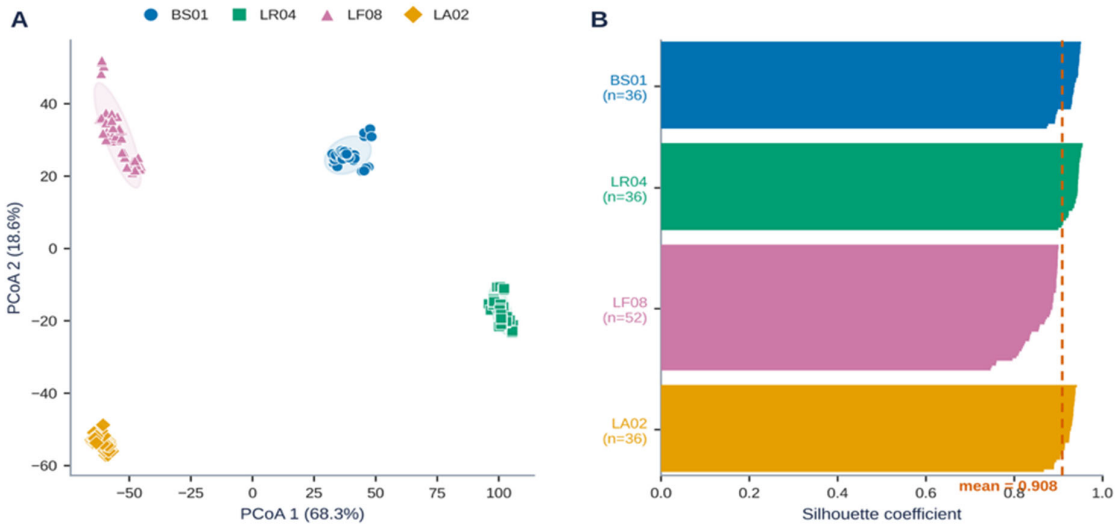

Figure 3. Multivariate visualisation of FTIR spectral separation between strains. (**A**) Principal coordinates analysis (PCoA) of the FTIR Euclidean distance matrix (n = 160 spectra) showing four spectrally distinct strain clusters: *B. animalis* subsp. *lactis* BS01 (blue, circle), *L. rhamnosus* LR04 (green, square), *L. fermentum* LF08 (reddish-purple, triangle), and *L. acidophilus* LA02 (orange, diamond). Strain identity is dual-encoded by colour and marker shape using the Okabe–Ito colourblind-safe palette. PCoA 1 and PCoA 2 explain 68.3% and 18.6% of total variance, respectively (cumulative: 86.9%). Ellipses denote 95% confidence regions. (**B**) Silhouette plot for the four-strain partition, with global mean silhouette = 0.908 (red dashed line). All strains achieved a silhouette $\geq$ 0.86, classified as a "strong cluster structure" [35].

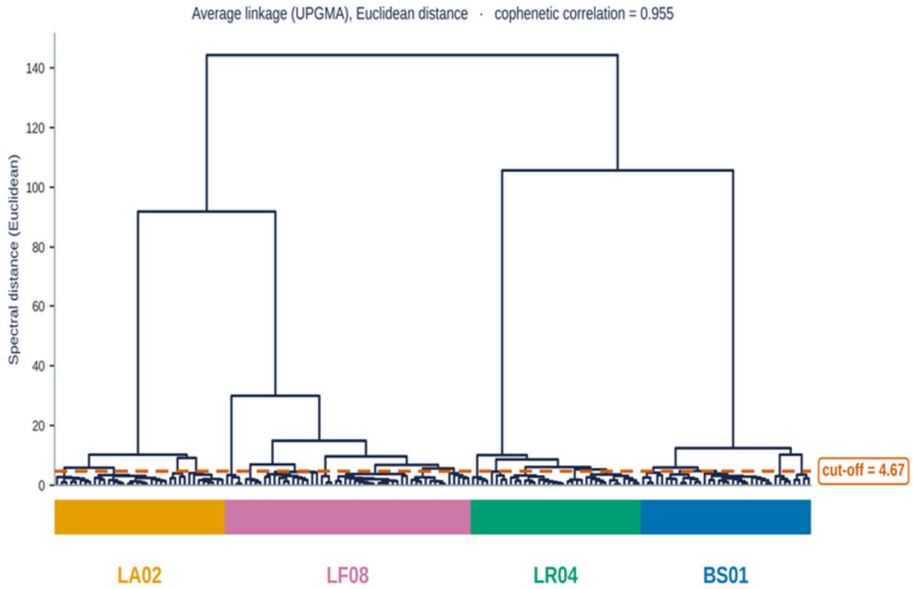

Figure 4. Hierarchical cluster analysis (HCA) of FTIR spectra (n = 160) using Euclidean distance and average linkage (UPGMA). Each leaf represents a single spectrum from one of 40 colony isolates (4 spectra per isolate). The coloured strip below the leaves indicates strain assignment. Red dashed line: software-calculated cut-off (4.67) optimising the SDI $\times$ mC criterion (0.919). Cophenetic correlation coefficient = 0.955, indicating excellent dendrogram fidelity to the original distance matrix. The four major branches correspond exactly to the four reference strains, providing data-driven confirmation of strain-level discrimination.

