



# Comparison of methods for the identification and sub-typing of O157 and non-O157 *Escherichia coli* serotypes and their integration into a polyphasic taxonomy approach

M.A. Prieto-Calvo<sup>1,2</sup>, M.K. Omer<sup>3</sup>, O. Alvseike<sup>3</sup>, M. López<sup>1,2</sup>, A. Alvarez-Ordóñez<sup>1,2†</sup>, M. Prieto<sup>1,2</sup>

<sup>1</sup>Department of Food Hygiene and Food Technology, Faculty of Veterinary Medicine, University of León, León, Spain

<sup>2</sup>Institute of Food Science and Technology, University of León, León, Spain

<sup>3</sup>Animalia – Norwegian Meat and Poultry Research Center, Oslo, Norway

## Abstract

Phenotypic, chemotaxonomic and genotypic data from 12 strains of *Escherichia coli* were collected, including carbon source utilisation profiles, ribotypes, sequencing data of the 16S–23S rRNA internal transcribed region (ITS) and Fourier transform-infrared (FT-IR) spectroscopic profiles. The objectives were to compare several identification systems for *E. coli* and to develop and test a polyphasic taxonomic approach using the four methodologies combined for the sub-typing of O157 and non-O157 *E. coli*. The nucleotide sequences of the 16S–23S rRNA ITS regions were amplified by polymerase chain reaction (PCR), sequenced and compared with reference data available at the GenBank database using the Basic Local Alignment Search Tool (BLAST). Additional information comprising the utilisation of carbon sources, riboprint profiles and FT-IR spectra was also collected. The capacity of the methods for the identification and typing of *E. coli* to species and subspecies levels was evaluated. Data were transformed and integrated to present polyphasic hierarchical clusters and relationships. The study reports the use of an integrated scheme comprising phenotypic, chemotaxonomic and genotypic information (carbon source profile, sequencing of the 16S–23S rRNA ITS, ribotyping and FT-IR spectroscopy) for a more precise characterisation and identification of *E. coli*. The results showed that identification of *E. coli* strains by each individual method was limited mainly by the extension and quality of reference databases. On the contrary, the polyphasic approach, whereby heterogeneous taxonomic data were combined and weighted, improved the identification results, gave more consistency to the final clustering and provided additional information on the taxonomic structure and phenotypic behaviour of strains, as shown by the close clustering of strains with similar stress resistance patterns.

## Keywords

FT-IR spectroscopy • genotyping • ITS sequencing • polyphasic taxonomy • VTEC

## Introduction

Bacterial taxonomy generally makes use of an exhaustive set of genotypic and phenotypic data obtained from isolates (Vandamme *et al.*, 1996; Vandamme and Peeters, 2014). Taxonomical studies use a polyphasic approach combining phenotypic and genotypic characteristics, but this strategy is barely used in identification or typing schemes, in which strains are normally studied using a single method. Combined use of these types of data can contribute to improving identification systems and obtaining a more complete picture of the taxonomic relationships among isolates. Studies of isolates using a set of simultaneous or successive procedures should be designed to yield complementary genotypic and phenotypic data of high quality. Heterogeneous data can be then integrated in a consensus type of classification or polyphasic taxonomical scheme that considers all available phenotypic, chemotaxonomic and genotypic

information. Considering the large amount of quality data of complementary nature generated by current phenotypic, chemotaxonomic and genotypic systems, combination and integration of all methodologies into a polyphasic scheme will probably constitute a precise typing approach that allows better understanding of population structure and taxonomic relationships.

Verocytotoxigenic *Escherichia coli* (VTEC) are zoonotic agents characterised by the production of Shiga-like toxins commonly associated with foodborne disease episodes that can lead to severe health complications and, sometimes, death. Correct VTEC classification and sub-typing is crucial for regulatory authorities and food business operators because false positive claims might be detrimental to business and affect food safety assurance systems. Although the majority of reported VTEC cases have been linked with strains of serotype O157, other serotypes, such as O45, O26, O91,

<sup>†</sup>Corresponding author: Avelino Alvarez-Ordóñez

E-mail: aalvo@unileon.es

O103, O111, O121 and O145, are emerging as causative agents of foodborne disease (EFSA and ECDC, 2015).

Phenotypic identification systems are still routinely used in clinical and food microbiology settings even though their limitations are well known (Kootallur *et al.*, 2011). Biolog Microstation™ ID System (Biolog, Hayward, CA, USA) is an automated system that tests the ability of bacteria to utilise 95 various carbon sources and assigns unknown strains to a species by mathematically comparing their utilisation pattern (metabolic fingerprint) with the Biolog database (Odumeru *et al.*, 1999; Kootallur *et al.*, 2011).

Fourier transform-infrared (FT-IR) spectroscopy is considered a chemotaxonomic method that is able to provide a great amount of taxonomic data. Spectra obtained from cellular biomass are very characteristic and can be considered as 'fingerprints' or molecular patterns distinctive of a particular bacterial strain (Naumann, 2000; Alvarez-Ordóñez *et al.*, 2011). Because infrared spectra reflect the global chemical composition of the samples, FT-IR spectroscopy provides information on strain-related differences in chemical composition and, therefore, it can be used in classification and identification.

Ribotyping generates a banding pattern that can be used to differentiate closely related bacteria based on differences in rRNA genes (Grimont and Grimont, 1986). The bar code-like pattern obtained can be compared with patterns already included in a database, for identification and typing purposes, based on similarities in the position and the size of the probed fragments.

Analysis of the internal transcribed spacer (ITS) region of the 16S–23S rRNA genes offers a partial view of relationships between different taxonomic units. Sequence polymorphisms in the ITS region are used for species and subspecies differentiation of bacteria and yeasts (Gurtler and Stanisich, 1996; Guasp *et al.*, 2000; Motoyama and Ogata, 2000; Schoch *et al.*, 2012). The differences in the sequence of the spacer region can be used as a molecular tool for species identification.

This work was undertaken to develop and test a polyphasic approach using all these characterisation systems, encompassing genotypic, phenotypic and chemotaxonomic information. This combination of methods should supply extra information on the behaviour of strains in foods and the environment while revealing the taxonomic structure and the underlying phylogenetic relationships between strains. For this purpose, 12 VTEC strains from various sources, including a diversity of serotypes, were comprehensively characterised using carbon source utilisation, ribotyping, sequencing of the ITS region of the 16S–23S rRNA genes and FT-IR spectroscopy.

