



**HAL**  
open science

## **Proteolytic activity of Lactobacillus strains isolated from Mongolian traditional dairy products: A multiparametric analysis**

Cyril Raveschot, Benoit Cudennec, Barbara Deracinois, Marc Frémont, Maxime Vaeremans, Jamyang Dugersuren, Shirchin Demberel, Djamel Drider, Pascal Dhulster, François Coutte, et al.

### ► To cite this version:

Cyril Raveschot, Benoit Cudennec, Barbara Deracinois, Marc Frémont, Maxime Vaeremans, et al.. Proteolytic activity of Lactobacillus strains isolated from Mongolian traditional dairy products: A multiparametric analysis. Food Chemistry, 2020, 304, pp.125415. <10.1016/j.foodchem.2019.125415>. <hal-03488186>

**HAL Id: hal-03488186**

**<https://hal.science/hal-03488186v1>**

Submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

1 **Proteolytic activity of *Lactobacillus* strains isolated from Mongolian traditional dairy**  
2 **products: a multiparametric analysis**

3

4 **Cyril Raveschot <sup>ab</sup>, Benoit Cudennec <sup>a</sup>, Barbara Deracinois <sup>a</sup>, Marc Frémont <sup>b</sup>, Maxime**  
5 **Vaeremans<sup>b</sup>, Jamyan Dugersuren <sup>c</sup>, Shirchin Demberel <sup>c</sup>, Djamel Drider <sup>a</sup>, Pascal**  
6 **Dhulster <sup>a</sup>, François Coutte <sup>a</sup>, Christophe Flahaut <sup>a</sup>**

7

8 <sup>a</sup> Université Lille, INRA, ISA, Université d'Artois, Université Littoral Côte d'Opale, EA  
9 7394-ICV Institut Charles Viollette, F-59000 Lille, France

10 <sup>b</sup> VF Bioscience, Parc Eurasanté, F-59120 Loos-lez-Lille, France

11 <sup>c</sup> Probiotic Research Center, Mongolian Veterinary Institute, 17024 Ulaanbaatar, Mongolia

12

13 **\* Corresponding author:**

14 Dr. Christophe FLAHAUT

15 Tel.: +33 3 21 79 17 80

16 Fax: +33 3 21 79 17 00

17 E-mail address: christophe.flahaut@univ-artois.fr

18

19 **Keywords**

20 Cell envelope associated peptidases, fermented milk, *Lactobacillus*, peptides

21

22 **Abbreviations**

23 ACE: Angiotensin Converting Enzyme

24 ACN: Acetonitrile

25 CEP: Cell Envelope associated Peptidase

26 EDTA: EthyleneDiamineTetraAcetic acid

27 FA: Formic Acid

28 FC Reagent: Folin-Ciocalteu Reagent

29 FPLC: Fast Protein Liquid Chromatography

30 HMWPs: High Molecular Weight Proteins/Peptides

31 LAB: Lactic Acid Bacteria

32 LMWPs: Low Molecular Weight Peptides

33 MRS: De Man, Rogosa, Sharpe

34 NISM: Non Inoculated Skim Milk

35 OD: Optical Density

36 PCA: Principal Component Analysis

37 PCR: Polymerase Chain Reaction

38 SEC: Size Exclusion Chromatography

39 TCA: TriChloroAcetic acid

40 TFA: TriFluoroAcetic acid

41 UHT: Ultra High Temperature

42

43

44

45

46 **ABSTRACT**

47 The aim of our study was to characterize the proteolytic activity of 170 *Lactobacillus* strains  
48 isolated from traditional Mongolian dairy products (yogurt and fermented milk), and to  
49 investigate their capacity to generate bioactive peptides during milk fermentation. All isolates  
50 were screened for proteolytic activity using skim milk agar-well diffusion test. Fifteen strains  
51 (9 *Lactobacillus helveticus* and 6 *Lactobacillus delbrueckii subsp. bulgaricus*) were then  
52 selected and further evaluated using an original strategy based on multiparametric analysis,  
53 taking into account growth rate, acidification capacity, proteolytic activity, cell envelope  
54 associated peptidase (CEP) profile and LC-MS/MS analysis of peptides. All parameters were  
55 analyzed using principal component analysis (PCA). Results showed that strain growth and  
56 acidification correlate with peptide production and that Mongolian *L. helveticus* strains differ  
57 from Western strains in terms of CEP distribution. The PCA revealed that CEP profiles are  
58 major determinants of  $\beta$ -casein hydrolysis patterns. Strains with distinctive proteolytic  
59 activities were identified.

