Multi-Loci Gene Sequencing and Identification of Bifidobacteria Strains Isolated from Dairy and Pharmaceutical Sources in South Africa

Richard Nyanzi[a], Piet J. Jooste[a], and Jacobus N. Eloff[b]

[a]Department of Biotechnology and Food Technology, Tshwane University of Technology, Pretoria, South Africa;

[b]Phytomedicine Programme, Department of Paraclinical Sciences, University of Pretoria, Onderstepoort, South Africa

CONTACT Richard Nyanzi, Private Bag x680, Pretoria, 0001, South Africa, email: ricardas62@hotmail.comi

ABSTRACT

The 16S rRNA gene sequence analysis of *Bifidobacterium* species reveals high interspecies sequence similarity in the range of 87.7–99.5%. This study illustrated the extent of superiority of a multigenic approach, involving protein-coding genes, in comparison to the 16S rRNA gene, to precisely delineate presumptive *Bifidobacterium* isolates obtained from probiotic milk beverages, culture collections and pharmaceutical probiotic preparations. Oligonucleotide pairs PurF-rev/PurF-uni; RpoC-uni/RpoC-rev; DnaB-uni/DnaB-rev; DnaG-uni/DnaG-rev; and ClpC-uni/ClpC-rev amplified housekeeping genes while 27F/1492R amplified the 16S rRNA gene of the presumptive bifidobacteria in a polymerase chain reaction. Sequences of 16S rRNA gene and some protein-coding genes effectively identified the isolates. Phylogenetic analysis together with concatenation showed that *clpC, purF* and *dnaG* genes had over 8-fold better discriminatory power than the 16S rRNA gene in discriminating between *Bifidobacterium* isolates. However, phylogenetic analysis involving *dnaB and rpoC* gene sequences or their concatenated trees showed discrepancies in clustering isolates with designated type strains.

Introduction

Probiotics are defined as viable microbial dietary supplements that beneficially affect the host beyond basic nutrition when consumed adequately and regularly (Buruleanu et al. 2012; Sadaghdar et al., 2012). The species and strains used as probiotics mostly belong to the genera Lactobacillus and Bifidobacterium (Vasconcelos et al., 2014). More than 30 Bifidobacterium species are known, some of which occur naturally in the human gastrointestinal tract (GIT) (Ventura et al., 2006). Bifidobacteria are the most predominant in the digestive tract of infants, while in adults a third of the intestinal bacterial population comprises bifidobacteria (Ventura et al., 2005; Buruleanu et al. 2012). Several prophylactic and therapeutic health benefits in

humans have been attributed to the existence of probiotic Bifidobacterium species and other probiotic species in the GIT (Ventura et al., 2005; Ventura et al., 2006; Delétoile et al., 2010; Kim et al., 2012; Buruleanu et al. 2012; Sadaghdar et al., 2012). On the basis of such effects, bifidobacteria are increasingly being incorporated in probiotic beverages and supplements (Chang et al., 2010; Koh et al., 2014; Vasconcelos et al., 2014). Bifidobacteria, particularly Bf. lactis and Bf. animalis, are increasingly being applied in commercial fermented dairy products based on technological aspects and probiotic properties (Mayer et al., 2007; Buruleanu et al., 2012; Kim et al., 2012). However, technological and health beneficial properties are strain dependent, which dictates the need to use competent techniques with high discriminatory power to differentiate and identify the strains to be used. Identification based on phenotypic characterization of Bifidobacterium strains is unreliable due to possible morphological changes related to growth and culturing conditions (Mayer et al., 2007). The 16S rRNA gene sequence analysis consequently forms the basis for most of the molecular techniques used in the identification of Bifidobacterium strains (Ventura et al., 2006). One of the limitations of the 16S rRNA gene sequence analysis in this regard is the high interspecies sequence similarity in the range of 87.7–99.5% (Ventura et al., 2006; Kim et al., 2012). Ventura et al. (2006) reported several studies in which subspecies of Bf. animalis and Bf. longum could not be differentiated using their 16S rRNA gene sequences. According to Ventura and Zink (2003), the use of a single ribosomal gene, like the 16S rRNA gene, in resolving taxonomic challenges has, therefore, increasingly become highly limited. A multigenic approach, involving several housekeeping (protein-coding) genes is reported to be superior to the 16S rRNA gene sequence analysis in the typing of close related strains (Ventura et al., 2006; Dale and Park, 2010; Delétoile et al., 2010;). It was reported that protein-coding genes appear in single copies in the genome, they are highly conserved and accumulate mutations at a slow rate hence they are capable of revealing evolutionary divergences (Ventura and Zink, 2003; Dale and Park, 2010). Ventura et al. (2006) observed that certain housekeeping gene markers enabled successful discrimination between *Bifidobacterium* species.

Due to lactose intolerance and/or cultural reasons, some people do not consume probiotic dairy products. A variety of cereals, however, are consumed globally. It would therefore make sense to develop a cereal-based synbiotic beverage containing multiple probiotic bacterial strains. In line with the ultimate objective of developing cereal-based synbiotic beverages containing multiple probiotic bacterial strains, this study used oligonucleotides of certain housekeeping genes (Ventura et al., 2006) to determine enhancement of the discriminatory power in terms of identification as well as the phylogenetic analysis of *Bifidobacterium* isolates obtained from different food and pharmaceutical sources.

Table 1. Isolates from different probiotic food and pharmaceutical sources.

Isolate code	Source of strain	Strain claimed on label
BA	*Carst & Walker (Frozen)	LAFTI B94
BB	*Lake Foods, (Freeze-dried)	Bb12
BC	Pharma Supplier (Frozen)	BLC
BD	Probiotic pharmaceutical liquid	35624
BE	Health tablet A	R-175
BF	Health capsule A	–
BG	Company B yoghurt	HN019
BH	Health tablet C	AY735403
BI	Health capsules D	Bi-07
BJ	University of the Free State	UFSBC503
BK	University of the Free State	UFSBC505
H14	Yoghurt B (month 1)	HN019
H22	Yoghurt B (month 2)	HN019
H32	Yoghurt B (month 3)	HN019
H42	Yoghurt B (month 4)	HN019
A1	Yoghurt E (month 1)	DN 173010
A2	Yoghurt E (month 2)	DN 173010
A3	Yoghurt E (month 3)	DN 173010
A4	Yoghurt E (month 4)	DN 173010

*Reference strains.

Materials and methods

Bacterial isolates

The sources for the bacterial isolates included probiotic supplements from pharmaceutical outlets, probiotic food beverages and supplier companies. The bacterial isolates' codes and sources are shown in Table 1. Bifidobacterial strains were isolated on a monthly basis over a period of four months from two popular local fermented dairy products (brands). *Bifidobacterium animalis* subsp. *lactis* Bb12 (Lake Foods, SA) was used as a reference strain because it has been a commercial strain for over 25 years and is cited in more than 200 scientific publications (Su et al., 2005; Barrangou et al., 2009; Buruleanu et al., 2012).