3.2. Statistical Validation of Strain Separation

To characterise the spectral separation observed in the exploratory analysis, multiple complementary cluster metrics were applied to the IR Biotyper Euclidean distance

matrix. The global silhouette coefficient was 0.908, well above the threshold for "strong cluster structure" (>0.7) [35]. Per-strain silhouette values ranged from 0.863 (LF08) to 0.939 (LR04), with all individual spectra above the strong-cluster threshold. The Davies–Bouldin index of 0.127 indicated excellent cluster compactness and separation (values < 0.5 are considered excellent). The optimal number of clusters, determined independently by silhouette maximisation across k = 2–10, was k = 4 (silhouette = 0.908; Supplementary Figure S2A), corresponding exactly to the four reference strains and indicating that the FTIR data structure of this dataset is organised into four phenotypic groups. Within this dataset, every within-strain spectral distance was smaller than every between-strain distance, and the visual block-diagonal structure of the pairwise distance matrix sorted by strain (Supplementary Figure S3) confirms that within-strain distances were systematically lower than between-strain distances; this complete separation is reported as a descriptive property of the data rather than as the outcome of a statistical test. Cluster stability under bootstrap resampling was maximal, with all four strain clusters recovered in every one of 200 bootstrap iterations (mean Jaccard similarity = 1.000; Supplementary Figure S2B), indicating that the observed partition is not driven by a small number of influential spectra. Together, these descriptive metrics indicate that in this dataset, the four-strain partition reflects intrinsic spectral structure rather than imposed labels, and provide the basis for the classifier described below. Because all spectra originated from a single blend preparation, these metrics characterise the present dataset and do not by themselves establish that comparable separation would be observed across independently prepared batches.

3.3. Colony Identification by the FTIR Classifier

The trained ANN classifier was applied to FTIR spectra from 36 individual colonies (12 colonies per each of 3 experimental runs) grown from the multi-strain blend. The classifier correctly identified 100% of the colonies to their strain of origin (36/36; 95% Clopper–Pearson CI: 90.3–100%). Per-class sensitivity and specificity were both 100% for all four strains. All 36 colony spectra received "green" high-confidence identifications for one of the four strains (LA02, LR04, LF08, or BS01), with no "yellow" intermediate scores and no "red" unreliable classifications. Figure 5 presents these results in confusion matrix format, demonstrating perfect classification accuracy with all colonies on the diagonal and zero misidentifications.

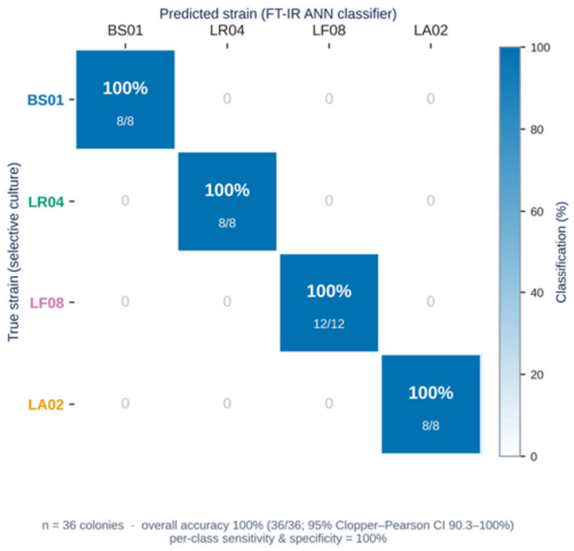

Figure 5. Confusion matrix of automated FTIR classifier results. All 36 colonies appear on the diagonal (100% accuracy; 95% Clopper–Pearson CI: 90.3–100%), with predicted identities matching true identities in every case. All classifications received high-confidence "green" scores.

3.4. Comparison of FTIR and Plate Count Results

Quantitative performance of FTIR-based identification was compared to conventional plate counting. For each strain, colony counts detected by selective plate count were compared to estimates from FTIR classifier assignments. Table 1 and Figure 6 present observed versus estimated counts ($\log_{10}$ CFU/mL) across three technical replicates. Mean counts obtained by FTIR closely matched the plate count results for all strains, with differences between method means ranging from 0.01 to 0.07 $\log_{10}$ units.

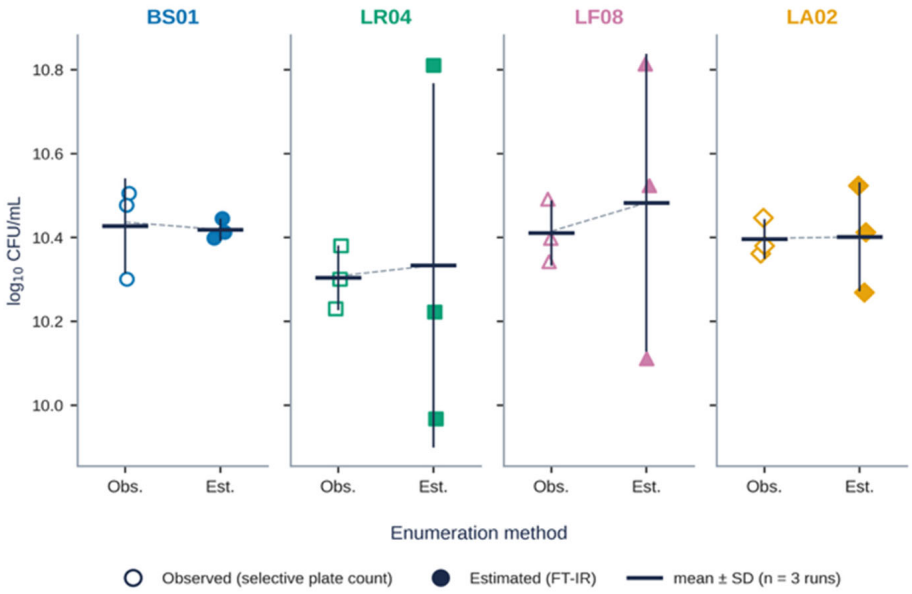

Figure 6. Comparison of observed (selective plate count) versus FTIR-estimated counts, expressed as $\log_{10}$ CFU/mL, for each strain (BS01, LA02, LF08, LR04). Individual points show the three technical replicates (R&S runs 1–3) for each method; horizontal bars indicate the mean and error bars ± 1 SD. Because only three technical replicates are available per strain, individual values with their mean and standard deviation are shown rather than box-and-whisker quartiles. Dotted lines indicate the theoretical expected values. The close overlap of observed and estimated values illustrates method agreement for this single blend preparation.

