Food Microbiology 60 (2016) 137–141

Contents lists available at ScienceDirect

Food Microbiology

journal homepage: www.elsevier.com/locate/fm

A peptidome-based phylogeny pipeline reveals differential peptides at the strain level within *Bifidobacterium animalis* subsp. *lactis*

Aitor Blanco-Míguez [a], Alberto Gutiérrez-Jácome [a], Florentino Fdez-Riverola [a], Anália Lourenço [a, b], Borja Sánchez [c, *]

[a] ESEI – Department of Computer Science, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain
[b] CEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[c] Department of Microbiology and Biochemistry of Dairy Products, Dairy Research Institute (IPLA-CSIC), Paseo Río Linares s/n, 33300 Villaviciosa, Asturias, Spain

ARTICLE INFO

Article history:
Received 7 February 2016
Received in revised form
23 May 2016
Accepted 26 June 2016
Available online 29 June 2016

Keywords:
Bifidobacterium animalis subsp. *lactis*
Peptidome
Specific peptides

ABSTRACT

Bifidobacteria are gut commensal microorganisms belonging to the Actinobacteria group. Some specific strains of *Bifidobacterium animalis* subsp. *lactis* are used in functional foods as they are able to exert health-promoting effects in the human host. Due to the limited genetic variability within this subspecies, it is sometimes difficult for a manufacturer to properly track its strain once included in dairy products or functional foods. In this paper, we present a peptidome-based analysis in which the proteomes of a set of *B. animalis* subsp. *lactis* strains were digested *in silico* with human gut endopeptidases. The molecular masses were compared along all the strains to detect strain-specific peptides. These peptides may be interesting towards the development of methodologies for strain identification in the final product.

Bifidobacteria are members of the, one of the main microbial taxa of the human gut microbiota. Their abundance varies depending on age, being one of the most prevalent microorganisms in newborns and infants. In addition, members of the species *Bifidobacterium longum*, *Bifidobacterium breve*, *Bifidobacterium bifidum* and *Bifidobacterium animalis* have been used as probiotics due to the scientific evidence of specific beneficial effects on the human host (Hill et al., 2014).

Within *B. animalis*, *B. animalis* subsp. *lactis* receives particular attention, due to its capacity to grow and survive in milk, and some strains such as CNCM I-2494 have been included as adjunct cultures of fermented milks since 1987 (Chervaux et al., 2011). This subspecies is by far the bifidobacteria more used in functional food products (Gueimonde et al., 2004), and it is usually the sole viable bifidobacteria species in fermented milks (Jayamanne and Adams, 2006). This is mainly due to the intrinsic resistance of this species to both technological and physiological stresses (Jayamanne and Adams, 2006), being moderately aerotolerant and able to adapt to both acidic and bile stress (Ruiz et al., 2012; Sánchez et al., 2007). Among the reported benefits of the strains of this species are decrease of serum cholesterol levels, protection against colorectal cancer, regulation of the gut transit time and constipation, and reduction of gut inflammation by maintaining a favorable balance of the microbiota (Tabbers et al., 2011; Veiga et al., 2010). In addition, *B. animalis* subsp. *lactis* is the bifidobacteria with higher number of clinical studies supporting its health-promoting attributes in gastrointestinal disorders and allergic processes (Agrawal et al., 2009; Isolauri et al., 2000; Weizman et al., 2005).

Multilocus sequence typing (MLST) showed that *Bifidobacterium animalis* is phylogenetically distant from other species of the genus (Deletoile et al., 2010). Indeed, strains belonging to *B. animalis* subsp. *lactis* are difficult to differentiate due to the high degree of genome identity among strains (99.975%) (Lomonaco et al., 2015). Proper strain identification is a very valuable trait for both producers and consumers, as close probiotic strains may have different effects on host health, notably at the immunomodulation level (Hill et al., 2014).