Purification and preservation of isolates

About 1g (if solid) or 1 mL (if liquid) of probiotic product sample was aseptically introduced into 9 mL of sterile *Bifidobacterium* selective medium (BSM). Anaerobic incubation (in anaerobic jars inserted with anaerobic catalysts from Merck, Germany) was done at 37°C for 24–48 h or until growth was observed. A loopful of broth culture was then inoculated into fresh sterile BSM broth and incubated for a further 24 h. The broth culture was streaked onto BSM agar in 90 mm Petri dishes and anaerobically incubated for 24–48 h until colonies were obtained. Representative colonies were anaerobically sub-cultured several times on MRS agar to purify the isolates. The purified isolates were then grown in MRS broth and aliquots preserved using freezing medium as described by Nyanzi (2007). Representative colonies were subjected to Gram staining and cell

morphology examination by light microscopy. Tetramethyl-*p*-phenylenedia-mine and hydrogen peroxide (3%) were used to conduct the oxidase and catalase tests respectively. Isolates that were oxidase- and catalase negative Gram positive rods were subjected to further phenotypic characterisation using the API rapid ID 32 A gallery specific for bifidobacteria (Biomérieux, South Africa).

Detection of fructose-6-phosphate phosphoketolase (F6PPK) activity

Fructose-6-phosphate phosphoketolase (F6PPK) activity was used as a screening test for bifidobacteria according to the method of Orban and Patterson (2000). The control tubes consisted of everything except bacterial cells. A reddish-violet color was the positive test for strains of species belonging to the genus *Bifidobacterium* (Orban and Patterson, 2000).

Primers used in the investigation

Primers used in the multigenic sequence typing of Bifidobacterium and other bacterial isolates included 27F/1492R(1) (Guo et al., 2010); PurF-rev/PurF-uni, RpoC-uni RpoC-rev, DnaB-uni/DnaB-rev, DnaG-uni/DnaG-rev, and ClpC-uni/ClpC-rev (Ventura et al., 2006). Their nucleotide sequences and sizes of the primers were reported by Ventura et al. (2006).

Isolation of DNA

The DNA was isolated from purified cultures using the QIAGEN® kit by following the manufacturer's protocol in the DNeasy® Blood & Tissue leaflet (Whitehead Scientific, Cape Town, South Africa). A NanoDrop 2000 spectro-photometer (Thermo Scientific, Wilmington, Delaware, USA) was used to determine the concentration of the eluted DNA and the DNA was then stored at -20°C until needed.

DNA amplification and sequencing

The 25.0 μL reaction mixture for the amplification of the 16S rRNA gene contained 1.0 X PCR (polymerase chain reaction) buffer, 1.5 mM $MgCl_2$, 1.0 μL (99%) dimethyl sulphoxide, 0.4 mM dNTPs, 0.4 μm of each primer, and 2.5 U *Taq* DNA Super-Thermpolymerase (Whitehead Scientific, SA) together with 1.0 μL of DNA template. The PCR conditions were as follows: initial denaturation (94.0°C for 15 min), 30 cycles of denaturation (95.0°C for 1.0 min), annealing (50°C for 40.0 sec), extension (72.0°C for 1.0 min 30 sec), and final extension (72.0°C for 5.0 min).

The 25.0 μL reaction mixture for the amplification of the housekeeping genes contained master mix 12.5 μL, forward primer and reverse primer, 1

µL each; RNase free water, 7.5 µL and 3 µL of DNA template. The PCR conditions used were as described by Ventura et al. (2006).

The PCR amplicons, including the negative control, were separated on 1.5% (m/v) agarose gel (Seakem® L E Agarose) stained with 0.75 µL 10 mg/mL ethidium bromide in 0.5 X TBE electrophoresis buffer at 90 mV for 45 min. Visualization of the PCR fragments was done under an ultraviolet transilluminator (VilberLourmat, Marne La Vallee, France) and a Bio-Rad Gel Doc™ XR+ Imaging system transilluminator.

The PCR amplicons were sequenced in a Genetic Analyser ABI PRISM™ 3100 (Applied Biosystems, Foster City, California, USA). Sequences were edited using the software Chromas Lite 2.0 and BioEdit v. 7.0.9 (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Preliminary identification was done by comparing the isolates' sequences with those in the nucleotide database of the National Centre for Biotechnology Information (NCBI) using the search option of blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Phylogenetic analysis

In order to confirm identities and show evolutionary relationships among operational taxonomic units (OTUs), the 16S rDNA sequences and the five housekeeping gene (*purF; rpoC; dnaB; dnaG;* and *clpC*) sequences were independently subjected to phylogenetic analysis. Initial alignment was done by ClustalW in BioEdit v. 7.0.9 (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and the rest of the analyses, including evolutionary divergence estimation, was done by using MEGA 5.05 software (Tamura et al., 2011). The isolates' sequences were arranged in a similar order for each of the targeted housekeeping genes and were then aligned, cut to size and concatenated using MAFFT v. 7.245 software (http://mafft.cbrc.jp/alignment/software). The combined sequences were used to construct a concatenated phylogenetic tree, calculate the overall mean evolutionary divergence, and conduct Tajima's D test by using MEGA 5.05 software. The isolates' gene sequences were deposited at the nucleotide sequence databases (EMBL and DDBJ) and allocated accession numbers: DDBJ, AB759515 – AB759533 (16S rDNA sequences); EMBL, HF548598 – HF548610 (*clpC* gene sequences); HF548611 – HF548623 (*dnaG* gene sequences); HF548624 – HF548636 (*dnaB* gene sequences); HF548637 – HF548649 (*rpoC* gene sequences); HF548650 – HF548662 (*PurF* gene sequences).

Results and discussion

Detection of fructose-6-phosphate phosphoketolase (F6PPK) activity

Most of the isolates shown in Table 1 (BA, BB, BC, BD, BG, A1, A2, A3, A4, H14, H22, H32, H42) were positive for the F6PPK enzyme. Strains BE, BF,

BH, BI, BJ and BK were negative for F6PPK activity. The presence of the F6PPK enzyme (characteristic of the bifid-shunt in *Bifidobacterium*) was indicated by the development of a reddish-violet color (Orban and Patterson, 2000; Bevilacqua et al., 2003), while pale yellow reactions were observed for the strains without the F6PPK enzymatic activity. The isolates that were positive for F6PPK activity and the Gram stain were presumed to belong to the genus *Bifidobacterium*.

The strains which were negative for F6PPK activity highlighted the lack of sufficient inhibition (by mupirocin; Bunesova et al., 2015) of non-bifidobacteria by *Bifidobacterium* Selective Medium (BSM), reported to be selective for bifidobacteria (Leuschner et al., 2003; Ashraf & Shah, 2011). This may imply the absence of *Bifidobacterium* strains in some probiotic supplements tested despite the labelling claiming their presence. The F6PPK reaction, however, is not exclusive as there are other non-*Bifidobacterium* species, such as *Gardnerella vaginalis*, that can give a positive reaction (Gavini et al., 1996). The F6PPK reaction was therefore regarded as a screening test prior to definitive identification. All isolates, positive or negative for the F6PPK reaction, were further subjected to gene sequencing and phylogenetic analysis for precise identification.