## Materials and methods

### Strains

The bacterial strains used in this study are shown in Table 1. In total, 12 strains of *E. coli* were used, of which 10 strains were clinical strains associated with outbreaks or individual clinical cases and two strains were non-pathogenic. The cultures were maintained in cryovials at –20 °C. The bacteria were resuscitated in tubes containing 10 mL of brain heart infusion (BHI) (Oxoid, Basingstoke, UK) by incubation at 37 °C for 24 h, followed by streaking on BHI agar plates, which were incubated under the same conditions. Stationary-phase inocula were prepared by inoculating 10 mL of fresh BHI with an isolated colony and incubating it overnight at 37 °C. This suspension was then used to inoculate 50 mL of sterile BHI with approximately 10<sup>3</sup> cells/mL, followed by incubation at 37 °C for 24 h, which resulted in a stationary phase culture with approximately 10<sup>9</sup> cells/mL.

### Carbon source utilisation profiles

Metabolic fingerprints were obtained with the Biolog Microstation ID System (Biolog, USA) according to the manufacturer's instructions. Bacteria grown as previously described were streaked onto BHI agar plates (37 °C for 24 h). Bacterial suspensions were prepared by removing bacterial colonies from the plate surface with a sterile cotton swab, followed by agitation in 5 mL of 0.85% saline. The suspension was adjusted to 52% transmittance at 590 nm (Beckman DU 7400 spectrophotometer; Beckman Coulter, Inc., Brea, CA, USA). An aliquot (150 µL) of this suspension was dispensed into each well of a Biolog GN2 plate. The plates were incubated for 6–24 h at 35°C and then read with the Biolog microplate reader using the Biolog Microstation 3.5 software. Biolog software was used to compare the results obtained with the tested strains against the software database, which allowed bacterial identification based on distance calculation. To create a distance matrix, correlation (Pearson coefficient *r*) between strains was calculated, and the formula 1 – *r* was applied. The MultiVariate Statistical Package (MVSP) software, version 3.13f (Kovach Computing Services, Anglesey, UK), was used for calculations.

### Ribotyping

Isolated colonies from BHI agar plates (grown for 18–20 h at 35 °C) were used for ribotyping. Lysis of cells with release of DNA, EcoRI digestion of the chromosomal DNA, separation of the fragments by agarose gel electrophoresis and hybridisation with a chemiluminescent agent-labelled

DNA probe containing the *E. coli* ribosomal RNA operon were carried out with an automated RiboPrinter® Microbial Characterization System (Dupont Qualicon; Wilmington, DE, USA). The position and relative intensities of the rRNA operon-specific DNA fragments were estimated automatically by the RiboPrinter® analysis software, and a digital record for each sample was obtained.

#### Identification process

Banding treatment was done using RiboPrinter software, which normalises the digitised patterns. Band analysis was performed after an automatic band search by using the following parameters: minimum profiling parameter of 5.0% and grey-zone parameter of 5.0% relative to the maximal value, minimum area parameter of zero and shoulder sensitivity parameter of zero. The RiboPrinter Microbial Characterization System database, containing more than 8,500 RiboPrint™ patterns, was used for assignment of isolates. Identification considered a threshold value of 93% overall similarity to pool patterns into a single ribogroup.

#### Cluster analysis

Normalised patterns were exported, and a distance matrix between strains using the formula  $1 - r$  ( $r$ , Pearson coefficient) was calculated. Clustering was performed using the Ward algorithm method based on the Pearson correlation (Statistica for Windows, version 7.0; Statsoft Inc., Tulsa, OK, USA).

#### Analysis by FT-IR spectroscopy

Cultures were grown aerobically in BHI broth (Oxoid) at 37 °C for 24 h. An inoculum was then transferred with a sterile platinum loop from the broths to BHI agar (37 °C, 24 h). For each strain, at least two replicates were grown and processed in independent assays. Cells (~10–60 mg dry weight) were removed with a platinum loop and suspended in 100 mL distilled sterile water, placed in a ZnSe window and oven-dried (10 min, 50 °C). Infrared spectra were obtained using an FT-IR spectroscope (Spectrum Systems 2000 FTIR, Perkin-Elmer, Waltham, MA, USA). Measurements were recorded in the range 4,000–500  $\text{cm}^{-1}$  (interval 1  $\text{cm}^{-1}$ ). Transformation of FT-IR profiles included normalisation, smoothing and first derivative (Savitzky–Golay algorithm).

#### Cluster analysis

Hierarchical cluster analysis (calculation of coefficients using  $1 - r$ , where  $r$  is the Pearson coefficient; joining of variables using the Ward method) was carried out with Statistica for Windows version 7.0 (Statsoft Inc.).

#### Sequencing of the Internal Transcribed Spacer region

The amplified PCR product of the 16S–23S ITS region of the 12 isolates was sequenced. Forward primer (ITSA1), with sequence 5'-GGC TTG TAG CTC AGG TGG TTA GA-

3', and reverse primer (ITSB1), with sequence 5'-CYR YTG CCA AGG CAT CCA CC-3' (Y = C or T; R = A or G), were used (Barry *et al.*, 1991) for amplification of the ITS region (Gradient Thermal Cycler PTC-200, MJ Research Inc., Reno, NV, USA). For PCR, target bacteria cells were grown in BHI for 12 h at 37 °C and 100 mL aliquot was transferred to 3 mL of BHI. Cells were centrifuged at 10,000  $g$  for 5 min, collected and lysed by heating (100 °C, 10 min). Then, 10 mL of the cell suspension (approximately  $1 \times 10^5$  colony-forming units/mL) was mixed with 25 mL PCR solution containing reaction buffer (10 mM Tris–HCl, pH 8.8, 1.5 mM  $\text{MgCl}_2$ , 50 mM KCl and 0.1% Triton X-100), 200 mM deoxynucleotide triphosphates (dNTPs), 50 pmol for each of the primers and 0.4 U ProZyme. PCR cycling conditions were as follows: 5 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 60 °C and 50 s at 72 °C; and 10 min at 72 °C for the last cycle. The amplicons were electrophoresed in a 2% agarose gel. PCR products were purified (Montage® PCR Centrifugal Filter devices; Millipore, Bedford, MA, USA) according to the manufacturer's protocol. The purified products were sequenced with the MegaBACE™ 500 (Amersham Biosciences, Uppsala, Sweden) using the sequencing MegaBACE DYEnamic ET dye terminator kit (Amersham Biosciences).