Correlation analysis showed close agreement between methods. Mean counts by FTIR and plate count were correlated (Pearson r = 0.78, 95% CI [−0.72, 1.00] via Fisher z-transformation, n = 4 strain means), as shown in Figure 7. The wide confidence interval, which spans zero, reflects the fact that the coefficient is computed from only four strain means; r is therefore reported as a descriptive index of association and not as evidence of a statistically established relationship. The linear regression slope approximated 1.0, and data points for all strains lay near the ideal 1:1 line. Bland–Altman analysis (Supplementary Figure S1), presented as an illustrative exercise given the four paired strain means available, indicated minimal systematic bias, with a mean difference of −0.028 $\log_{10}$ CFU/mL (SD = 0.039) and nominal 95% limits of agreement from −0.103 to +0.048 $\log_{10}$ units. These values are consistent with the close agreement between the two methods, of a magnitude smaller than the precision typically attributed to colony counting; however, with n = 4, the limits of agreement are statistically unstable and should not be read as establishing equivalence between the methods.

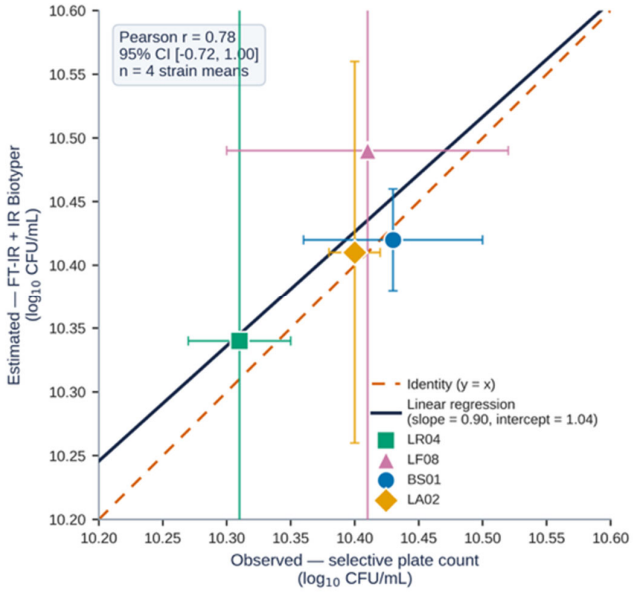

Figure 7. Correlation between FTIR enumeration and plate count. Scatter plot shows mean $\log_{10}$ CFU/mL per strain with error bars (±1 SD). Blue solid line: linear regression; red dashed line: identity line (y = x). Pearson r = 0.78 indicates strong concordance. Data points lie near the identity line, suggesting no systematic bias.

3.5. PCR Confirmation of Strain Identifications

FTIR classifier identifications were validated using species-specific PCR on 13 colony isolates from the first experimental run. PCR results showed 100% agreement with FTIR assignments (Figure 8). Each colony's DNA amplified only the expected strain-specific product with no cross-amplification. Specifically, colonies identified as *L. rhamnosus* yielded the 290 bp band, *L. fermentum* colonies the 337 bp band, *L. acidophilus* colonies the 750 bp fragment, and *B. lactis* colonies the 740 bp fragment. These confirmations substantiate that FTIR combined with ANN classification correctly identified each colony to the strain level.