Phylogeny on the basis of 16S rDNA sequences

The data of the studied bacterial isolates together with selected GenBank 16S rDNA sequences were used to construct a phylogenetic tree (Fig. 1). All the strains studied that were positive for the F6PPK activity associated and clustered with *Bf. lactis* [X89513] and *Bf. animalis* subsp. *lactis* strains JCM 10602 [AB507074] and YIT 4121[AB050136]. The latter two strains are synonyms of strain DSM 10140 (https://www.ncbi.nlm.nih.gov/nuccore/15823597), the designated type strain for *Bf. animalis* subsp. *lactis*.

The above-mentioned cluster was clearly separate from strain JCM 1190 [AB507070 or AB116277], which is synonymous with ATCC 25527, the designated type strain for *Bf. animalis* subsp. *animalis* (Dong et al., 2000). This confirmed that all the *Bifidobacterium* isolates studied were in fact *Bf. animalis* subsp. *lactis* as they associated more closely with strain YIT 4121 (=DSM 10140, the designated type strain). Possibly, strains JCM 1253 [AB507071] and JCM 7117 [AB507072] were inaccurately registered as *Bf. animalis* subsp. *animalis* at the time their sequences were deposited at the GenBank. If JCM 1253 and JCM 7117 were used alone, in this study, to infer identity to the isolates, they would have been erroneously identified as *Bf. animalis* subsp. *animalis*. This highlighted the relevance of using sequences of uncontested designated type strains, among others. The isolates (BE, BF, BH, BI, BJ, and BK) that were negative for F6PPK activity did not associate with any *Bifidobacterium* species proving that determination of the F6PPK
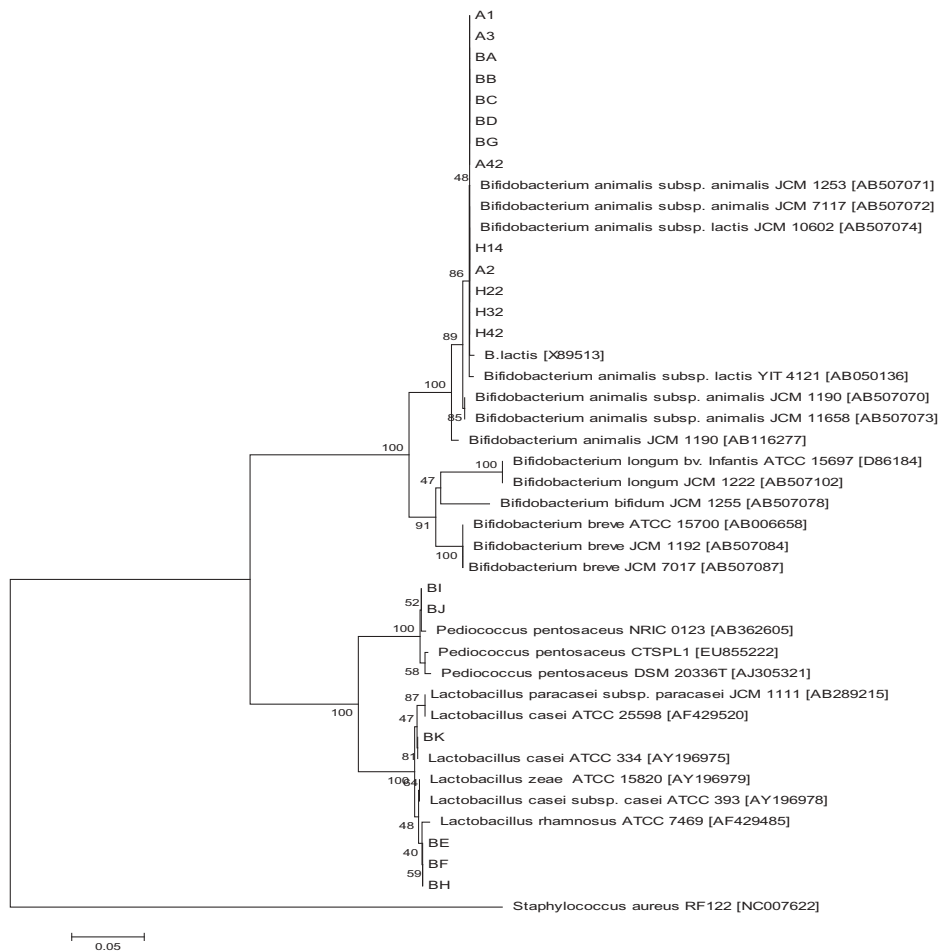
Figure 1. The phylogenetic tree for bacterial isolates constructed using the 16S rDNA sequences and the Neighbor-Joining method. The percentages of the bootstrap value (1000 replications) by which the associated strains clustered together are shown at the internodes. Accession numbers (in parentheses) are for sequences from the NCBI GenBank. Bar, 5% substitutions per site.

activity is effective in presumptive identification of bifidibacteria. The phylogenetic tree topology provided by the Neighbor-Joining method was comparable to the one constructed using the maximum Likelihood method (not shown). The non-bifidobacterial isolates were not subjected to further investigations.

Phylogeny on the basis of the protein-coding gene sequences

All isolates identified as *Bifidobacterium* strains on the basis of their 16S rDNA sequences were subjected to phylogenetic analysis in a multigenic sequence analysis (MLSA) on the basis of five selected protein-coding genes. Phylogeny based on *clpC, purF and dnaG* gene sequences of the
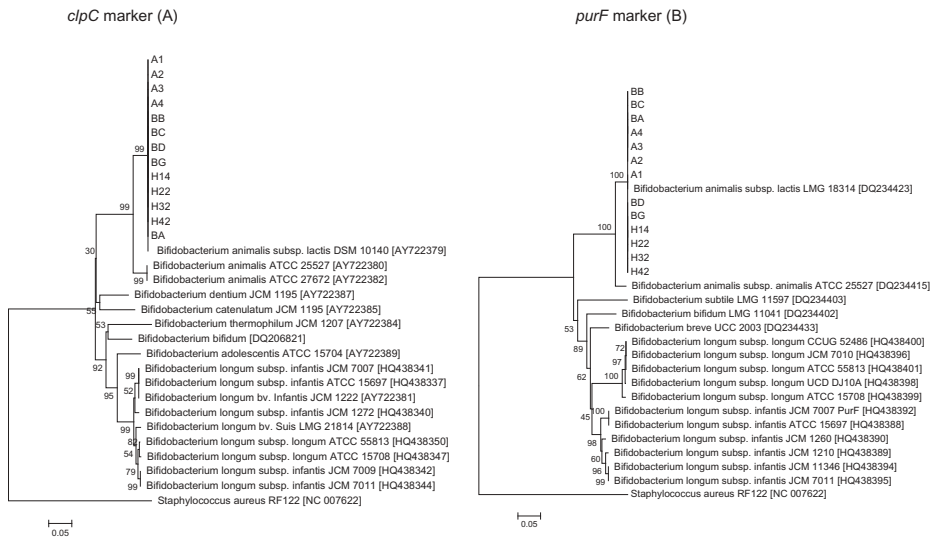
Figure 2. The phylogenetic trees for *Bifidobacterium* isolates constructed with the *clpC* gene (A) and the *purF* gene (B) sequences using the Neighbor-Joining method and the p-distance model. The bootstrap values are indicated at the internodes. Accession numbers (in parentheses) are for the sequences retrieved from the GenBank. *Staphylococcus aureus* RF 122 was used to root the trees.