In recent years, multiple efforts have been focused on developing methods for strain differentiation within *B. animalis* subsp. *lactis*. First attempts included application of relatively classical techniques, such as pulse-field gel electrophoresis, restriction fragment length polymorphisms or MLST using housekeeping genes such as *gyrB* or *rpoB* (Briczinski and Roberts, 2006; Deletoile et al., 2010; Ventura and Zink, 2002). Later, the public availability of

* Corresponding author.
E-mail address: borja.sanchez@csic.es (B. Sánchez).

http://dx.doi.org/10.1016/j.fm.2016.06.015
0740-0020/

multiple genome sequences enabled the use of other genotypic methods based on single nucleotide polymorphisms and insertion-deletion polymorphisms (Briczinski et al., 2009; Lomonaco et al., 2015; Tmanova et al., 2012). Alternatively, a number of mass spectrometry strategies have also been proposed in order to offer a more rapid analysis, higher reproducibility and more accurate results at species and strain levels (PMID: 26510657, 26537565, 26300860). In particular, previous works show that protein biomarkers are able to accurately differentiate *B. animalis* strains at the subspecies level (PMID: 22365357, 22417598, 21598393).

In this work, we evaluate the discriminative availability of an *in silico* proteomic analysis of *B. animalis* strains, i.e. a fully computational analysis based on public mass spectra and free software. The aim is to be able to generate a valid and manageable list of potentially specific peptides for each strain, which can be further investigated using *in vitro* approaches, such as LC-MS/MS, towards the identification of biomarkers and the development of application-specific detection methods.

We have obtained the bacterial proteomes of all the *B. animalis* subsp. *lactis* strains whose genomes are completely sequenced and available at the National Center for Biotechnology Information (NCBI) servers (Table 1). All the sequence data used in this study were retrieved from the BioProject collection of the NCBI, using the public FTP site (ftp://ftp.ncbi.nih.gov/genomes/bacteria/) (Wheeler et al., 2000). Peptidomes were basically obtained through the following pipeline: i) retrieval of proteins encoded in the publicly available and completely sequenced genomes, ii) *in silico* digestion of proteins using human gut endopeptidases, and iii) comparison of the peptides according to their theoretical mass (peaks) and subsequent computation of consensus peak sets.

The subcellular localizations of the proteins were predicted using the standalone version of the PSortB v3.0 tool, following the developer guidelines (Yu et al., 2010). The molecular weight of the peptides and isoelectric point of the proteins were calculated using in-house customised scripts. Peptidomes were generated for each strain using the open-source Java library mzJava from ExPASy (http://mzjava.expasy.org), which supports *in silico* protein digestion (Horlacher et al., 2015). For the purposes of the present analysis, three proteases representing the major intestinal endoproteases were used: trypsin, chymotrypsin and pepsin (low specificity model, ph > 2).

In order to reduce the proteomic data input we sampled the peptidomes according to different parameters such as peptide length, subcellular localization of the source protein and its isoelectric point (see Supplementary Material 1). Two different datasets or subpeptidomes were used in our study: i) peptides longer than 50 amino acids obtained from cytoplasmic proteins included in the pI range 4.5–5.5 (Cyto_50_more dataset), and ii) peptides obtained from extracellular proteins (Extracellular dataset). The rationale behind the selection of these two subpeptidomes was XXX. These datasets corresponded to about 1000 different peptides per strain (see Supplementary Material 2). These datasets were submitted to an in-house compiled version of SPECLUST (Alm et al., 2006), which enabled the identification of representative and reproducible peak masses in all spectral profiles. SPECLUST calculates the mass difference between two peaks taken from different peak lists and determines if the two peaks are identical taking into account some measurement uncertainty (σ; set empirically to 3.0 Da). In addition, a pairwise cut-off parameter determines whether a peptide is shared between two spectra. In our work, a peak match score greater than 0.6 (corresponding to a 0.5 Da mass difference) was set to consider two peak masses to be the same in two different profiles.

Using this methodology a consensus table summarizing the shared and differential peptides among the strains was obtained. The consensus mass peptide matrix was translated into a binary matrix (0's and 1's, representing absence or presence of a given peptide mass respectively) in NEXUS file format. An in-house script allowed us to retrieve the sequence of any differential peptide from its position in the matrix, as well as the source protein and bifidobacterium strain (see Fig. 1).