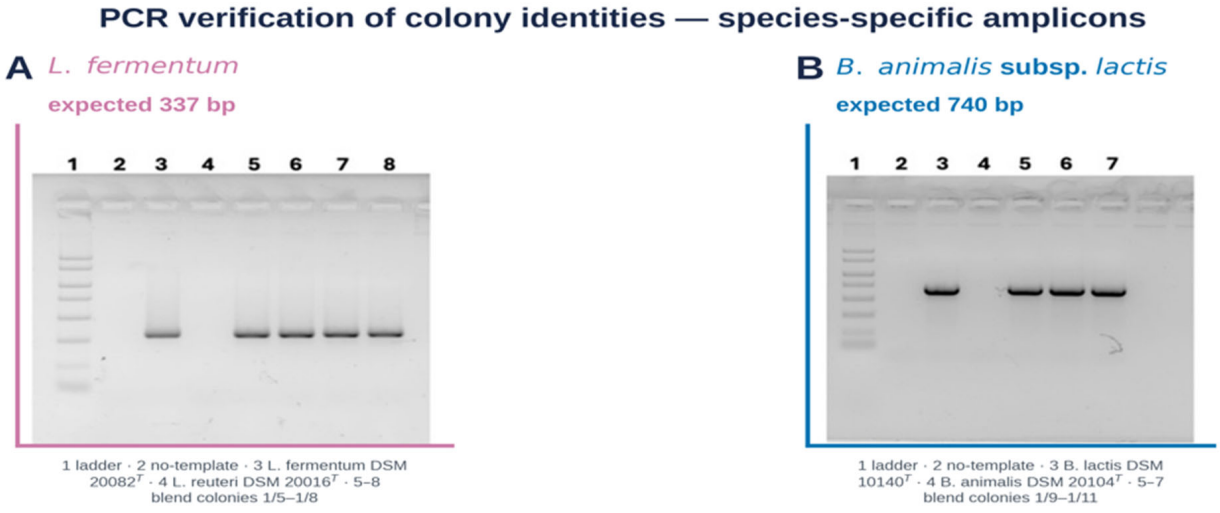

Figure 8. *Cont.*



Figure 8. PCR-based verification of colony identities. Agarose gel electrophoresis of species-specific amplicons: (**A**) *L. fermentum*, expected 337 bp; (**B**) *B. animalis* subsp. *lactis*, expected 740 bp; (**C**) *L. acidophilus*, expected 750 bp; (**D**) *L. rhamnosus*, expected 290 bp. In every panel, lane 1 is the 50–2000 bp ladder, lane 2 the no-template control, lanes 3 and 4 type-strain references, and the remaining lanes blend-derived colonies. All colonies produced expected amplicons matching the FTIR classifications, with no cross-amplification.

4. Discussion

FTIR spectroscopy is emerging as a powerful tool for bacterial identification and enumeration, offering an attractive phenotypic alternative to traditional DNA-based methods and MALDI-TOF MS [6,7]. In this proof-of-concept study, we applied FTIR to identify and count strains within a four-strain probiotic consortium. The results demonstrated that FTIR-based enumeration corresponds closely with conventional plate counting, with mean differences below 0.1 $\log_{10}$ CFU/mL per strain and minimal systematic bias in this single-batch experiment. It should be stated explicitly that the workflow is not a direct enumeration method. FTIR is not applied to the product itself: an initial plate-count and cultivation step is required to obtain isolated colonies, and the spectroscopic measurement is performed on the resulting isolates. The approach is therefore best described as a hybrid PC + FTIR workflow, in which plate count supplies viable, culturable counts and colony isolates, while FTIR supplies the strain-level assignment that plate count on a non-selective medium cannot provide. This architecture carries an intrinsic throughput constraint: total time-to-result remains bounded by the 72-h anaerobic incubation, and the number of colonies that can be characterised per run is bounded by the isolation, sub-culture, and measurement steps. Molecular methods such as qPCR or ddPCR, which can interrogate DNA extracted directly from the product matrix without cultivation, are consequently faster in time-to-result and are not subject to the culturability bias inherent to any cultivation-based workflow. The practical advantages we observed for the hybrid approach lie elsewhere: it reports culturable counts, which remain the regulatory unit of measure for probiotic products [3], while adding strain-level identification at low per-sample consumable cost and without strain-specific primer development.

The descriptive characterisation of strain separation observed here addresses three requirements for FTIR-based selective enumeration. First, the four strains are clearly separated in the spectral space of this dataset: the global silhouette of 0.908 places it in the "strong cluster structure" category [35], Davies–Bouldin (0.127) indicates excellent compactness, and all between-strain distances exceeded all within-strain distances. Second, this separation is internally consistent under resampling: bootstrap analysis (200 iterations)

recovered all four clusters in every iteration (Jaccard = 1.000), indicating that the partition is not an artefact of individual influential spectra, while the cophenetic correlation of 0.955 indicates that the hierarchical structure faithfully preserves the underlying distance relationships. Third, the ANN classifier's 100% accuracy (95% Clopper–Pearson CI: 90.3–100%) is coherent with this structure: when k = 4 emerges as the data-driven optimal partition by silhouette maximisation, and when the underlying distance structure already cleanly separates the four strains, high classification accuracy is the expected outcome of any reasonable supervised learning algorithm. This coherence should not, however, be read as excluding overfitting or as establishing generalisable accuracy. The classifier was trained and tested on material originating from the same batch, laboratory, and protocol, so the test colonies occupy the same phenotypic space as the training library; and the point estimate of 100% rests on 36 test colonies, with the lower confidence bound of 90.3% reflecting that limited test-set size. The honest interpretation is that for four spectrally distinct strains under controlled single-batch conditions, the approach performed without error; establishing whether that performance persists across independently produced batches, operators, and time points requires the external validation outlined below. These convergent descriptive results therefore support a proof-of-concept conclusion—that FTIR can in principle serve as a selective enumeration tool for multi-strain probiotic blends comprising spectrally distinct strains, provided that strain-specific reference libraries are properly trained—rather than a claim of method validation.

Unlike genotypic techniques such as pulsed-field gel electrophoresis, multilocus sequence typing, or whole-genome sequencing, FTIR spectroscopy captures a comprehensive phenotypic fingerprint of microorganisms [6]. By measuring infrared light absorption by various cellular components—lipids, carbohydrates, proteins, and nucleic acids—FTIR produces unique spectra reflecting overall biochemical composition [22,23]. This molecular fingerprinting capability enables discrimination at the species, subspecies, and strain levels with high specificity. For the probiotic industry, FTIR offers notable advantages: assays can be completed in hours without complex sample processing, instrumentation costs per sample are low once established, and the technique inherently captures phenotypic traits that purely genetic methods would miss [21].

The effectiveness of FTIR for fine microbial discrimination has been documented in several studies. Deidda et al. [6] successfully differentiated strains within the genus *Bifidobacterium*, achieving typing results comparable to or superior to PFGE and MLST. Li et al. [36] reported that FTIR matched or surpassed the discriminatory power of whole-genome sequencing for differentiating *Lactiplantibacillus plantarum* strains. Such strain-level discrimination is critical in probiotics, given that different strains of the same species can vary in stability, safety, and functional efficacy.