#### Identification process

ITS sequences were identified by searching the National Center for Biotechnology Information database using the Basic Local Alignment Search Tool (BLAST) sequence analysis tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and were compared using nucleotide–nucleotide BLAST (blastn) with default settings.

#### Phylogenetic analysis

Chroma software (Goodstadt and Ponting, 2001) was used for formatting annotated multiple sequence alignments. Sequences were aligned together using the ClustalW multiple alignment application. Phylogenetic analysis of sequences was performed using MEGA 4 (Tamura *et al.*, 2007) with the Tajima–Nei distance measure (Tajima and Nei, 1984) and the neighbour-joining clustering method.

#### Integration of results and polyphasic analysis

Four distance matrices stemming from the different analyses (carbon use, ITS sequencing, FT-IR and ribotyping), obtained using the procedures described herein, were merged to create a polyphasic cluster. First, weighting of two sets of data (carbon use profile and sequencing) was carried out to allow for comparison of different value scales. A specific weight was given to carbon use profile ( $\times 0.1$ ) and sequencing ( $\times 10$ ). Second, the sum of the four individual distance values between two given strains was calculated, obtaining a global distance matrix. Finally, multidimensional scaling was carried

out (Statistica for Windows, version 7.0; Statsoft Inc.) to graphically present the underlying dimensions considering distances between strains globally.

## Results and discussion

### Carbon source utilisation results

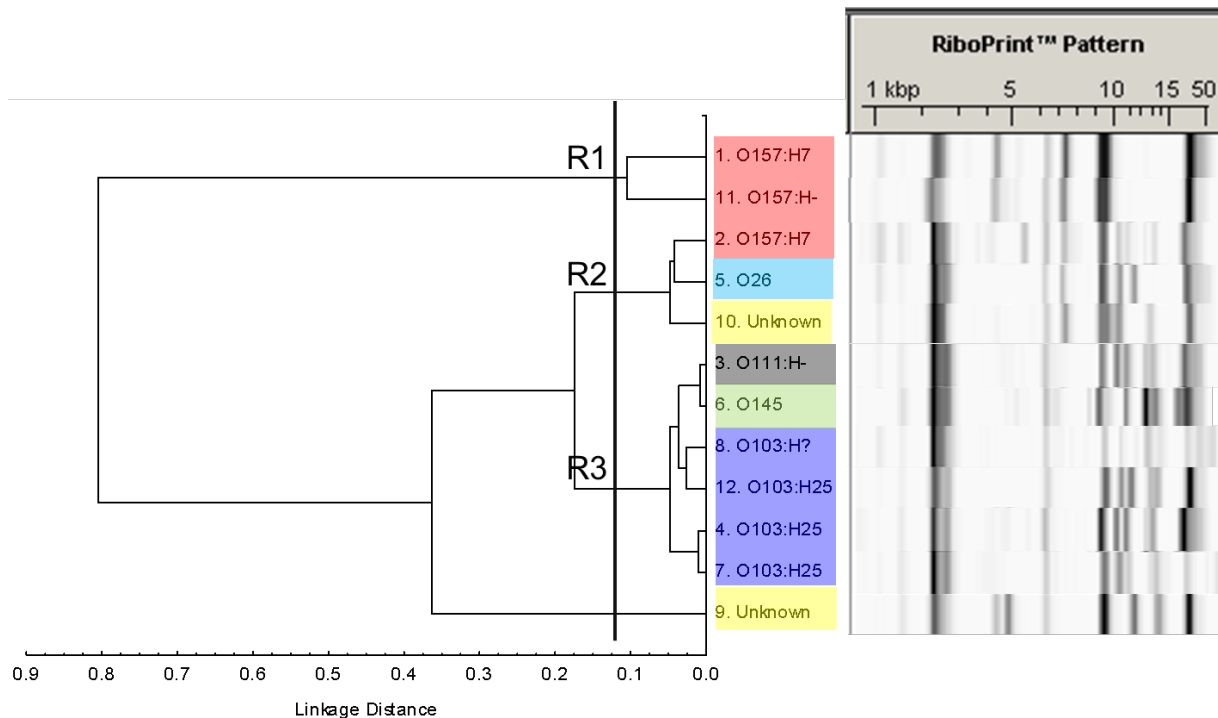
The metabolic patterns of the *E. coli* isolates were compared with reference data available in the database of the Biolog system. The results of the identification are presented in Table 1. The system correctly identified seven out of the 12 isolates at the species level (*Escherichia coli*). Two strains (MF2522 and MF2499, serotype O103) were incorrectly identified as O157:H7. Of the remaining five isolates, two were identified at the genus level as *Escherichia* spp. (MF2486 and MF3578), while three strains (MF2411, C-600 and MF3582) could not be correctly identified. Strains correctly classified as *E. coli* (EDL933, E218/02, MF2494, MF2493, MF2522, MF2499 and M23) utilised 18 carbon sources, with a utilisation rate of 100%. None of the 12 strains could use 42 carbon sources. The efficacy of the methods based on phenotypic traits depends on the extension and quality of the reference database, frequently requiring additional tests for a correct

identification (Kootallur *et al.*, 2011; Sellyei *et al.*, 2011). Although the Biolog database includes carbon source data obtained from *E. coli* O157:H7, other serotypes are not represented, precluding correct identification at the subspecies level. Carbon source utilisation data should be used cautiously to identify *E. coli* at the subspecies level, and additional methods are needed for subtype differentiation.

### Ribotyping

Table 1 displays the EcoRI RiboPrint identification results for the 12 isolates tested. Assignment to database elements (ribogroups) was based on the unknown isolate's distance to the RiboPrinter patterns using the instrument's computer algorithm to pool patterns into ribogroups. Eight isolates could be identified at the species level (Table 1). One of them (EDL933) was assigned to the correct serotype (O157:H7). A hierarchical classification based on digitised ribotyping values was also performed (Figure 1). Clustering yielded three groups: cluster R1 (EDL933 and MF3582), cluster R2 (strains E218/02, MF2486 and M23) and cluster R3 (strains MF2411, MF2493, MF2499, MF3578, MF2494 and MF2522). Strain C-600 was distantly located.