Bifidobacterium isolates clearly confirmed the previous findings in Fig. 1, namely that they were homogeneous with *Bf. animalis* subsp *lactis* DSM 10140[T] [AY722379] (Fig. 2A), *Bf. animalis* subsp *lactis* LMG 18314T [DQ234423] (Fig. 2B), and *Bf. animalis* subsp *lactis* LMG 18314T [DQ234376] (Fig. 3C), respectively. Phylogenetic trees constructed using *clpC*, *purF* and *dnaG* gene sequences clearly differentiated the isolates, identified as *Bf. animalis* subsp. *lactis*, from *Bf. animalis* subsp *animalis* ATCC 25527[T], the designated type strain. Ventura et al. (2006) reported the difficulty in differentiating subspecies of *Bf. animalis* due to their great homogeny. This study confirmed the power of *clpC*, *purF* and *dnaG* molecular markers in differentiating between *Bf. animalis* subspecies.

However, *dnaB* and *rpoC* gene markers (Fig. 3D and Supplementary Fig. 4, respectively) were less effective in associating the *Bifidobacterium* isolates with *Bf. animalis* subsp. *lactis* type strains in a manner that *clpC*, *purF* and *dnaG* molecular markers did. This was despite the observation that the isolates clustered together homogeneously in the *dnaB* and *rpoC* phylogenetic trees (Fig. 3D and Supplementary Fig. 4, respectively). Hence, a multigenic approach is relevant in the identification and classification of closely related strains. Had *dnaB* and *rpoC* been the only chromosomal genes targeted in this study, the conclusions would have been inaccurate. The *dnaB* and *rpoC* genes ambiguously and inexplicably failed to provide phylogenies, for isolates, similar to topologies provided by the other protein-coding gene markers used in the study.
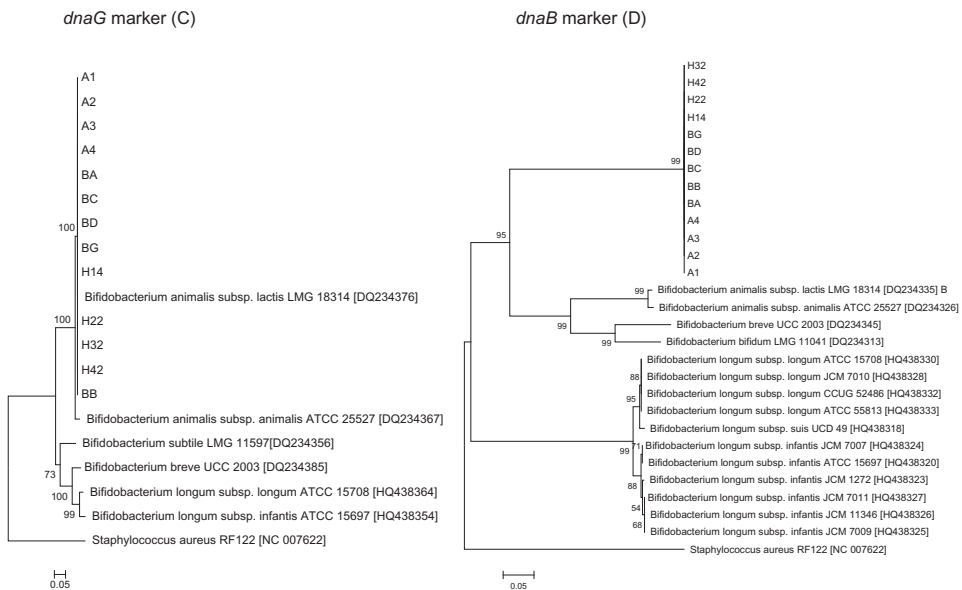
Figure 3. The phylogenetic trees for *Bifidobacterium* isolates constructed with *dnaG* gene (C) and *dnaB* gene (D) sequences using the Neighbor-Joining method and the p-distance model. The bootstrap values are indicated at the internodes. Accession numbers (in parentheses) are for the sequences retrieved from the GenBank. *Staphylococcus aureus* RF 122 was used to root the trees.
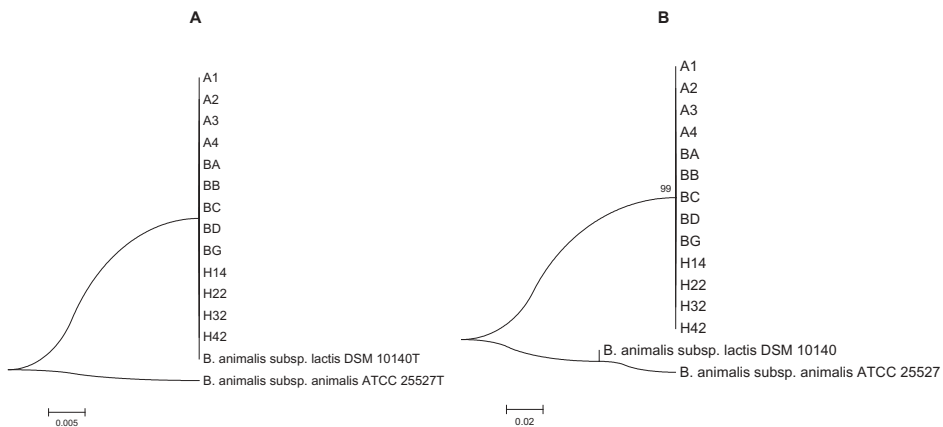

Figure 4. The curved traditional phylogenetic trees, for *Bifidobacterium* isolates and two designated type strains, constructed from the concatenated sequences of *dnaG, clpC* and *purF* genes (in the case of Fig. 4A) a n d *dnaG, clpC, purF, rpoC* and *dnaB* genes (in the case of Fig. 4B). The phylogenetic trees were drawn using the neighbour-joining method and the Jukes-Cantor model as the substitution model. Bar, 0.005 and 0.02 were base substitutions per site. The phylogenetic trees were confirmed by the maximum likelihood method and the Tamura-Nei substitution model.