MrBayes, the model-based phylogenetic inference tool based in Bayesian statistics, was utilized to generate consensus trees from the NEXUS files (Huelsenbeck and Ronquist, 2001). Phylogeny was inferred through the restriction data type implemented in MrBayes (with state 0 or 1 representing the absence or presence of a consensus peptide throughout all the strain peptidomes). A majority-rule consensus tree (50%) was obtained after discarding

Table 1
Bifidobacterium strains used in this study. Genome and protein data were retrieved from the BioProject collection of the NCBI.

Strain	BioProject
Bifidobacterium adolescentis ATCC 15703	PRJNA16321
Bifidobacterium animalis subsp. *lactis* AD011	PRJNA19423
Bifidobacterium animalis subsp. *lactis* ATCC 27673	PRJNA215974
Bifidobacterium animalis subsp. *lactis* B420	PRJNA156973
Bifidobacterium animalis subsp. *lactis* BB-12	PRJNA42883
Bifidobacterium animalis subsp. *lactis* BLC1	PRJNA71815
Bifidobacterium animalis subsp. *lactis* Bi-07	PRJNA156975
Bifidobacterium animalis subsp. *lactis* Bl12	PRJNA186412
Bifidobacterium animalis subsp. *lactis* Bl-04	PRJNA32897
Bifidobacterium animalis subsp. *lactis* BS 01	PRJNA59607
Bifidobacterium animalis subsp. *lactis* CNCM I-2494	PRJNA67865
Bifidobacterium animalis subsp. *lactis* DSM 10140	PRJNA32893
Bifidobacterium animalis subsp. *lactis* HN019	PRJNA28807
Bifidobacterium animalis subsp. *lactis* V9	PRJNA32515

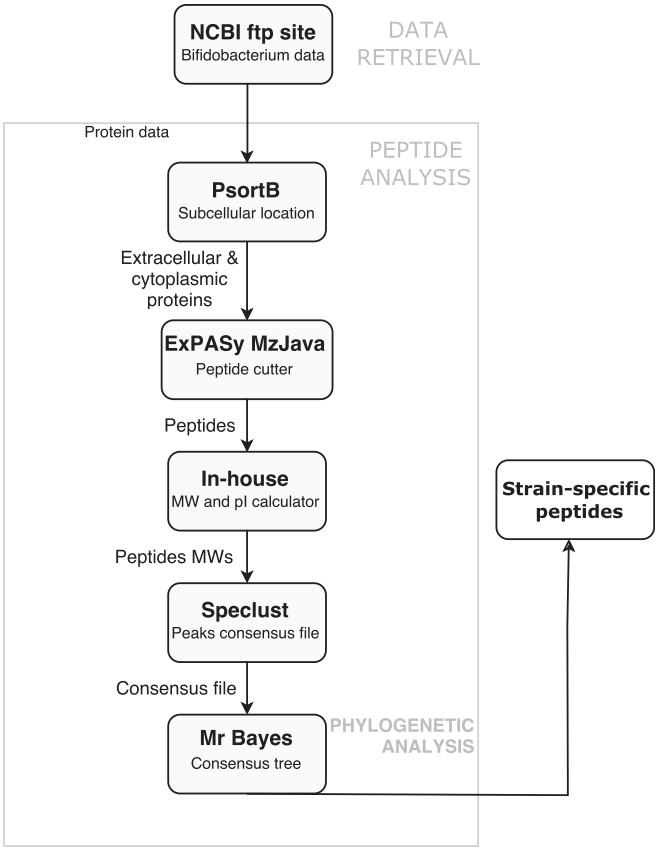

Fig. 1. Data workflow of the proposed peptidome-based strain comparison pipeline.