60

61

62

63

64

65

66

67

## 68 **1. Introduction**

69 Milk proteins are precursors of many different peptides, some of which display interesting  
70 bioactivities: peptides with antimicrobial, antioxidant, opioid, immunostimulating or calcium  
71 binding activities have already been described (Nielsen, Beverly, Qu, & Dallas, 2017;  
72 Nongonierma & FitzGerald, 2015; Park & Nam, 2015). Bioactive peptides can be generated  
73 from milk proteins by enzymatic hydrolysis during gastrointestinal digestion, or produced *in*  
74 *vitro* with the use of commercial enzymes. However, industrial production of peptides is  
75 limited by the lack of suitable large-scale technologies, and by the high cost of enzymes (Wu  
76 et al., 2013).

77 Milk fermentation by *Lactobacillus* bacteria represents an inexpensive way to generate  
78 bioactive peptides (Giacometti & Buretić-Tomljanović, 2017). In order to source all the  
79 amino acids required for their growth, *Lactobacillus* bacteria indeed hydrolyze milk proteins  
80 during fermentation, and a large variety of peptides is generated in this process. The  
81 enzymatic system responsible for milk protein degradation is composed of cell envelope  
82 associated peptidases (CEPs) and peptidases. The CEPs first hydrolyze proteins into peptides,  
83 which are then transported into the cell by a peptide transport system; intracellular peptidases  
84 finally degrade these peptides into smaller molecules and amino acids (Savijoki, Ingmer &  
85 Varmanen, 2006). Only one CEP (named PrtB) has been described for *Lactobacillus*  
86 *delbrueckii subsp. bulgaricus* (*L. delbrueckii*) (Hou, Liu, Ren, Han, & Du, 2015) whereas 4  
87 CEPs (PrtH1, PrtH2, PrtH3 and PrtH4) can be expressed, in a strain dependent manner, by  
88 *Lactobacillus helveticus* (*L. helveticus*) strains (Griffiths & Tellez, 2013). The proteolytic  
89 activity of a given strain finally depends on its CEP and peptidase profiles, on gene variations,  
90 and on the conditions required for optimal enzyme activity (Alcántara et al., 2016).

91

92 *L. helveticus* LBK-16H is an example of strain that generates bioactive peptides upon milk  
93 fermentation: two tripeptides isoleucyl-prolyl-proline (IPP) and valyl-prolyl-proline (VPP)  
94 were identified, which have been shown to act as angiotensin converting enzyme (ACE)  
95 inhibitors (Narva, Halleen, Väänänen, & Korpela, 2004). Another example is the  
96 antioxidative peptide present in milk fermented by *L. delbrueckii* IFO13953 (Korhonen &  
97 Pihlanto, 2006).

98 This study aimed at evaluating the peptide production capacity of *Lactobacillus* strains  
99 isolated in three different areas of Mongolia. Production of dairies through spontaneous milk  
100 fermentation has a long history in this country. Herders use different milks (mare, cow, yak,  
101 camel, sheep, goat) to produce a variety of fermented food such as yogurt, kefir, or airag  
102 (fermented horse milk). These products were reported to contain a wide diversity of LAB,  
103 especially *Lactobacillus* species (Watanabe et al., 2008; Yu et al., 2011). Mongolian *L.*  
104 *helveticus* strains isolated from mare's fermented milk present a different CEP distribution,  
105 and generate different peptides from casein hydrolysis, when compared to European and  
106 North American strains (Miyamoto et al., 2015). This suggests that *Lactobacillus* species  
107 from different origins could hydrolyze milk proteins in different ways, and therefore generate  
108 peptides with original bioactivities. Actually, new peptides with ACE inhibitory activity have  
109 been identified in milk fermented by the Mongolian strain *L. helveticus* 130B4 (Schuangquan,  
110 Tsuda & Miyamoto, 2008).

111 The screening was carried on 170 *Lactobacillus* isolates, and involved a multiparametric  
112 analysis based on the correlation between i) fermentation characteristics (strain growth,  
113 acidification pattern), ii) CEP gene distribution and iii) peptide heterogeneity obtained after  
114 milk fermentation. The study was able to distinguish strains with specific proteolytic  
115 activities, generating distinct populations of peptides upon fermentation of milk.