Several limitations of this proof-of-concept study should be acknowledged, and we note at the outset that the central ones are structural consequences of the study design rather than remediable analytical choices. First, and most importantly, evaluation was performed on a single four-strain blend preparation, with three technical replicates per method. Because the blend was prepared as a single batch, these replicates do not constitute independent biological observations. This is a major limitation: it means that no inferential statistic reported on these data would be valid, and accordingly none is reported; that between-batch variability—the dominant source of variance in industrial probiotic manufacture—was entirely unobserved in this study; and that the agreement metrics presented describe this particular preparation rather than the method in general. Method agreement was therefore assessed descriptively through correlation, Bland–Altman visualisation, and practical comparison of mean counts, with the limits of agreement explicitly presented as illustrative given $n = 4$ strain means. Furthermore, individual technical replicates from selective plate

counting and FTIR classification could not be paired across methods, as each method processed colonies from independent plates; therefore, strain means were compared rather than replicate-level observations. While this design is sufficient to demonstrate proof-of-concept feasibility, it is not sufficient to validate the method, and we make no such claim. Future studies must include multiple independently prepared blends (biological replicates) to enable formal statistical validation. A post hoc power analysis indicated that given the observed between-method variability (SD of differences = 0.039 log units), a study with $n = 6$ independently prepared blends would achieve 80% power to detect a mean bias of 0.05 log units at $\alpha = 0.05$; such a design ($n \geq 6$ blends $\times$ 4 strains, $n \geq 24$ paired observations) would also yield substantially tighter confidence intervals around the correlation coefficient, which here spans $[-0.72, 1.00]$. Second, the size and provenance of the reference library constrain the strength of the classification claim. The library comprised 128 spectra (32 per strain), which is modest by the standards of neural-network training. Two features of the design mitigate, but do not remove, this concern: the classification task is a four-class problem over strains that the unsupervised analysis shows to be already well-separated in spectral space, which is a substantially easier task than the many-class discrimination for which large libraries are typically required; and the library was structured rather than opportunistic, sampling 4 independent isolates $\times$ 2 culture matrices $\times$ 4 replicate spots per strain, so that isolate-level and matrix-level variability were both represented. Nonetheless, a library of this size samples only a narrow portion of the phenotypic variability that a production strain expresses across lots, media, and storage times, and expanding it is a prerequisite for routine deployment. Third, the ANN classifier was evaluated on colonies originating from the same laboratory, batch, and protocol used to train the reference library. While internal fourfold cross-validation confirmed model robustness, cross-validation performed within a single phenotypic space cannot detect the failure mode that matters most in practice, namely spectral drift between production lots. External validation on independently sourced colonies—from different production lots, different operators, and different time points—will therefore be required before deployment in routine quality control settings, both to detect batch-to-batch spectral drift and to establish the long-term stability of the reference library. Fourth, the selective media used in this study were adopted from established ISO and published protocols and were not themselves experimentally validated for specificity against this particular four-strain combination. The concordance we observed is indirect evidence that they behaved as intended: species-specific PCR on colony isolates from Run 1 agreed with the FTIR assignment in 13 out of 13 cases, and the selective counts summed consistently with the total counts. However, a formal specificity assessment—in which each strain is plated individually on each selective medium and its recovery quantified, together with cross-inhibition controls—was not performed here and should accompany any future validation study. Fifth, our consortium comprised four different genera deliberately selected for distinguishability using selective culture techniques, and the four strains were combined on an equal-total-cell basis (Section 2.2), which yielded culturable proportions spanning 16–33% of the total count. Both choices were made for conceptual clarity: the aim was to present the principle of pairing cultivation with spectroscopic identification against a reference method that could independently resolve the same strains, not to stress-test the approach. Consequently, the present data speak only to blends of this composition. Performance on strongly skewed blends—where a minor strain contributes a small percentage of the total count and may be under-sampled in a colony-picking scheme—is not addressed by these data and cannot be extrapolated from them; the number of colonies required to characterise a minor component with acceptable precision grows as its proportion falls, and this sampling relationship should be established explicitly in future work. Likewise, this design does not exploit FTIR's potential for differ-

entiating strains within the same species. Future work should challenge the method with more complex consortia containing multiple strains of identical or closely related species, and with deliberately unbalanced strain ratios—scenarios where FTIR has demonstrated capability in prior studies [7,36].

A final limitation concerns the reference method itself rather than the FTIR workflow. Plate count enumerates only those cells able to form colonies under the conditions provided, and lactic acid bacteria and bifidobacteria are well-documented to enter a viable but non-culturable (VBNC) state: a physiologically active condition in which cells retain membrane integrity, gene expression, and metabolic activity but no longer produce colonies on standard media [37,38]. This is not a marginal phenomenon for the organisms studied here. Loss of cultivability has been reported for *Bifidobacterium animalis* subsp. *lactis* and for several *Lactobacillus* sensu lato species under stresses directly relevant to probiotic manufacture and storage, including freeze-drying, low-temperature storage in fermented matrices, and acid exposure [37]. This study neither measured the non-culturable fraction nor was designed to, and we draw no inference about it from our data. The flow-cytometry values reported in Section 2.1 are total fluorescent units, acquired as a formulation input; because TFU counts every cell irrespective of membrane integrity, the arithmetic difference between TFU and CFU conflates dead cells, membrane-compromised cells, and any VBNC subpopulation, and cannot be used to quantify the last of these. Doing so would require the AFU determination (TO-positive/PI-negative events), from which the non-culturable viable fraction is conventionally estimated as AFU minus CFU [39]; this measurement was outside the scope of this work and would be a prerequisite for any claim about the viable population of these preparations.