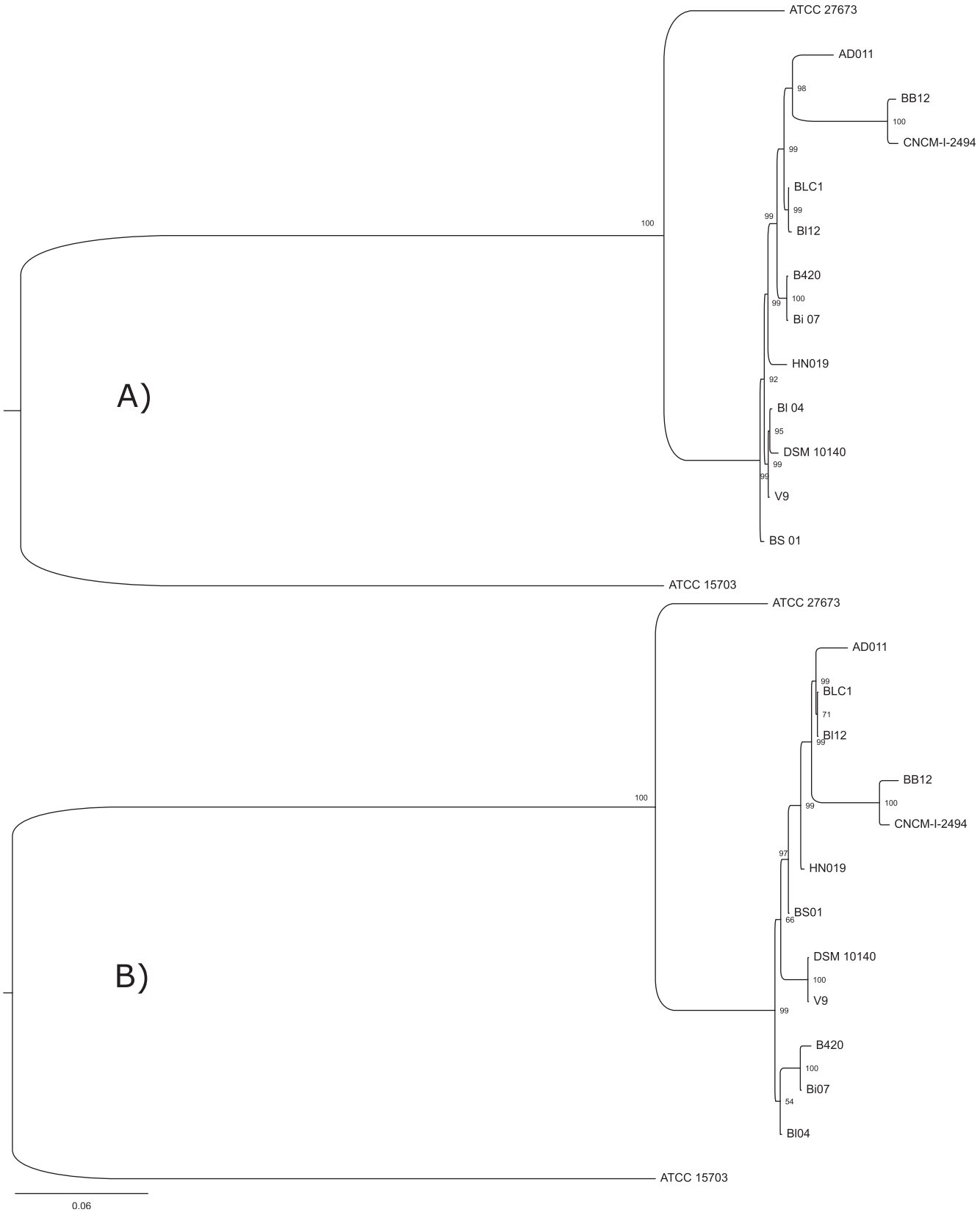

Fig. 2. Bayesian phylogenetic trees obtained with the "Cyto_50_more" (A) or the "Extracellular" (B) peptidome datasets.

the initial 25% of the trees (burnin = 250), where the log likelihood values of the analysis were not yet stabilized. *Bifidobacterium adolescentis* ATCC 15703 was selected as outgroup for phylogenetic tree rooting. The resulting peptide-based phylogenetic trees are shown in Fig. 2, one corresponding to the "Cyto_50_more" sub-peptidome (Fig. 2-A) and the other corresponding to the

"Extracellular" peptidome (Fig. 2-B). The trees did not show major discrepancies regarding strain clusters nor in average branch support. However, the extracellular dataset was more resolutive in terms of strain differentiation and branch lengths were higher in comparison with the cytoplasmic tree. This is interesting, because the extracellular subproteome contains many of the proteins supporting beneficial effects on the human host, notably in terms of immunomodulation (Sánchez et al., 2010). For that reason, one may hypothesize that specific strains with specific benefits harbor unique extracellular proteins encoded within their genomes.