Determining the evolutionary divergences between the strains

Table 2 shows that the 16S rRNA gene and *clpC, purF,* and *dnaG* gene markers, which effectively discriminated between isolates, had 0.0008, 0.0085, 0.0071, and 0.0044, respectively as overall mean evolutionary divergence (in base pairs). Although *dnaG* marker was of lower discriminatory power compared to *purF* and *clpC* gene markers, it was more than 5-fold better than the 16S rRNA gene marker in discriminating between *Bf. animalis* subspecies. The *dnaB and rpoC* gene markers, which had overall mean evolutionary divergences of 0.2252 and 0.1748, respectively, could erroneously be seen as being better at differentiating the *Bififdobacterium* isolates. However, as already stated, the isolates did not precisely cluster with the designated type strains in the *dnaB and rpoC* phylogenetic trees (Fig. 3D and Supplementary Fig. 4) and as such could not be as effective as the *clpC, purF* and *dnaG* gene markers on the basis of mean evolutionary divergence.

Concatenation of *clpC, purF,* and *dnaG* gene sequences resulted in an overall mean evolutionary divergence of 0.0070, which was more than 8-fold better than the overall mean evolutionary divergence of 0.0008 provided by 16S rRNA gene. These findings were confirmed by Tajima's neutrality test results. The higher the mean divergence (equivalent to nucleotide diversity (π)), the better the discriminative power except for the *rpoC* and *dnaB* genes and/or the concatenation that included these genes' sequences (Table 2). Similarly, Tajima's Ps values (Table 2) were equivalent to the pairwise distance (Supplementary Table 3) between sequences of the same gene. The guanine-cytosine (G + C) content of the *Bifidobacterium* strains was in the range of 59–65.29%. The average G + C content for bifidobacteria is 55–67% (Lee and O'Sullivan, 2010), implying that the genes were not acquired recently. The three concatenated protein-coding genes (*clpC, PurF,* and *dnaG*) had a better selection potential than the 16S rRNA gene to purify the strains on the basis of a more negative Tajima test statistic (D).

When the *dnaB* and *rpoC* gene sequences were included, along with the other three, in the concatenation (Fig. 4B, Table 2), the overall mean evolutionary divergence was erroneously raised to 0.0528. However, this was largely attributed to the bigger pairwise distances between the *dnaB* or *rpoC* gene sequences of the type strains and the sequences of the studied isolates for the same genes (Supplementary Table 3). In Supplementary Table 3, isolates assigned numbers 1–13 were too homologous with *Bf. animalis* subsp. *lactis* designated type for each gene. Hence, only strains numbered 14 and 15 were included in Supplementary Table 3 as they were observed to be the centre of variance. Supplementary Table 3 shows that, in the case of *dnaB* and *rpoC* genes, the evolutionary divergences were 0.9006 and 0.7031, respectively, between sequences of the isolates and the *B. animalis* subsp. *lactis* designated type strains. The sequences of the type strain

)

Table 2. Evolutionary divergence estimates and application of Tajima's neutrality test in descriptive analysis of data used in MLSA.

Gene	Final data set size (bp)	Mean divergence	G + C Content (mol %)	Tajima's Neutrality Test results					
				M	*S*	*Ps*	θ	π	*D*
clpC	608	0.0085	59.00	15	37	0.060855	0.018716	0.008114	−2.40945
dnaG	370	0.0044	65.10	15	12	0.032432	0.009974	0.004324	−2.20684
purF	446	0.0071	61.22	15	23	0.051570	0.015860	0.006876	−2.34457
dnaB	206	0.2252	61.98	15	110	0.533981	0.164223	0.130559	−0.89987
rpoC	675	0.1748	65.26	15	313	0.463704	0.142610	0.113354	−0.91035
X*	1424	0.0070	61.45	15	72	0.050562	0.015550	0.006742	−2.46531
XX**	2277	0.0528	62.29	15	462	0.202899	0.062400	0.045988	−1.16944
16S rRNA	570	0.0008	60.22	16	2	0.003509	0.001057	0.000819	−0.57783

X*, concatenated protein-coding genes' sequences excluding *dnaB* and *rpoC* gene sequences of isolates and type strains.
XX**, concatenation of all the chromosomal gene sequences of isolates and type strains.
m = number of sequences, S = Number of segregating sites, *Ps* = S/m, θ = *Ps*/a_1, π = nucleotide diversity *and D* is the Tajima test statistic.

for *B. animalis* subsp. *animalis* diverged from each of the isolate sequences by 0.9169 and 0.7069 in the case of the *dnaB* and *rpoC* genes respectively. This implied divergences of 0.0147 and 0.0195 between the two type strains on the basis of the *dnaB* and *rpoC* gene sequences, respectively. The foregoing respective divergences were different from the divergences recorded between sequences of isolates and *Bf. animalis* subsp. *lactis* designated type strains for the same genes. This shed light on the failure of the *dnaB* and *rpoC* genes to provide phylogenies in which isolates effectively associated with *Bf. animalis* subsp. *lactis* as was the case with *clpC, purF* and *dnaG* genes (see concatenated phylogenetic tree, Fig. 4A). With the latter three genes, the sequences of the isolates and *Bf. animalis* subsp. *lactis* designated type strains were too homogeneous (zero divergence distance, Supplementary Table 3). This implied that the divergence between sequences of isolates and *Bf. animalis* subsp. *animalis* was equivalent to divergence between the two designated type strains for each of the chromosomal genes targeted in the study, except *dnaB* and *rpoC* genes. To avoid and/or detect errors such as those presented by the *dnaB* and *rpoC* gene sequences of the type strains, it was important to first construct individual phylogenetic trees for each of the genes targeted. Such prior analysis would reveal sources from which eventual taxonomic and clustering challenges in the phylogenetic trees of the concatenated sequences would emanate.

Figure 4 shows the curved phylogenetic trees for *Bifidobacterium* isolates and two designated type strains. Constructed trees from the concatenated sequences of *dnaG, clpC* and *purF* genes (Fig. 4A) and the additional *dnaB* and *rpoC* gene sequences (Fig. 4B) are clearly illustrated. In Fig. 4A, the isolates clustered with *B. animalis* subsp. *lactis* type strain, confirming their identification, as was the case with the 16S rRNA gene tree. However, in Fig. 4B, the isolates did not cluster with the designated type strains, confirming the failure of the *dnaB* and *rpoC* gene sequences to confer precise identities to the isolates.

Association of the isolates with Bf. animalis *subsp.* lactis

On the basis of the 16S rDNA sequences, the phylogenetic analysis demonstrated that the *Bifidobacterium* strains clustered with *Bf. animalis* subsp. *lactis* strain YIT 4121, a synonym of *Bf. animalis* subsp. *lactis* DSM 10140, the designated type strain (https://www.ncbi.nlm.nih.gov/nuccore /15823597). The use of 16S rRNA gene has the advantage that they appear as several copies in a bacterial cell unlike the housekeeping genes (Ventura et al., 2001). Hence, the initial phylogeny of the presumptive Bifidobacterium strains in this study was based on the 16S rDNA sequences, as Holzapfel et al. (2001) recommended. However, the slow evolution of 16S rRNA genes creates limitations when identifying closely related species or strains

(Holzapfel et al., 2001; Ventura & Zink, 2003). In this study though, the phylogenetic tree based on 16S rDNA sequences demonstrated that the strains, identified as *Bf. animalis* subsp. *lactis* were distinguished from *Bf. animalis* subsp. *animalis* JCM 1190 (=ATCC 25527[T], Dong et al., 2000) (see Fig. 1), even though a close relatedness was observed.