Two considerations bound the impact of this limitation on the present study. First, this work did not set out to quantify the total viable population, and no claim to that effect is made: the question addressed is whether colonies recovered by plate count can be assigned to their strain of origin by a trained FTIR classifier, and every count reported here is explicitly a plate count. Second, and more importantly for the validity of the comparison, VBNC cells form no colonies and are therefore invisible to both arms of this study in equal measure. The reference method (selective plate count) and the test method (FTIR classification of isolates recovered from the total plate count) both sampled the same population of colonies that had grown; a VBNC subpopulation, however large, was absent from both and cannot bias the agreement between them. What it does bound is the external interpretation of the counts: neither method reports the viable load of the product, and neither should be read as doing so. Any future extension of this workflow toward a viability-inclusive claim would need to be anchored to a cultivation-independent method such as flow cytometry, with the CFU-to-AFU relationship established empirically for each strain rather than assumed [3,39].

Another consideration for broader implementation involves the requirement for reference spectral libraries or machine learning models tailored to specific probiotic blends. We envision that for each unique strain combination, an FTIR classifier could be trained and validated during product development and quality control. Establishing a repository of FTIR spectral profiles for probiotic strains—potentially maintained by an independent body—could facilitate method standardisation and enable laboratories worldwide to verify product composition by matching spectra to known fingerprints.

5. Conclusions

This proof-of-concept study demonstrates that a four-strain probiotic blend comprising different genera can be enumerated and identified using FTIR spectroscopy with results closely comparable to conventional selective plate counting. The integration of FTIR with

plate count methods offers a novel workflow whereby colonies can be counted on agar and rapidly identified by their infrared spectral signature. This approach is efficient—providing both count and identity in a single streamlined procedure—and cost-effective once the spectral classifier is established. With further validation across additional strains and products, FTIR spectroscopy paired with machine learning has the potential to become a valuable asset in the quality control of multi-strain probiotic and live biotherapeutic products, ensuring label compliance and strain-specific efficacy assessment.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/spectroscj4030013/s1, Figure S1: Bland–Altman plot of agreement between FTIR and selective plate count enumeration; Figure S2: Extended cluster analysis diagnostics (silhouette across k, bootstrap cluster stability); Figure S3: Full distance matrix heatmap (n = 160 spectra) sorted by strain; Figure S4: Experimental plan diagram for the test blend including serial dilution scheme and selective vs. total plating workflow.

Author Contributions: Conceptualisation, F.D., M.C. (Miriam Cordovana) and M.P.; methodology, F.D., M.C. (Miriam Cordovana) and M.B.; software, M.C. (Matteo Calgaro) and N.V.; validation, M.C. (Matteo Calgaro), N.V. and S.A.; formal analysis, M.C. (Matteo Calgaro) and N.V.; investigation, F.D., C.M., S.A. and M.B.; resources, M.C. (Miriam Cordovana) and M.P.; data curation, C.M. and S.A.; writing—original draft preparation, F.D.; writing—review and editing, M.C. (Miriam Cordovana), M.P. and M.C. (Matteo Calgaro); visualisation, F.D. and M.C. (Matteo Calgaro); supervision, N.V. and M.P.; project administration, M.P.; funding acquisition, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by Probiotical Research Srl.

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals.

Informed Consent Statement: Not applicable. This study did not involve humans.

Data Availability Statement: All raw and processed data supporting the conclusions of this article are openly available. The raw FTIR spectra (n = 160; 4000–500 cm^{-1}), pre-processed datasets (1300–800 cm^{-1} fingerprint region), distance matrices, PCoA coordinates, silhouette values, bootstrap iterations, plate count records, and IR Biotyper classification outputs can be obtained with written request from the corresponding author.

Acknowledgments: The authors thank Probiotical SpA for providing the lyophilised probiotic strain stocks used in this study, and Bruker Daltonics for technical support with the IR Biotyper system. During the preparation of this manuscript, the authors used Claude (Anthropic, San Francisco, CA, USA; models Opus 4.6/4.7) for the purposes of (i) assisting with the statistical re-analysis pipeline for cluster validation and Bland–Altman reproducibility, (ii) drafting and editing of the methods and statistical-analysis subsections, and (iii) drafting reference and figure-caption formatting. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: F.D., C.M., S.A., M.B. and M.P. are employees of Probiotical Research Srl. M.Co. is an employee of Bruker Daltonics GmbH. M.Ca. and N.V. declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
CFUs	Colony-Forming Units
CI	Confidence Interval
CV	Coefficient of Variation

ddPCR	Digital Droplet PCR
FT IR	Fourier Transform Infrared (spectroscopy)
HCA	Hierarchical Cluster Analysis
ISAPP	International Scientific Association for Probiotics and Prebiotics
LBP	Live Biotherapeutic Product
LDA	Linear Discriminant Analysis
LoA	Limits of Agreement
LoD	Limit of Detection
MALDI-TOF MS	Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry
PBS	Phosphate-Buffered Saline
PC	Plate Count
PCA	Principal Component Analysis
PCoA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
qPCR	Quantitative PCR
R&S	Repeat Sampling (experimental session replicates)
SD	Standard Deviation
TFUs	Total Fluorescent Units