Compared to other molecular approaches, our peptidome-based trees are fairly similar to those generated by single nucleotide polymorphisms (SNP)/insertion-deletion polymorphism-based allelic typing (Briczinski et al., 2009; Tmanova et al., 2012). Briefly, all strains were very close due to their high genetic similarity with the exception of *B. animalis* subsp. *lactis* ATCC 27673, which clustered apart with a 100% posterior probability. It has been reported that strain ATCC 27673 harbors unique genome traits into genomic islands, making this strain very different from the rest (Loquasto et al., 2013).

There are two main hypotheses supporting our results that indeed re-iterate previous findings from previous research. The first is that the high genetic similarity might be the result of a bias in strain sequencing, notably the fact that almost all sequenced strains come from commercial products; granted that the natural reservoir of this species is not yet know, we would lack most of the wild-type strains. The second is that it might be the consequence of a recent adaptation of an ancestor strain to the dairy environment, which would have eliminated many of the genetic diversity within the subspecies, an assessment perfectly consistent with the Founder effect theory of Population Genetics (Milani et al., 2013). Given this scenario of high genetic similarity, it is very hard to find new polymorphisms among *B. animalis* subsp. *lactis* strains; indeed occurrence of single nucleotide variants has been stimated in just 29 per 100 kb (Loquasto et al., 2013). Therefore, it is very difficult for an industry to differentiate its strain from the rest of competitors, both in their stock cultures and in their final products.

Within this context, our peptidome-based strategy is very valuable, as a SNP may be silent. However, some SNPs can lead to the generation of premature stop codons while others may cause a frameshift and produce altered amino acid sequences. All these changes have a direct reflect in the peptidomes, since proteins with different amino acid content or with different length will produce peptides with distinct molecular masses. All these potential differences are detected by our analysis, granted that any changes in the molecular mass of the peptides lead to a lack of consensus among strains.

With our methodology, we have been able to identify specific peptides in each of the strains included in our analysis. This was the case of peptide SCATPPMNGMSSMAR, specific of the strain *B. animalis* subsp. *lactis* DSM 10140 and encoded in long-chain-fatty-acid-CoA ligase, a protein that can be detected in the cytoplasmic fraction. Other example is the peptide RSSARPTPPRRIRRSAVSNQQF, encoded in a small and extracellular hypothetical protein of *B. animalis* subsp. *lactis* BB12. These are just two examples, but in fact more than 50 specific peptides per strain were identified using our approach, both in the cytoplasmic and extracellular proteomes (see Supplementary Material 3).

Knowledge on strain-specific peptides may facilitate the development of methodologies focused on the detection of the strain, e.g. in the final product. This is a very interesting applied area from the point of view of probiotic quality. Notably, antibody-based tests targeting these specific peptides may efficiently detect a defined strain in fermented milks or within the gut microbiota during clinical trials. This is extremely important in the case of *B. animalis* subsp. *lactis* given its global distribution in human food and low genetic variability. Therefore, we consider our peptidome-based approach of practical and important relevance to the industry, with the advantage of minimal implementation requirements and costs whilst providing considerable analysis power and flexibility.

Acknowledgements

This research was funded by Grant AGL2013-44039-R from the Spanish "Plan Estatal de I+D+I", and by Grant EM2014/046 from the "Plan Galego de investigación, innovación e crecemento 2011–2015". Borja Sánchez was recipient of a Ramón y Cajal postdoctoral contract from the Spanish Ministry of Economy and Competitiveness (RYC-2012-10052).

This work was also partially funded by the [14VI05] Contract-Programme from the Unixikversity of Vigo and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273). The research leading to these results has also received funding from the European Union's Seventh Framework Programme FP7/REGPOT-2012-2013.1 under grant agreement no 316265, BIOCAPS. This document reflects only the author's views and the European Union is not liable for any use that may be made of the information contained herein.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fm.2016.06.015.