116

## 117 **2. Materials and methods**

### 118 **2.1. Sample collection**

119 A total of 67 dairy samples were collected from three regions of Mongolia, corresponding to  
120 different ecosystems and environmental conditions: Gobi desert (16 samples), Northern  
121 Mongolia Hangai region (12 samples) and Altaï mountains (39 samples). Collected samples  
122 included mare (*Equus caballus*) fermented milk, camel (*Camelus bactrianus*) fermented milk,  
123 as well as cow (*Bos taurus*), goat (*Capra hircus*), yak (*Bos grunniens*) and hainakh (*Bos*  
124 *taurus* × *Bos grunniens*) yogurts. Samples were collected in sterile containers and kept  
125 refrigerated for a maximum of two days before analysis.

126

### 127 **2.2. Isolation and identification of LAB**

128 Samples were serially diluted in saline solution (NaCl 0.85% w/v), streaked on MRS plates,  
129 incubated for 48h at 37°C under anaerobic conditions (De Man, Rogosa & Sharpe, 1960).

130 Morphologically distinct colonies were picked and streaked again until purification. Strains  
131 were examined with an optical microscope for Gram staining. Catalase assay was performed  
132 by application of 3 % (v/v) hydrogen peroxide solution on cultures. Bubbling indicates  
133 catalase activity. A total of 170 isolates were finally retained in the collection. Strains were  
134 stored in MRS media and glycerol (15% v/v final concentration) at -80°C.

135 Identification of strains was performed by 16S rDNA amplification and sequencing. Strain  
136 DNA was extracted from bacterial colonies using QIAamp DNA Mini Kit (Qiagen, Hilden,  
137 Germany). DNA amplification and sequencing were performed by GATC biotech (Konstanz,  
138 Germany). Species were identified by comparing the 16S sequence with bacterial sequences  
139 in Medline databases, a score higher than 99% was considered significant.

140

### 141 **2.3. Screening for proteolytic activities on skim milk agar**

142 Proteolytic activity was determined using skim milk agar medium. Strains were grown in  
143 MRS at 37°C for 48h, washed twice in the same medium and adjusted to a final optical  
144 density at 600 nm (OD<sub>600</sub>) of 1. Twenty microliters of the suspension were placed in skim  
145 milk agar (5% w/v) wells (diameter 4 mm) and inoculated plates were incubated at 37°C for  
146 72h. Proteolysis resulted in the formation of a clear zone around the wells; diameter of this  
147 clear zone was measured for proteolytic activity quantification.

148

### 149 **2.4. Fermentation of Ultra High Temperature (UHT) skim milk with selected strains**

150 In order to assess the strains proteolytic activity, milk cultures were prepared in a media  
151 composed of 100% UHT skim milk (Cora, France). Strains were inoculated to 30 mL of UHT  
152 skim milk, to reach an initial OD<sub>600</sub> of 0.3, and grown at 37°C for 48h without agitation.  
153 Concomitantly, a non-inoculated UHT skim milk (NISM) was incubated in the same  
154 conditions and served as negative control. After incubation, the pH value of culture media  
155 was determined using a pH meter (Mettler Toledo, Greifensee, Swiss) and all milk cultures  
156 were kept at -20°C for further analysis. Experiments were performed in triplicate and the  
157 results were expressed as the mean plus standard error.

158

#### 159 **2.4.1. Bacterial growth determination**

160 Bacterial concentration in skim milk cultures was determined as follows: a volume of culture  
161 was diluted 10 times in 2% ethylenediaminetetraacetic acid (EDTA), pH 12, and then  
162 centrifuged at 13,000 × g for 10 min (Abraham, De Antoni & Anon, 1993). This process was

163 repeated and the pellet was resuspended in phosphate buffer solution (PBS) at pH 7. OD of  
164 the solution was measured at 600 nm.