The phylogenetic tree based on 16S rRNA gene sequences unexpectedly yielded additional revelations. Isolate BI and BJ, assigned to the *Pediococcus pentosaceus* taxon, would not usually be expected to be included in probiotic products albeit they are LAB. It was also observed that *Bf. animalis* subsp. *animalis* strains JCM 1253 [AB507071] and JCM 7117 [AB507072] retrieved from the GenBank, clustered with the strains identified as *Bf. animalis* subsp. *lactis*. This could be ascribed to a hypothesis that the sequences of the two strains, JCM 1253 and JCM 7117, were possibly inaccurately registered as *Bf. animalis* subsp. *animalis* at the time that their sequences were deposited at the nucleotide database. Similar incorrect sequence registrations and taxonomic challenges have been encountered in a plethora of other research outcomes. It was for example discovered that *Bf. pseudolongum* ATCC 25526[T] (M58742) and *Bf. indicum* ATCC 25912[T] (M58737) were phylogenetically closely related to *Bf. longum* and *Bf. infantis* respectively due to incorrect registration of sequences (Dong et al., 2000).

In a study by Ventura et al. (2001), *Bf. animalis* ATCC 27536 was observed to be more similar to *Bf. lactis* DSM 10140 than to its type strain, *Bf. animalis* ATCC 25527, on the basis of pulsed-field gel electrophoresis patterns. Two years later, the potential of housekeeping genes to resolve taxonomic challenges was revealed when certain *Bf. animalis* strains (ATCC 27536, ATCC 27674 and ATCC 27673) were observed to cluster with type strain *Bf. lactis* DSM 10140 on the basis of their *tuf* and *recA* gene sequences (Ventura and Zink, 2003). In the same study, however, it was seen that *Bf. animalis* ATCC 27672 and *Bf. animalis* ATCC 25527[T] formed a cluster that was separate from that of *Bf. lactis*. Some of the foregoing reported taxonomic and phylogenetic challenges are attributed to incorrect sequence naming and/or mislabeling, an issue that may require, as shown in this study, further investigation involving a multigenic approach. In support of the need of the foregoing suggestion, different authors (Marteau et al., 2002; Van der Meulen et al., 2004; Collado et al., 2006; Mayer et al., 2007; Tabbers et al., 2009) found more variation in labelling and/or nomenclature of the same strains. However, over the four months period of product sampling in this study, the identity of isolates remained the same.

All isolates' sequences were 100% homogeneous with the type strain (DSM 10140[T]) sequence for *Bf. animalis* subsp. *lactis* and the control, *Bf. animalis* subsp. *lactis* Bb12 (denoted as BB) (Garrigues et al., 2010). Mayer et al.

([2007](#)) used RAPD-PCR and PFGE techniques to show that *Bf. animalis* strains (DN-173010, Danone; Bb12, Chr. Hansen; DSM 21105, ATCC 27536; CB 120, SKW Bio-systems; LMG 18906, ATCC 27674) were initially not assigned to the *Bf. lactis* taxon as they should have been. In the present study, strain Bb12 was included as the reference strain.

Identification based on protein-coding gene sequences

Compared with 16S rDNA, chromosomal DNA is more unstable and discriminates more effectively between closely related species and strains (Holzapfel et al., [2001](#)). Hence, the phylogeny based on the 16S rRNA gene sequences was complemented with phylogenetic analyses of DNA sequences of multiple housekeeping genes in a so-called multilocus sequence analysis (MLSA) approach. The MLSA approach is one of the best techniques that can be used for the phylogenetic analysis of strains of closely related species (Delétoile et al., [2010](#)). Accuracy in probiotic strain identification is imperative during probiotic food product development as well as in clinical trials (Su et al., [2005](#)). The MLSA approach was also applied in the present study because the precisely identified strains were intended to be included in a probiotic fermented maize beverage, in later work. The multigene approach, which in this study involved five protein-coding genes, resulted in the precise identification and classification of strains.

Among the five housekeeping genes used in this study, topologies of phylogenies based on *clpC, purF,* and *dnaG* gene sequences were comparatively similar to the topology of the phylogeny based on 16S rDNA sequences. Phylogenies based on *dnaB* and *rpoC* gene sequences, however, presented topologies that were not similar to those of the 16S rDNA sequences and the other chromosomal gene sequences. In the case of the *dnaB* and the *rpoC* phylogenies, the studied strains clustered together, but not with *Bf. animalis* subsp. *lactis* LMG 18314 (= DSM 10140$^\text{T}$, the designated type strain). This requires further investigation as the *xfp* molecular marker similarly could not amplify the targeted gene of the strains under the many parameters and conditions attempted. It was nevertheless clear that the *clpC, purF* and *dnaG* gene loci, individually as well as in their concatenated sequences, had, 8-fold more discriminative power than the 16S rDNA sequences ([Table 2](#) and Supplementary Table 3).

The *Bifidobacterium* isolates in this study were confirmed as *Bf. animalis* subsp. *lactis,* separate from, but closely related to, *Bf. animalis* subsp. a*nimalis* when *clpC, purF* and *dnaG* gene sequences were subjected to phylogenetic analysis. Ventura et al. ([2005](#)) made a similar observation when they used *dnaJ* gene sequences for the phylogenetic analysis of different bacteria. Similarly, other workers (Mayer et al., [2007](#)) observed that *Bf. animalis* subsp. *lactis* strains from different commercial probiotic dairy products,

were genetically homogeneous when compared to type and reference strains on the basis of DNA fingerprints by RAPD-PCR even though *Bifidobacterium* strain LAFTI B94 was slightly different due to an extra band in its RAPD profile. In the present study, strain LAFTI B94 (isolate BA) did not exhibit any differences from the other *B. animalis* subsp. *lactis* strains.