References

Agrawal, A., Houghton, L.A., Morris, J., Reilly, B., Guyonnet, D., Goupil Feuillerat, N., Schlumberger, A., Jakob, S., Whorwell, P.J., 2009. Clinical trial: the effects of a fermented milk product containing Bifidobacterium lactis DN-173 010 on abdominal distension and gastrointestinal transit in irritable bowel syndrome with constipation. Aliment. Pharmacol. Ther. 29, 104–114. http://dx.doi.org/10.1111/j.1365-2036.2008.03853.x.

Alm, R., Johansson, P., Hjernø, K., Emanuelsson, C., Ringnér, M., Häkkinen, J., 2006. Detection and identification of protein isoforms using cluster analysis of MALDI-MS mass spectra. J. Proteome Res. 5, 785–792. http://dx.doi.org/10.1021/pr050354v.

Briczinski, E.P., Loquasto, J.R., Barrangou, R., Dudley, E.G., Roberts, A.M., Roberts, R.F., 2009. Strain-specific genotyping of Bifidobacterium animalis subsp. lactis by using single-nucleotide polymorphisms, insertions, and deletions. Appl. Environ. Microbiol. 75, 7501–7508. http://dx.doi.org/10.1128/AEM.01430-09.

Briczinski, E.P., Roberts, R.F., 2006. Technical note: a rapid pulsed-field gel electrophoresis method for analysis of bifidobacteria. J. Dairy Sci. 89, 2424–2427. http://dx.doi.org/10.3168/jds.S0022-0302(06)72315-3.

Chervaux, C., Grimaldi, C., Bolotin, A., Quinquis, B., Legrain-Raspaud, S., van Hylckama Vlieg, J.E.T., Denariaz, G., Smokvina, T., 2011. Genome sequence of the probiotic strain Bifidobacterium animalis subsp. lactis CNCM I-2494. J. Bacteriol. http://dx.doi.org/10.1128/JB.05716-11.

Deletoile, A., Passet, V., Aires, J., Chambaud, I., Butel, M.J., Smokvina, T., Brisse, S., 2010. Species delineation and clonal diversity in four Bifidobacterium species as revealed by multilocus sequencing. Res. Microbiol. 161, 82–90. http://dx.doi.org/10.1016/j.resmic.2009.12.006.

Gueimonde, M., Delgado, S., Mayo, B., Ruas-Madiedo, P., Margolles, A., De Los Reyes-Gavilán, C.G., 2004. Viability and diversity of probiotic Lactobacillus and Bifidobacterium populations included in commercial fermented milks. Food Res. Int. 37, 839–850. http://dx.doi.org/10.1016/j.foodres.2004.04.006.

Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., Morelli, L., Canani, R.B., Flint, H.J., Salminen, S., Calder, P.C., Sanders, M.E., 2014. 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. 11, 9. http://dx.doi.org/10.1038/nrgastro.2014.66.

Horlacher, O., Nikitin, F., Alocci, D., Mariethoz, J., Müller, M., Lisacek, F., 2015. MzJava: an open source library for mass spectrometry data processing. J. Proteomics. http://dx.doi.org/10.1016/j.jprot.2015.06.013.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755. http://dx.doi.org/10.1093/bioinformatics/17.8.754.

Isolauri, E., Arvola, T., Sütas, Y., Moilanen, E., Salminen, S., 2000. Probiotics in the management of atopic eczema. Clin. Exp. allergy J. 30, 1604–1610 doi:cea943