The Riboprinter database used in this study included 241 ribotypes belonging to *E. coli*, of which only two corresponded to serotype O157:H7. The absence of *E. coli* ribotypes from



**Figure 1.** Dendrogram of 12 ribotypes generated from the *E. coli* strains used in this study. Ribotypes of the strains are included at the right of the dendrogram. Cluster analysis was performed using similarity values calculated with the Pearson product moment correlation coefficient ( $r$ ) and the Ward algorithm grouping method.

**Table 1.** *Escherichia coli* strains used throughout the study and results from identification using BIOLOG, Ribotyping and 16S–23S ITS region gene sequence similarity

Isolate n°	Strain	Serogroup/type	stx gene profile	BIOLOG			RIBOTYPING			SEQUENCING 16S–23S ITS		
				BIOLOG ID result	identification	ID #1 SIM	ID #1 DIST	Ribotyping ID result	Ribotyping identification	BP	Blast hit (Similarity)	Bacterial name
#1	EDL933	O157:H7	1+2	Species ID	<i>Escherichia coli</i> O157:H7	0.649	5.00	Species ID	<i>Escherichia coli</i> O157:H7	376	CP008957.1 (100%)	<i>Escherichia coli</i> O157:H7
#2	E218/02 <sup>1</sup>	O157:H7	2	Species ID	<i>Escherichia coli</i> O157:H7	0.554	4.07	No ID		382	LM997402.1 (98%)	<i>Escherichia coli</i>
#3	MF2411 <sup>2</sup>	O111:H-	1	No ID	<i>Citrobacter freundii</i>	0.256	7.32	Species ID	<i>Escherichia coli</i>	386	LM997402.1 (99%)	<i>Escherichia coli</i>
#4	MF2494 <sup>3</sup>	O103:H25	2	Species ID	<i>Escherichia coli</i> (USP5-7085)	0.530	4.45	Species ID	<i>Escherichia coli</i>	379	LM997402.1 (99%)	<i>Escherichia coli</i>
#5	MF2486 <sup>3</sup>	O26	1	Genus ID: <i>Escherichia</i>	<i>Escherichia coli</i> O157:H7	0.475	6.27	Species ID	<i>Escherichia coli</i>	393	CP009578.1 (95%)	<i>Escherichia coli</i>
#6	MF2493 <sup>3</sup>	O145	1	Species ID	<i>Escherichia coli</i>	0.654	4.12	No ID		374	LM997016.1 (99%)	<i>Escherichia coli</i>
#7	MF2522 <sup>4</sup>	O103:H25	-	Species ID	<i>Escherichia coli</i> O157:H7	0.783	3.26	No ID		375	LM997402.1 (99%)	<i>Escherichia coli</i>
#8	MF2499 <sup>4</sup>	O103:H?	1	Species ID	<i>Escherichia coli</i> O157:H7	0.710	2.21	No ID		376	CP009578.1 (99%)	<i>Escherichia coli</i>
#9	C-600 (1021)	Unknown	-	No ID	<i>Enterobacter nimirpressurales</i>	0.371	5.75	Species ID	<i>Escherichia coli</i>	408	LM995446.1 (98%)	<i>Escherichia coli</i>
#10	M23 <sup>5</sup>	Unknown	-	Species ID	<i>Escherichia coli</i>	0.791	3.08	Species ID	<i>Escherichia coli</i>	384	CP009578.1 (99%)	<i>Escherichia coli</i>
#11	MF3582 <sup>3</sup>	O157:H-	2	No ID	<i>Escherichia coli</i>	0.411	3.38	Species ID	<i>Escherichia coli</i>	382	CP008957.1 (99%)	<i>Escherichia coli</i> O157:H7
#12	MF3578	O103:H25	2	Genus ID: <i>Escherichia</i>	<i>Escherichia coli</i> (USP5-7085)	0.239	5.84	Species ID	<i>Escherichia coli</i>	382	LM997402.1 (99%)	<i>Escherichia coli</i>

1. Kindly received from Dr. S. Löfdahl, Swedish Institute for Infectious Disease Control, Sweden
2. Kindly received from Dr. F. Scheutz, Statens Serum Institut, Denmark
3. Kindly received from Prof. G. Kapperud, Norwegian Institute of Public Health, Norway
4. Kindly received from Dr. C. Sekse, Norwegian School of Veterinary Science, Norway
5. Kindly received from Dr. T. Ross, Tasmanian Institute of Agricultural Research, University of Tasmania, Hobart Tasmania, Australia

serotypes other than O157:H7 in the database is probably the main cause for the inaccurate identifications, taking into account the fact that some serotypes tested (e.g., O111:H-, O103:H25, O26 and O145) yielded atypical ribotyping profiles (Figure 1). Accurate identification at the subspecies level is hampered by the presence of a few atypical profiles in the Riboprinter databases, considering the large variability of ribotyping profiles among VTEC strains. Assay conditions (e.g., restriction enzymes) could allow better discrimination between and within heterogeneous serotypes, including different clones and genetic lineages within the VTEC group.