References

1. FAO/WHO. *Guidelines for the Evaluation of Probiotics in Food*; Report of a Joint FAO/WHO Working Group: London, ON, Canada, 2002.
2. Hill, C.; Guarner, F.; Reid, G.; Gibson, G.R.; Merenstein, D.J.; Pot, B.; Morelli, L.; Canani, R.B.; Flint, H.J.; Salminen, S.; et al. Expert consensus document: The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat. Rev. Gastroenterol. Hepatol.* **2014**, *11*, 506–514. [CrossRef] [PubMed]
3. Boyte, M.-E.; Benkowski, A.; Pane, M.; Shehata, H.R. Probiotic and postbiotic analytical methods: A perspective of available enumeration techniques. *Front. Microbiol.* **2023**, *14*, 1304621. [CrossRef]
4. Liu, J.; Yang, L.; Kjellerup, B.V.; Xu, Z. Viable but Nonculturable (VBNC) State, an Underestimated and Controversial Microbial Survival Strategy. *Trends Microbiol.* **2023**, *31*, 1013–1023. [CrossRef]
5. Song, S.; Wood, T.K. 'Viable but Non-Culturable Cells' Are Dead. *Environ. Microbiol.* **2021**, *23*, 2335–2338. [CrossRef] [PubMed]
6. Deidda, F.; Bozzi Cionci, N.; Cordovana, M.; Campedelli, I.; Fracchetti, F.; Di Gioia, D.; Ambretti, S.; Pane, M. Bifidobacteria strain typing by Fourier transform infrared spectroscopy. *Front. Microbiol.* **2021**, *12*, 692975. [CrossRef] [PubMed]
7. Deidda, F.; Cordovana, M.; Bozzi Cionci, N.; Graziano, T.; Di Gioia, D.; Pane, M. In-process real-time probiotic phenotypic strain identity tracking: The use of Fourier-transform infrared spectroscopy. *Front. Microbiol.* **2022**, *13*, 1052420. [CrossRef] [PubMed]
8. Coryell, M.P.; Sava, R.L.; Hastie, J.L.; Carlson, P.E., Jr. Application of MALDI-TOF MS for enumerating bacterial constituents of defined consortia. *Appl. Microbiol. Biotechnol.* **2023**, *107*, 4069–4077. [CrossRef] [PubMed]
9. Scheier, T.C.; Franz, J.; Boumasmoud, M.; Andreoni, F.; Chakrakodi, B.; Duvnjak, B.; Egli, A.; Zingg, W.; Ramette, A.; Wolfensberger, A.; et al. Fourier-transform infrared spectroscopy for typing of vancomycin-resistant *Enterococcus faecium*: Performance analysis and outbreak investigation. *Microbiol. Spectr.* **2023**, *11*, e00984-23. [CrossRef] [PubMed]
10. Kon, H.; Temkin, E.; Elmalih, P.; Keren-Paz, A.; Ben-David, D.; Najjar-Debbiny, R.; Gottesman, T.; Carmeli, Y. Analysis of four carbapenem-resistant *Acinetobacter baumannii* outbreaks using Fourier-transform infrared spectroscopy. *Infect. Control Hosp. Epidemiol.* **2023**, *44*, 991–993. [CrossRef] [PubMed]
11. Hu, Y.; Zhou, H.; Lu, J.; Sun, Q.; Liu, C.; Zeng, Y.; Zhang, R. Evaluation of the IR Biotyper for *Klebsiella pneumoniae* typing and its potentials in hospital hygiene management. *Microb. Biotechnol.* **2021**, *14*, 1343–1352. [CrossRef] [PubMed]
12. Martak, D.; Valot, B.; Sauget, M.; Cholley, P.; Thouverez, M.; Bertrand, X.; Hocquet, D. Fourier-transform infrared spectroscopy can quickly type Gram-negative bacilli responsible for hospital outbreaks. *Front. Microbiol.* **2019**, *10*, 1440. [CrossRef] [PubMed]
13. Savini, F.; Romano, A.; Giacometti, F.; Indio, V.; Pitti, M.; Decastelli, L.; Devalle, P.L.; Gorrasi, I.S.R.; Miaglia, S.; Serraino, A. Investigation of a *Staphylococcus aureus* sequence type 72 food poisoning outbreak associated with food-handler contamination in Italy. *Zoonoses Public Health* **2023**, *70*, 411–419. [CrossRef] [PubMed]
14. Oberreuter, H.; Dyk, M.; Rau, J. Validated differentiation of *Listeria monocytogenes* serogroups by FT-IR spectroscopy using an artificial neural network-based classifier in an accredited official food control laboratory. *Clin. Spectrosc.* **2023**, *5*, 100030. [CrossRef]