The findings in the present study regarding phylogenetic clustering and evolutionary divergence, based on the *clpC, purF,* and *dnaG* gene sequences, of *Bf. animalis* and *Bf. lactis* strains are in agreement with the findings of several workers (Masco et al., 2004; Ventura et al., 2006). It was reported that, on the basis of 16S rDNA sequences, type strains *Bf. animalis* ATCC 25527[T] and *Bf. lactis* DSM 10140[T] were highly homologous (98.8% similarity) and could not be clearly distinguished (Masco et al., 2004). Against this background, Table 2 shows that concatenated sequences of *clpC, purF* and *dnaG* chromosomal genes were more than 8-fold more effective than the 16S rRNA gene sequences in discriminating between *Bf. animalis* and *Bf. lactis* strains. To the best of our knowledge, this is so far the highest mean discriminative power provided for *Bf. animalis* and *Bf. lactis* strains by concatenated sequences of *clpC, purF* and *dnaG* genes in comparison with 16S rDNA sequences. In a study by Ventura et al. (2006), if the same genes were to be considered, a simple calculation would indicate that the concatenated *clpC, purF* and *dnaG* genes would be 4.1 to 4.8 times more discriminative than the 16S rRNA gene. The discrepancy in the discriminatory power would be attributed to the variety of species included in the study by Ventura *et al.* (2006) while in the present study only strains of *Bf. animalis* subsp. *lactis* and *Bf. animalis* subsp. a*nimalis* were considered. In another study, the average evolutionary divergence (K_{nue} value) for the 16S rDNA sequences of *Bifidobacterium* strains was found to be 3-fold less than the mean value obtained from *tuf* and *recA* housekeeping genes gene sequences (Ventura and Zink, 2003). Some workers have reported success in discriminating between closely related bifidobacteria (Ventura and Zink, 2003; Masco et al., 2004). However, very few authors present measurable degrees of evolutionary divergence between strains, enabled by the specific genes employed, as is the case in the present study which involved closely related *Bifidobacterium* strains. The superiority of the multiple protein-coding genes over the 16S rRNA gene, to accurately discriminate between closely related strains, has been noted by several workers and taken advantage of in resolving taxonomic challenges (Holzapfel et al., 2001; Masco et al., 2004; Thompson et al., 2005; Ventura et al., 2005; Singh et al., 2009; Delétoile et al., 2010).

In the present study, phylogenetic analysis of *clpC, purF* and *dnaG* gene sequences led to distinct clusters of *Bf. animalis* subsp. *lactis* and *Bf. animalis* subsp. *Animalis,* respectively, even though they are closely related. In this study, the observed close relatedness between the two clusters strengthens the

suggestion of a single taxon separated at the subspecies level as recommended by other workers (Ventura and Zink, 2003; Masco et al., 2004). It is widely accepted that probiotic and technological properties are strain-dependent. A multigenic approach in the present study provided a definite identification of the isolates. The identification also confirmed that *Bf. animalis* subsp. *lactis* may be the most used *Bifidobacterium* strain in marketed probiotic dairy products (Su et al., 2005; Collado et al., 2006; Barrangou et al., 2009), perhaps due to technological reasons. *Bifidobacterium lactis* for example tolerates oxygen better than *Bf. animalis* which enables the former species to attain higher population levels in probiotic products (Ventura et al., 2001; Masco et al., 2004). In addition to oxygen tolerance, *Bf. lactis* also has better acid tolerance (Ventura et al., 2001), a property that is necessary for survival at higher than the required therapeutic levels during storage in high acid products.

Conclusions

The multigenic approach in this study enabled precise and accurate identification of closely related *Bifidobacterium* isolates from probiotic milk beverages, culture collections, and pharmaceutical probiotic supplements.

Some isolates, including BE, BF, BH, BK, BI, and BJ, which were obtained using bifidobacteria selective medium (BSM) containing mupirocin (50 mg/L), turned out to be non-*Bifidobacterium* isolates on the basis of 16S rRNA gene sequencing. The amount of selective agent, mupirocin, incorporated in BSM, it appears, may not fully inhibit *Lactobacillus* and other LAB strains.

Concatenation of sequence data from the protein-coding genes (*clpC, PurF* and *dnaG* genes) used in this study resulted in a more reliable and discriminative concatenated tree compared to the 16S rRNA gene tree in terms of mean evolutionary divergence. The latter gene was more than eight-fold less discriminative than the concatenated data of the *clpC, PurF* and *dnaG* genes.

The concatenated tree from the data involving *rpoC* and *dnaB* genes could not provide reliable phylogenetic relationships for the isolates and could possibly conceal problems emanating from the individual genes. Constructing individual gene phylogenetic trees prior to sequence concatenation is recommended in order to detect sources of conflict in constructing a concatenated phylogenetic tree, as was the case in this study. The use of *rpoC* and *dnaB* gene markers therefore cannot be recommended on the basis that the isolates were shown not to associate with *Bf. animalis* subsp *lactis* designated type strains when these markers were used.

In envisaged future studies, the identified and classified isolates will be investigated for their technological suitability, safety and the potential to impart health benefits prior to their use in the development of cereal-based probiotic beverages.

Funding

References

Ashraf, R., Shah, N.P. (2011). Selective and differential enumeration of *Lactobacillus delbrueckii* subsp. *bulgaricus, Streptococcus thermophilus, Lactobacillus acidophilus, Lactobacillus casei* and *Bifidobacterium* spp. in yoghurt – a review. *Int. J. Food Microbiol.* 149:194–208.

Barrangou, R., Briczinski, E.P., Traeger, L.L., Loquasto, J.R., Richards, M., Horvath, P. (2009). Comparison of complete genome sequences of *Bifidobacterium animalis* subsp. *lactis* DSM 10140 and B1-04. *J. Bacteriol.* 191:4144–4151.

Bevilacqua, L., Ovidi, M., Di Mattia, E., Trovatelli, L.D., Canganella, F. (2003). Screening of *Bifidobacterium* strains isolated from human faeces for antagonistic activities against potentially bacterial pathogens. *Microbiol. Res.* 158:179–185.

Bunesova, V., Musilova, S., Geigerova, M., Pechar, R., Rada, V. (2015). Comparison of mupirocin-based media for selective enumeration of bifidobacteria in probiotic supplements. *J. Microbiol. Meth.* 109:106–109.

Buruleanu, C.L., Nicolescu, C.L, Avram, D., Manea, I., Bratu, M.G. (2012). Effects of yeast extract and different amino acids on the dynamics of some components in cabbage juice during fermentation with *Bifidobacterium lactis* BB-12. *Food Sci. Biotech.* 21 (3):691–699.

Chang, S.Y., Kim, D.-Y., Han, M.J. (2010). Physicochemical and sensory characteristics of soy yogurt fermented with *Bifidobacterium breve* K-110, *Streptococcus thermophilus* 3781, or *Lactobacillus acidophilus* Q509011. *Food Sci. Biotech.* 19(1):107–113.

Collado, M.C., Moreno, Y., Cobo, J.M., Mateos, J.A., Harnández, M. (2006). Molecular detection of *Bifidobacterium animalis* DN-173010 in human feces during fermented milk administration. *Food Res. Int.* 39:530–535.

Dale, J.W., Park, S.F. (2010). *Molecular Genetics of Bacteria*. Hoboken, NJ: John Wiley and Sons.

Delétoile, A., Passet, V., Aires, J., Chambaud, I., Butel, M., Smokvina, T., Brisse, S. (2010). Species delineation and clonal diversity in four *Bifidobacterium* species as revealed by multilocus sequencing. *Res. in Microbiol.* 161:82–90.

Dong, X., Xin, Y., Jian, W., Liu, X., Ling, D. (2000). *Bifidobactreium thermocidophilum* sp. Nov., isolated from an anaerobic digester. *Int. J. System. Evol. Microbiol.* 50:119–125.