**Analysis by FT-IR spectroscopy**

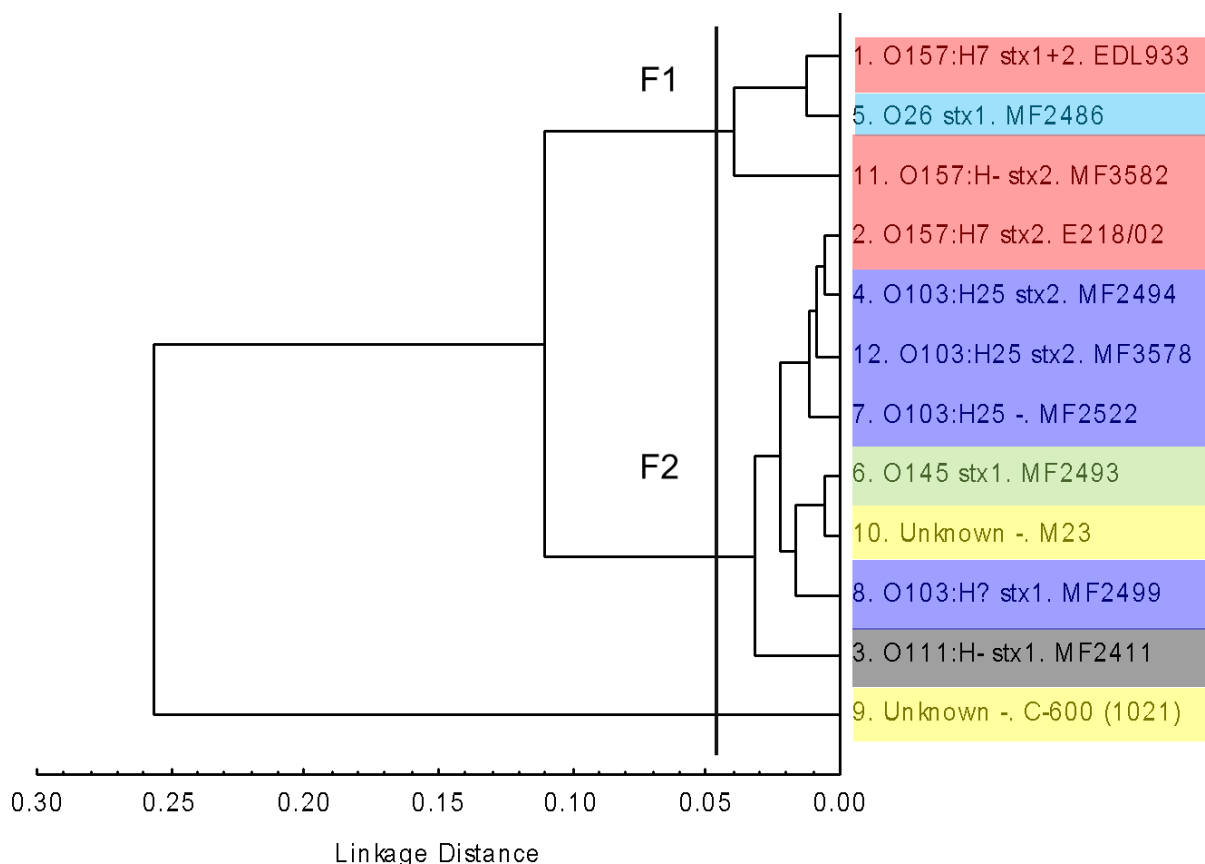
A classification scheme yielding a hierarchical cluster (Pearson correlation coefficient *r*, Ward cluster algorithm) based on the first-derivative FT-IR spectra was tested (Figure 2). A sub-range of the spectrum (1,000–950 cm<sup>-1</sup>), with a global linkage distance of 0.6, was the most discriminant, yielding two clusters: F1 (EDL933, MF2486 and MF3582); and F2, comprising most of the strains (E218/02, MF2411,

MF2494, MF2493, MF2522, MF2499, M23 and MF3578). One strain (C-600) remained unclustered.

FT-IR spectroscopy has high discriminatory power at the subspecies level (Mouwen *et al.*, 2005; Mouwen *et al.*, 2011), including for *E. coli* (Davis *et al.*, 2012; Sousa *et al.*, 2013; Dawson *et al.*, 2014; Prieto-Calvo *et al.*, 2014). The current study also showed a significant level of agreement between certain groupings (strains MF2494, MF2522 and MF3578; strains EDL933 and MF3582) when using FT-IR spectra and ribotyping or ITS rRNA gene sequencing (Figure 2).

**Sequence diversity of the ITS region**

The 16S–23S rRNA ITS region of the 12 strains of *E. coli* was sequenced and studied for intra-species relationships. Comparative sequence analysis of the ITS region (encompassing between 376 bp and 408 bp) of the 12 isolates was used to infer the phylogenetic relationships between all taxa. The strains of *E. coli* studied showed a very limited variation at the level of the ITS region, indicated by an average pairwise nucleotide difference (*p*, average pairwise



**Figure 2.** Dendrogram, obtained from 1st derivative FT-IR spectra, from the *E. coli* strains used in this study. Cluster analysis was performed using distance indexes calculated with the formula  $1 - r$  (*r*, Pearson product moment correlation coefficient) and the Ward algorithm grouping method.

nucleotide difference per site) of 0.011563 and a minimal linkage distance in the cluster (Figure 3).

The identification was performed by comparing the sequences obtained against the non-redundant data of the GenBank database using BLAST (Zhang *et al.*, 2000). The BLAST analysis allowed the correct identification of two strains (EDL933 and MF3582) as *E. coli* O157:H7, with more than 99% sequence similarity. The other 10 strains were identified at the species level, with sequence similarities in the range of 95%–99%.