[pii].
Jayamanne, V.S., Adams, M.R., 2006. Determination of survival, identity and stress resistance of probiotic bifidobacteria in bio-yoghurts. Lett. Appl. Microbiol. 42, 189–194. http://dx.doi.org/10.1111/j.1472-765X.2006.01843.x.
Lomonaco, S., Furumoto, E.J., Loquasto, J.R., Morra, P., Grassi, A., Roberts, R.F., 2015. Development of a rapid SNP-typing assay to differentiate Bifidobacterium animalis ssp. lactis strains used in probiotic-supplemented dairy products. J. Dairy Sci. 98, 804–812. http://dx.doi.org/10.3168/jds.2014-8509.
Loquasto, J.R., Barrangou, R., Dudley, E.G., Stahl, B., Chen, C., Roberts, R.F., 2013. Bifidobacterium animalis subsp. Lactis ATCC 27673 is a genomically unique strain within its conserved subspecies. Appl. Environ. Microbiol. 79, 6903–6910. http://dx.doi.org/10.1128/AEM.01777-13.
Milani, C., Duranti, S., Lugli, G.A., Bottacini, F., Strati, F., Arioli, S., Foroni, E., Turroni, F., van Sinderen, D., Ventura, M., 2013. Comparative genomics of Bifidobacterium animalis subsp. lactis reveals a strict monophyletic bifidobacterial taxon. Appl. Environ. Microbiol. 79, 4304–4315. http://dx.doi.org/10.1128/AEM.00984-13.
Ruiz, L., Gueimonde, M., Patricia, R.M., Ribbera, A., de los Reyes-Gavilán, C.G., Ventura, M., Margolles, A., Sánchez, B., 2012. Molecular clues to understand the aerotolerance phenotype of Bifidobacterium animalis subsp. lactis. Appl. Environ. Microbiol. 78, 644–650. http://dx.doi.org/10.1128/AEM.05455-11.
Sánchez, B., Champomier-Vergès, M.-C., Stuer-Lauridsen, B., Ruas-Madiedo, P., Anglade, P., Baraige, F., de los Reyes-Gavilán, C.G., Johansen, E., Zagorec, M., Margolles, A., 2007. Adaptation and response of Bifidobacterium animalis subsp. lactis to bile: a proteomic and physiological approach. Appl. Environ. Microbiol. 73, 6757–6767. http://dx.doi.org/10.1128/AEM.00637-07.
Sánchez, B., Urdaci, M.C., Margolles, A., 2010. Extracellular proteins secreted by probiotic bacteria as mediators of effects that promote mucosa-bacteria interactions. Microbiology 156, 3232–3242. http://dx.doi.org/10.1099/mic.0.044057-0.
Tabbers, M.M., Chmielewska, A., Roseboom, M.G., Crastes, N., Perrin, C., Reitsma, J.B., Norbruis, O., Szajewska, H., Benninga, M.A., 2011. Fermented milk containing Bifidobacterium lactis DN-173 010 in childhood constipation: a randomized, double-blind, controlled trial. Pediatrics 127, e1392–e1399. http://dx.doi.org/10.1542/peds.2010-2590.
Tmanova, L.L., Onyenwoke, A., Roberts, R.F., 2012. Short communication: identification and differentiation of bifidobacteria obtained from Ukraine. J. Dairy Sci. 95, 91–97. http://dx.doi.org/10.3168/jds.2011-4448.
Veiga, P., Gallini, C.A., Beal, C., Michaud, M., Delaney, M.L., DuBois, A., Khlebnikov, A., van Hylckama Vlieg, J.E.T., Punit, S., Glickman, J.N., Onderdonk, A., Glimcher, L.H., Garrett, W.S., 2010. Bifidobacterium animalis subsp. lactis fermented milk product reduces inflammation by altering a niche for colitogenic microbes. Proc. Natl. Acad. Sci. U. S. A. 107, 18132–18137. http://dx.doi.org/10.1073/pnas.1011737107.
Ventura, M., Zink, R., 2002. Rapid identification, differentiation, and proposed new taxonomic classification of Bifidobacterium lactis. Appl. Environ. Microbiol. 68, 6429–6434. http://dx.doi.org/10.1128/AEM.68.12.6429-6434.2002.
Weizman, Z., Asli, G., Alsheikh, A., 2005. Effect of a probiotic infant formula on infections in child care centers: comparison of two probiotic agents. Pediatrics 115, 5–9. http://dx.doi.org/10.1542/peds.2004-1815.
Wheeler, D.L., Chappey, C., Lash, A.E., Leipe, D.D., Madden, T.L., Schuler, G.D., Tatusova, T.A., Rapp, B.A., 2000. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 28, 10–14. http://dx.doi.org/10.1093/nar/gks1189, 10.1093/nar/28.1.10.
Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L.J., Brinkman, F.S.L., 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615. http://dx.doi.org/10.1093/bioinformatics/btq249.