Garrigues, C., Johansen, E., Pedersen, M.B. (2010). Complete genome sequence of *Bifidobacterium animalis* subsp. *lactis* Bb-12, a widely consumed probiotic strain. *J. Bacteriol.* 192:2467–2468.

Gavin, F., Van Esbroeck, M., Touzel, J.P., Fourment, A., Goossens, H. (1996). Detection of fructose-6-phosphate phosphoketolase (F6PPK), a key enzyme of the bifid-shunt, in *Gardnerella vaginalis. Anaerobe* 2:191–193.

Guo, X.H., Kim, J.M., Nam, H.M., Park, S.Y., Kim, J.M. (2010). Screening lactic acid bacteria form swine origins for multistrain probiotics based on *in vitro* functional properties. *Anaerobe* 16:321–326.

Holzapfel, W.H., Haberer, P., Geisen, R., Björkroth, J., Schillinger, U. (2001). Taxonomy and important features of probiotic microorganisms in food and nutrition. *Amer. J. Clin. Nutr.* 73(suppl):365S–373S.

Kim J.Y., Wang Y., Park S.J., Ji G.E., Park M.S. (2012). Cloning and expression of β-glucosidases from *Bifidobacterium lactis* AD011. *Food Sci. Biotechnol.* 21(3): 731–738.

Koh, E., Jang, K-H., Surh, J. (2014). Improvement of physicochemical properties of cereal based ready-to-eat *sunsik* using fermentation with *bionuruk* and *Bifidobacterium longum*. *Food Sci. Biotechnol.* 23(6):1977–1985.

Lee, J.-H., O'Sullivan, D.J. (2010). Genomic insights into bifidobacteria. *Microbiol. Mol. Biol. Rev*. 74(3):378–416.

Leuschner, R.G.K., Bew, J., Simpson, P., Ross, P.R., Stanton, C. (2003). A collaborative study of a method for the enumeration of probiotic bifidobacteria in animal feed. *Int. J. Food Microbiol.* 83:161–170.

Marteau, P., Cuillerier, E., Meance, S., Gerhardt, M.F., Myara, A., Bouvier, M., Bouley, C., Tondu, F., Bommelaer, G., Grimaud, J.C. (2002). *Bifidobacterium animalis* strain DN-173010 shortens the colonic transit time in healthy women: a double-blind, randomized, controlled study. *Alim. Pharm. Therapeutics* 16:587–593.

Masco, L., Ventura, M., Zink, R., Huys, G., Swings, J. (2004).Polyphasic taxonomic analysis of *Bifidobacterium animalis* and *Bifidobacterium lactis* reveals relatedness at thesubspecies level: reclassification of *Bifidobacterium animalis* as *Bifidobacterium animalis* subsp. *animalis* subsp. nov and *Bifidobacteriumanimalis* subsp. *lactis* subsp. nov. *International J. System. Evol. Microbiol.* 54:1137–1143.

Mayer, H., Amtmann, E., Philippi, E., Steinegger, G., Mayrhofer, S., Kneifel, W. (2007). Molecular discrimination of new isolates of *Bifidobacterium animalis* subsp. *lactis* from reference strains and commercial probiotic strains. *Int. Dairy J.* 17:565–573.

Nyanzi, R. (2007). *Enhancing the Functional Quality of Mageu*. (MTech dissertation). Tshwane University of Technology, South Africa.

Orban, J.I., Patterson, J.A. (2000). Modification of the phosphoketolase assay for rapid identification of bifidobacteria. *J. Microbiol. Meth.* 40(3):221–224.

Sadaghdar, Y., Mortazavian, A.M., Ehsani, M.R. (2012). Survival and activity of 5 probiotic lactobacilli strains in 2 types of flavoured fermented milk. *Food Sci. Biotechnol.* 21(1):151–157.

Singh, S., Goswami, P., Singh, R., and Heller, K.J. (2009). Application of molecular identification tools for *Lactobacillus*, with a focus on discrimination between closely related species: a review. *LWT–Food Sci. Tech.* 42:448–457.

Su, P., Henriksson, A., Tandianus, J.E., Park, J.H., Foong, F., Dunn, N.W. (2005). Detection and quantification of *Bifidobacterium lactis* LAFTI[®] B94 in human faecal samples from a consumption trial. *FEMS Microbiol. Lett.* 244:99–103.

Tabbers, M.M., Chmielewska, A., Roseboom, M.G., Boudet, C., Perrin, C., Szajewska, H., Benninga, M.A. (2009). Effect of consumption of a fermented dairy product containing *Bifidobacterium lactis* DN-173010 on constipation in children: a multicentre randomized controlled trial (NTRTC: 1571). *BioMed Cent. Ped.* 9:22.

Tamura, K., Peterson, N., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011). MEGA5: molecular evolutionary Genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol. Biol. and Evol* 28(10):2731–2739.

Thompson, F.L., Gevers, D., Thompson, C.C., Dawyndt, P., Naser, S., Hoste, B., Munn, C.B., Swings, J. (2005). Phylogeny and molecular identification of vibrios on the basis of multi-locus sequences analysis. *App. Environ. Microbiol.* 71(9):5107–5115.

Van der Meulen, R., Avonts, L., De Vuyst, L. (2004). Short fractions of oligofructose are preferentially metabolized by *Bifidobacterium animalis* DN-173010. *App. Environ Microbiol.* 70:1923–1930.

Vasconcelos, B.G., Martinez, R.C.R., de Castro, I.A., Saad, S.M.I. (2014). Innovative açaí (*Euterpe oleracea, Mart., Arecaceae*) functional frozen dessert exhibits high probiotic viability throughout shelf-life and supplementation with inulin improves sensory acceptance. *Food Sci. Biotechnol.* 23(5):1843–1849.

Ventura, M., Zink, R. (2003). Comparative sequence analysis of the tuf and recA genes and restriction fragment length polymorphism of the internal transcribed spacer region sequences supply additional tools for discriminating *Bifidobacterium lactis* from *Bifidobacterium animalis. Appl. Environ. Microbiol.* 69:7517–7522.

Ventura, M., Canchaya, C., Bernini, V., Del Casale, A., Dellaglio, F., Neviani, E., Fitzgerald, G. F., Van Sinderen, D. (2005). Genetic characterization of the *Bifidobacterium breve* UCC 2003 *hrcA* locus. *App. Environ. Microbiol.* 71(12):8998–9007.

Ventura, M., Canchaya, C., Bernini, V., Del Casale, A., Dellaglio, F., Neviani, E., Fitzgerald, G. F., Van Sinderen, D. (2006). Analysis of bifidobacterial evolution using a multilocus approach. *Int. J. System. Evol. Microbiol.* 56:2783–2792.

Ventura, M., Reniero, R., Zink, R. (2001). Specific Identification and targeted characterization of *Bifidobacterium lactis* from different environmental isolates by a combined multiplex-PCR approach. *Appl. Environ. Microbiol.* 67:2760–2765.