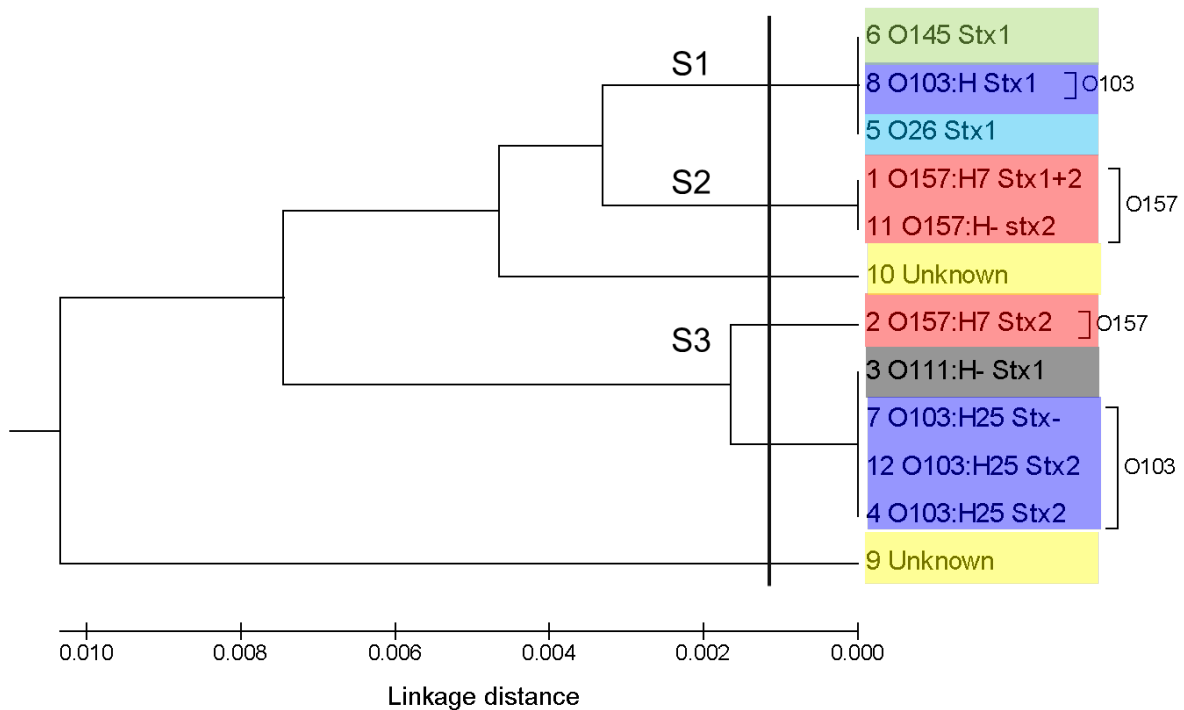
In the phylogenetic analysis (Figure 3), cluster S1 was formed by strains MF2493, MF2486 and MF2499. Cluster S2 contained two strains (EDL933 and MF3582) belonging to the O157 serotype, while cluster S3 was composed of four strains (MF2411, MF2494, MF2522 and MF3578), three of them from the O103 serotype. Three strains did not cluster and remained isolated (E218/02, C-600 and M23).

ITS regions are variable among species and can be useful for species recognition. Utility of rRNA genes to infer phylogenetic relationships has been previously established because of its clockwise pace of change (Woese, 1987). The ITS region of the 16S–23S rRNA gene is usually more variable than the flanking 16S and 23S ribosomal genes. The feasibility of sequence analysis of the rRNA genes' ITS regions for the

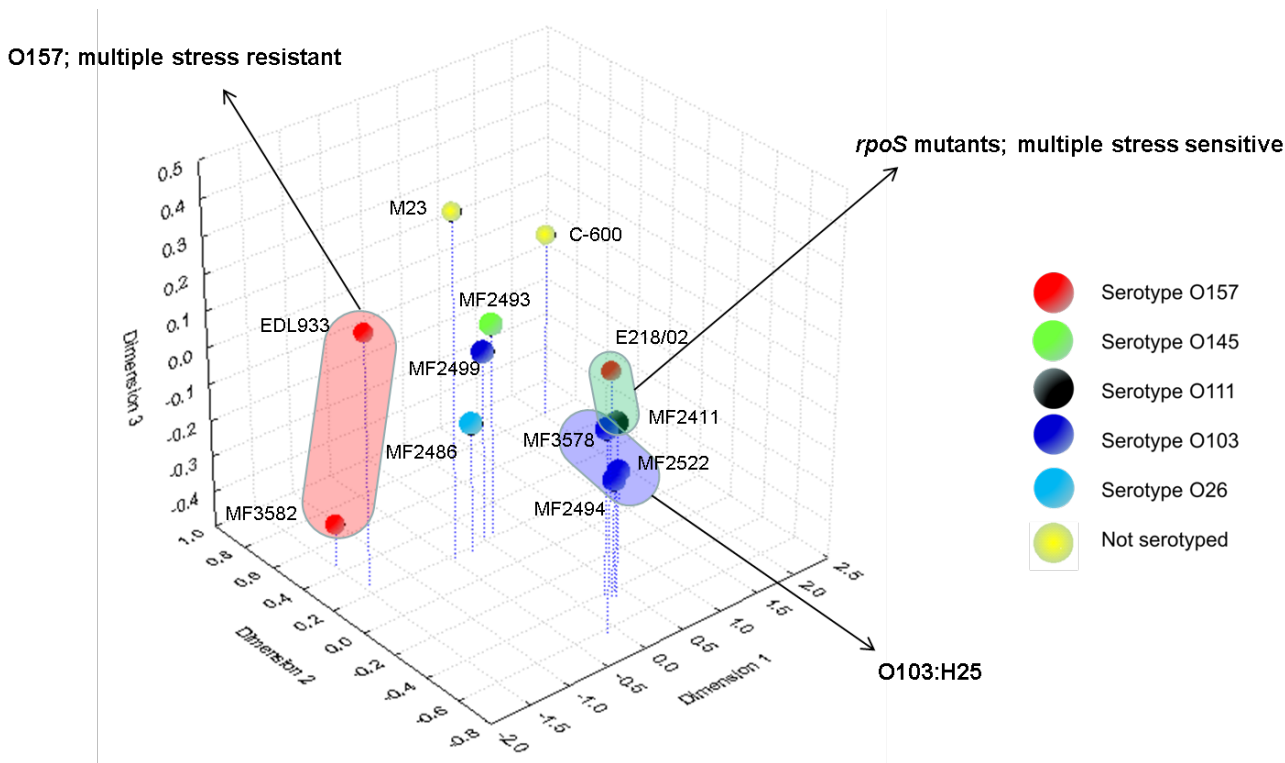
identification of both yeasts of clinical relevance (Leaw *et al.*, 2006) and bacteria (Perez-Luz *et al.*, 1998; Wang *et al.*, 2008; Fernandez and Vendano-Herrera, 2009) has been previously demonstrated. In genera such as *Salmonella*, the ITS region has a significant level of polymorphism across serotypes and is therefore useful for subtype discrimination (Perez-Luz *et al.*, 1998). The current study shows that sequencing of *E. coli* based on the ITS fragment is very accurate for identification purposes but does not allow a satisfactory sub-typing of *E. coli* at the subspecies level, because it does not discriminate between strains from different serotypes. Within closely related strains, the spacer region can be quite conserved. This is probably due to its location between two highly conserved genes as well as the concerted evolution of a multi-gene family that helps in homogenising and conserving the sequences not only within individual cells but also within groups among which horizontal genetic exchange is relatively common.