15. Radmehr, B.; Zaferanloo, B.; Tran, T.; Beale, D.J.; Palombo, E.A. Prevalence and characteristics of *Bacillus cereus* group isolated from raw and pasteurised milk. *Curr. Microbiol.* **2020**, *77*, 3065–3075. [CrossRef] [PubMed]
16. Zendri, F.; Schmidt, V.; Mauder, N.; Loeffler, A.; Jepson, R.E.; Isgren, C.; Pinchbeck, G.; Haldenby, S.; Timofte, D. Rapid typing of *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* by Fourier-transform Infrared spectroscopy informs infection control in veterinary settings. *Front. Microbiol.* **2024**, *15*, 1334268. [CrossRef] [PubMed]
17. Zendri, F.; Schmidt, V.; Mauder, N.; Loeffler, A.; Jepson, R.E.; Isgren, C.; Pinchbeck, G.; Haldenby, S.; Timofte, D. Corrigendum: Rapid typing of *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* by Fourier-transform Infrared spectroscopy informs infection control in veterinary settings. *Front. Microbiol.* **2024**, *15*, 1396367. [CrossRef] [PubMed]
18. Keinprecht, H.; Irimaso, E.; Cabal Rosel, A.; Stessl, B.; Ntakirutimana, C.; Marek, L.; Fischer, O.W.; Szostak, M.P.; Zöchbauer, J.; Wittek, T.; et al. Diversity of *Staphylococcus aureus* associated with mastitis from dairy cows in Rwanda. *J. Glob. Antimicrob. Resist.* **2024**, *36*, 326–335. [CrossRef] [PubMed]
19. Tata, A.; Marzoli, F.; Cordovana, M.; Tiengo, A.; Zacometti, C.; Massaro, A.; Barco, L.; Belluco, S.; Piro, R. A multi-center validation study on the discrimination of *Legionella pneumophila* sg.1, *Legionella pneumophila* sg. 2-15 and *Legionella* non-*pneumophila* isolates from water by FT-IR spectroscopy. *Front. Microbiol.* **2023**, *14*, 1150942. [CrossRef] [PubMed]
20. Pascale, M.R.; Bisognin, F.; Mazzotta, M.; Girolamini, L.; Marino, F.; Dal Monte, P.; Cordovana, M.; Scaturro, M.; Ricci, M.L.; Cristino, S. Use of Fourier-transform infrared spectroscopy with IR Biotyper® system for *Legionella pneumophila* serogroups identification. *Front. Microbiol.* **2022**, *13*, 866426. [CrossRef] [PubMed]
21. Baker, M.J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H.J.; Dorling, K.M.; Fielden, P.R.; Fogarty, S.W.; Fullwood, N.J.; Heys, K.A.; et al. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **2014**, *9*, 1771–1791. [CrossRef] [PubMed]
22. Naumann, D.; Helm, D.; Labischinski, H. Microbiological characterizations by FT-IR spectroscopy. *Nature* **1991**, *351*, 81–82. [CrossRef] [PubMed]
23. Lasch, P.; Naumann, D. Infrared spectroscopy in microbiology. In *Encyclopedia of Analytical Chemistry*; Meyers, R.A., Ed.; John Wiley & Sons: Chichester, UK, 2015; pp. 1–32.
24. *ISO 19344:2015/IDF 232:2015*; Milk and Milk Products—Starter Cultures, Probiotics and Fermented Products—Quantification of Lactic Acid Bacteria by Flow Cytometry. International Organization for Standardization: Geneva, Switzerland, 2015.
25. *ISO 29981:2010/IDF 220:2010*; Milk Products—Enumeration of Presumptive Bifidobacteria—Colony Count Technique at 37 °C. International Organization for Standardization: Geneva, Switzerland, 2010.
26. *ISO 7889:2003/IDF 117:2003*; Yogurt—Enumeration of Characteristic Microorganisms—Colony-Count Technique at 37 °C. International Organization for Standardization: Geneva, Switzerland, 2003.
27. *ISO 20128:2006/IDF 192:2006*; Milk Products—Enumeration of Presumptive *Lactobacillus acidophilus* on a Selective Medium—Colony-Count Technique at 37 °C. International Organization for Standardization: Geneva, Switzerland, 2006.
28. Walter, J.; Tannock, G.W.; Tilsala-Timisjarvi, A.; Rodtong, S.; Loach, D.M.; Munro, K.; Alatossava, T. Detection and identification of gastrointestinal *Lactobacillus* species by using denaturing gradient gel electrophoresis and species-specific PCR primers. *Appl. Environ. Microbiol.* **2000**, *66*, 297–303. [CrossRef] [PubMed]
29. Aureli, P.; Fiore, A.; Scalfaro, C.; Franciosa, G. Metodi microbiologici tradizionali e metodi molecolari per l'analisi degli integratori alimentari a base di o con probiotici per uso umano. *Rapp. Istisan* **2008**, *8*, 2.
30. Dickson, E.M.; Riggio, M.P.; Macpherson, L.M. A novel species-specific PCR assay for identifying *Lactobacillus* fermentum. *J. Med. Microbiol.* **2005**, *54*, 299–303. [CrossRef] [PubMed]
31. Ventura, M.; Reniero, R.; Zink, R. Specific identification and targeted characterization of *Bifidobacterium* lactis from different environmental isolates by a combined multiplex-PCR approach. *Appl. Environ. Microbiol.* **2001**, *67*, 2760–2765. [CrossRef] [PubMed]
32. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
33. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
34. Hennig, C. Cluster-wise assessment of cluster stability. *Comput. Stat. Data Anal.* **2007**, *52*, 258–271. [CrossRef]
35. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: New York, NY, USA, 1990. [CrossRef]
36. Li, X.; Zhu, L.; Wang, X.; Li, J.; Tang, B. Evaluation of IR Biotyper for *Lactiplantibacillus* plantarum typing and its application potential in probiotic preliminary screening. *Front. Microbiol.* **2022**, *13*, 823120. [CrossRef] [PubMed]
37. Pazos-Rojas, L.A.; Cuellar-Sánchez, A.; Romero-Cerón, A.L.; Rivera-Urbalejo, A.; Van Dillewijn, P.; Luna-Vital, D.A.; Muñoz-Rojas, J.; Morales-García, Y.E.; Bustillos-Cristales, M.d.R. The viable but non-culturable (VBNC) state, a poorly explored aspect of beneficial bacteria. *Microorganisms* **2024**, *12*, 39. [CrossRef] [PubMed]

38. Oliver, J.D. The viable but nonculturable state in bacteria. *J. Microbiol.* **2005**, *43*, 93–100. [PubMed]
39. Tracey, H.; Coates, N.; Hulme, E.; John, D.; Michael, D.R.; Plummer, S.F. Insights into the enumeration of mixtures of probiotic bacteria by flow cytometry. *BMC Microbiol.* **2023**, *23*, 48. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.