**Polyphasic analysis and integration of results**

A distance matrix was created with the weighted sum of values (1 – Pearson *r*) from the analysis using the four methods described herein. The resulting three-dimensional scatterplot (Figure 4) was prepared displaying the underlying dimensions calculated by means of a multidimensional scaling. Strains



**Figure 3.** Cluster showing the relationships of 12 taxa, inferred using the UPGMA method. The optimal tree with the sum of branch length = 0.03763754 is shown. A total of 304 positions were in the final dataset. Analyses were conducted in MEGA4.



**Figure 4.** Multidimensional analysis performed using four distance indexes (1-Pearson *r*), from four characterization methods, for grouping. Four distance matrices stemming from the different analysis (carbon use profile, sequencing, FT-IR and ribotyping) were merged together to create a polyphasic scatterplot.

EDL933 and MF3582 (both *E. coli* O157) clustered together. Interestingly, these two strains have been reported to present high resistance to multiple stresses (Alvarez-Ordóñez *et al.*, 2013). The other O157 strain (E218/02), which has a nonsense mutation in the gene encoding the general stress response regulator *rpoS*, was situated apart. This separated position seems logical considering that RpoS regulates the transcription of a wide range of genes, therefore, influencing cellular composition and growth behaviour (Alvarez-Ordóñez *et al.*, 2015), phenotypes that were both included in the polyphasic approach. Indeed, E218/02 was located close to MF2411, a strain from serotype O111, which also has a truncated RpoS. Also closely related to E218/02 and MF2411 were strains MF2494, MF2522 and MF3578, which all belonged to serotype O103:H25. All O103:H25 strains showed very similar genotypic and phenotypic traits and were grouped together by all characterisation systems. Strains MF2522 and MF2494 had very similar features, including identical ITS sequencing data, ribotype and FT-IR spectra. Both O103:H25 strains were, among other isolates, linked to the Norwegian outbreak in 2006 and show identical multiple-loci variable-number tandem-repeat analysis genotypes (Lindstedt *et al.*, 2007; Schimmer *et al.*, 2008). As expected,

the rifampicin-resistant variant of MF2494, MF3578, was consistently clustered with its wild-type strain. Strain MF2499, of serotype O103:H?, did not group, on the other hand, with O103:H25 strains.

The current study included a single representative strain from other serotypes (O26, O111 and O145), located somewhat closer to various strains of serotype O103. Monitoring of a larger collection of strains from serotype O157 and other clinically relevant non-O157 serotypes would be necessary to test the robustness of the proposed polyphasic approach for the sub-typing of *E. coli*.

Strains C-600 and M23 showed features that placed them apart from the remaining strains. These are non-pathogenic laboratory isolates that are phenotypically and genotypically distant from the clinical strains. They presented large variability in properties and were placed erratically in the clusters depending on the typing system. That was also the outcome obtained in the polyphasic analysis.

To sum up, a system integrating heterogeneous taxonomic information stemming from different methodologies (phenotypic, chemotaxonomic and genotypic) by using distance matrices and hierarchical clustering has been reported. Integration of heterogeneous information from

different sources successfully meets the challenges posed by bacterial identification and characterisation schemes (Valmorri *et al.*, 2006; Geider *et al.*, 2009). Polyphasic approaches are able to reduce the bias posed by a single method with too much emphasis on only phenotypic, chemotaxonomic or genotypic characteristics. Moreover, there is high level of agreement between the polyphasic arrangement obtained and the classification recently made based on the multivariate analysis of stress resistance patterns (Alvarez-Ordóñez *et al.*, 2015), which shows that this approach is able to not only reveal taxonomic relationships but also provide useful cues on strain ecology and physiology, giving information on strain behaviour in relation to food and the environment. To improve the taxonomic significance of strain arrangements, additional genotyping techniques such as multilocus sequence typing (MLST) could be integrated in the polyphasic approach. Limitations of the genotyping techniques used in this study in relation to their discriminatory power and reproducibility could be solved by incorporating taxonomic information extracted from MLST. More research is needed to extrapolate the current results by using a more extensive set of strains (including more strains per serotype and other serotypes) to improve the taxonomic picture and discard the misidentification of atypical strains.

## Acknowledgements

The authors acknowledge the financial contribution of the Spanish INIA, the Research Council of Norway (project 178230/I10), Foundation for Levy on Foods, the Norwegian Research Fees Fund for Agricultural Goods, the Norwegian Independent Meat and Poultry Association, Nortura SA and NHO Mat og Landbruk.

## Conflicts of interest

The authors declare no conflicts of interest.

## References

- Alvarez-Ordóñez, A., Mouwen, D.J.M., Lopez, M. and Prieto, M. 2011. Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria. *Journal of Microbiological Methods* **84**: 369–378.
- Alvarez-Ordóñez, A., Alvseike, O., Omer, M.K., Heir, E., Axelsson, L., Holck, A. and M. Prieto. 2013. Heterogeneity in resistance to food-related stresses and biofilm formation ability among verocytotoxigenic *Escherichia coli* strains. *International Journal of Food Microbiology* **161**: 220–230.
- Alvarez-Ordóñez, A., Broussolle, V., Colin, P., Nguyen-The, C. and Prieto, M. 2015. The adaptive response of bacterial food-borne pathogens in the environment, host and food: implications for food safety. *International Journal of Food Microbiology* **213**: 99–109.
- Barry, T., Colleran, G., Glennon, M., Dunican, L.K. and Gannon, F. 1991. The 16s/23s ribosomal spacer region as a target for DNA probes to identify eubacteria. *PCR Methods and Applications* **1**: 51–56.
- Davis, R., Paoli, G. and Mauer, L.J. 2012. Evaluation of Fourier transform infrared (FT-IR) spectroscopy and chemometrics as a rapid approach for sub-typing *Escherichia coli* O157:H7 isolates. *Food Microbiology* **31**: 181–190.
- Dawson, S.E., Gibreel, T., Nicolaou, N., AlRabiah, H., Xu, Y., Goodacre, R. and Upton, M.. 2014. Implementation of Fourier transform infrared spectroscopy for the rapid typing of uropathogenic *Escherichia coli*. *European Journal of Clinical Microbiology & Infectious Diseases* **33**: 983–988.
- EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control). 2015. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. *EFSA Journal* **13**: 3991.
- Fernandez, J. and Vendano-Herrera, R. 2009. Analysis of 16S-23S rRNA gene internal transcribed spacer of *Vibrio anguillarum* and *Vibrio ordalii* strains isolated from fish. *FEMS Microbiology Letters* **299**: 184–192.
- Geider, K., Auling, G., Jakovljevic, V. and Voelksch, B. 2009. A polyphasic approach assigns the pathogenic *Erwinia* strains from diseased pear trees in Japan to *Erwinia pyrifoliae*. *Letters in Applied Microbiology* **48**: 324–330.
- Goodstadt, L. and Ponting, C.P. 2001. CHROMA: Consensus-based colouring of multiple alignments for publication. *Bioinformatics* **17**: 845–846.
- Grimont, F. and Grimont, P.A. 1986. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. *Annales de l'Institut Pasteur/Microbiologie* **137B**: 165–175.
- Guasp, C., Moore, E.R., Lalucat, J. and Bennisar, A. 2000. Utility of internally transcribed 16S-23S rDNA spacer regions for the definition of *Pseudomonas stutzeri* genomovars and other *Pseudomonas* species. *International Journal of Systematic and Evolutionary Microbiology* **50(Pt 4)**: 1629–1639.
- Gurtler, V. and Stanisich, V.A. 1996. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **142**: 3–12.
- Kootallur, B.N., Thangavelu, C.P. and Mani, M. 2011. Bacterial identification in the diagnostic laboratory: how much is enough? *Indian Journal of Medical Microbiology* **29**: 336–340.
- Leaw, S.N., Chang, H.C., Sun, H.F., Barton, R., Bouchara, J.P. and Chang, T.C. 2006. Identification of medically important yeast species by sequence analysis of the internal transcribed spacer regions. *Journal of Clinical Microbiology* **44**: 693–699.
- Lindstedt, B.A., Brandal, L.T., Aas, L., Vardund, T. and Kapperud, G.

2007. Study of polymorphic variable-number of tandem repeats loci in the ECOR collection and in a set of pathogenic *Escherichia coli* and *Shigella* isolates for use in a genotyping assay. *Journal of Microbiological Methods* **69**: 197–205.
- Motoyama, Y. and Ogata, T. 2000. 16S-23S rDNA spacer of *Pectinatus*, *Selenomonas* and *Zymophilus* reveal new phylogenetic relationships between these genera. *International Journal of Systematic and Evolutionary Microbiology* **50**(Pt 2): 883–886.
- Mouwen, D., Horman, A., Korkeala, H., Alvarez-Ordóñez, A. and Prieto, M. 2011. [Applying Fourier-transform infrared spectroscopy and chemometrics to the characterization and identification of lactic acid bacteria](#). *Vibrational Spectroscopy* **56**: 193–201.
- Mouwen, D.J.M., Weijtens, M.J.B.M., Capita, R., Alonso-Calleja, C. and Prieto, M. 2005. Discrimination of enterobacterial repetitive intergenic consensus PCR types of *Campylobacter coli* and *Campylobacter jejuni* by Fourier transform infrared spectroscopy. *Applied and Environmental Microbiology* **71**: 4318–4324.
- Naumann, D. 2000. Infrared spectroscopy in microbiology. In: "Encyclopedia of Analytical Chemistry", (ed. R.A. Meyers), John Wiley & Sons Ltd, Chichester, pages 1–29.
- Odumeru, J.A., Steele, M., Fruhner, L., Larkin, C., Jiang, J., Mann, E. and McNab, W.B. 1999. Evaluation of accuracy and repeatability of identification of food-borne pathogens by automated bacterial identification systems. *Journal of Clinical Microbiology* **37**: 944–949.
- Perez-Luz, S., Rodriguez-Valera, F., Lan, R. and Reeves, P.R. 1998. Variation of the ribosomal operon 16S-23S gene spacer region in representatives of *Salmonella enterica* subspecies. *Journal of Bacteriology* **180**: 2144–2151.
- Prieto-Calvo, M., Prieto, M., Lopez, M. and Álvarez-Ordóñez, A. 2014. Effects of high hydrostatic pressure on *Escherichia coli* ultrastructure, membrane integrity and molecular composition as assessed by FTIR spectroscopy and microscopic imaging techniques. *Molecules* **19**: 21310–21323.
- Schimmer, B., Nygard, K., Eriksen, H.M., Lassen, J., Lindstedt, B.A., Brandal, L.T., Kapperud, G. and Aavitsland, P. 2008. Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive *Escherichia coli* O103:H25 traced to cured mutton sausages. *BMC Infectious Diseases* **8**: 41.
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A. and Chen, W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the USA* **109**: 6241–6246.
- Sellyei, B., Wehmann, E., Makrai, L. and Magyar, T. 2011. Evaluation of the Biolog system for the identification of certain closely related *Pasteurella* species. *Diagnostic Microbiology and Infectious Disease* **71**: 6–11.
- Sousa, C., Novais, A., Magalhaes, A., Lopes, J. and Peixe, L. 2013. Diverse high-risk B2 and D *Escherichia coli* clones depicted by Fourier transform infrared spectroscopy. *Scientific Reports* **3**: 3278.
- Tajima, F. and Nei, M. 1984. Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* **1**: 269–285.
- Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596–1599.
- Valmorri, S., Settanni, L., Suzzi, G., Gardini, F., Vernocchi, P. and Corsetti, A. 2006. Application of a novel polyphasic approach to study the lactobacilli composition of sourdoughs from the Abruzzo region (central Italy). *Letters in Applied Microbiology* **43**: 343–349.
- Vandamme, P. and Peeters, C. 2014. Time to revisit polyphasic taxonomy. *Antonie Van Leeuwenhoek* **106**: 57–65.
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. and Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiology Reviews* **60**: 407–438.
- Wang, M., Cao, B., Yu, Q., Liu, L., Gao, Q., Wang, L. and Feng, L.. 2008. Analysis of the 16S-23S rRNA gene internal transcribed spacer region in *Klebsiella* species. *Journal of Clinical Microbiology* **46**: 3555–3563.
- Woese, C.R. 1987. Bacterial evolution. *Microbiology Reviews* **51**: 221–271.
- Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**: 203–214.