165

#### 166 **2.4.2. Peptide extraction and purification**

167 Proteins were precipitated from 4 mL of milk cultures by addition of a trichloroacetic acid  
168 solution (TCA, 10% w/v in water) to reach a final TCA concentration of 1%. After  
169 centrifugation at  $10,000 \times g$  for 10 min at room temperature, the peptide-containing  
170 supernatant was collected. Sugars, salt and TCA were removed from supernatant using a C<sub>18</sub>  
171 solid phase extraction with Bond Elut C<sub>18</sub> 1000 mg microcolumns (Agilent Technologies,  
172 Santa Clara, USA). Elution of C<sub>18</sub>-retained peptides was performed with 2 mL of 80% v/v  
173 acetonitrile (ACN), 20% v/v water and 0.1% v/v trifluoroacetic acid (TFA).

174

#### 175 **2.4.3. Peptide quantification**

176 After extraction and purification, the assessment of peptide concentration in milk culture was  
177 performed using the Folin-Ciocalteu reagent (FC reagent – (ref F9252, Sigma-Aldrich, St  
178 Quentin Fallavier, France)) (McSweeney & Fox, 1997; Lowry, Rosebrough, Farr & Randall,  
179 1951). Briefly, the complexation reaction was performed in a final volume of 800  $\mu\text{L}$   
180 composed of 200  $\mu\text{L}$  of extracted peptides, 500  $\mu\text{L}$  of sodium carbonate solution (500 mM)  
181 and 100  $\mu\text{L}$  of FC reagent. After incubation at 37°C for 30 min in the dark, the OD<sub>750</sub> was  
182 measured to evaluate the peptide amount. Standard curve was established with a peptide  
183 digest standard (Peptide digest assay standard, ThermoFisher Scientific, Waltham, USA).

184

185

#### 186 **2.4.4. Molecular weight profiles of peptides obtained from milk hydrolysates**

187 Peptide content in milk hydrolysates was analyzed by size exclusion chromatography (SEC)  
188 based on a fast protein liquid chromatography (FPLC) using a Superdex Peptide 10/300 GL  
189 column (10 × 300-310 mm, 13 μm, GE Healthcare, Little Chalfont, UK) coupled to an AKTA  
190 Protein Purification System (GE Healthcare, Little Chalfont, UK). Twenty-five microliters of  
191 sample were loaded, eluted in isocratic conditions with 30% ACN, 0.1% TFA solution at a  
192 flow rate of 0.5 mL.min<sup>-1</sup> for 60 minutes and monitored at 214 nm. Five substances were used  
193 as internal standards to obtain a calibration curve for calculating molecular weight  
194 distribution: albumin (60 kDa), cytochrome C (12,400 Da), aprotinin (6,500 Da), vitamin B12  
195 (1,355 Da) and glutathione (307.3 Da). Peptide content in each sample was estimated by peak  
196 area integration. Two molecular weight ranges were distinguished, one corresponding to the  
197 high molecular weight proteins/peptides (HMWPs) with an apparent molecular weight higher  
198 than 1,700 Da, and one corresponding to the low molecular weight peptides (LMWPs) with  
199 an apparent molecular weight lower than 1,700 Da. Peptide content estimation for each range  
200 was expressed either as percentage of total profile area (%HMWPs and %LMWPs) or as  
201 percentage compared to non-inoculated milk sample (%HMWPs/Milk and %LMWPs/Milk).

202

#### 203 **2.4.5. Polymerase chain reaction (PCR)-based detection of CEP genes**

204 Bacteria from milk cultures were harvested as described in the 2.4.1. section. Cellular DNA  
205 was isolated using a Wizard Genomic DNA Purification Kit (Promega, Madison, USA). Each  
206 strain was tested with five different primer sets corresponding to different CEP genes. The  
207 primers described by Hou et al. (2015) were used for detection of *prtB* gene in *L. delbrueckii*  
208 strains. The primers used for the detection of *prtH1* and *prtH2* genes of *L. helveticus* were  
209 from Genay, Sadat, Gagnaire, & Lortal, (2009) and the primers for *prtH3* and *prtH4* (same

210 species) were from Broadbent et al. (2011). PCR products were detected with GelRed staining  
211 after electrophoresis on a 1% w/v agarose gel. An O'GeneRuler DNA ladder Mix  
212 (ThermoFisher Scientific, Waltham, USA) was used as molecular weight marker. The strain  
213 CNRZ32 CIRM-BIA 103 (provided by CIRM-BIA, INRA, Rennes, France) was used as  
214 control for *L. helveticus* strains.

215

#### 216 **2.4.6. Peptides identification by UPLC-ESI-qTOF-MS/MS**

217 The peptides purified on Bond Elut C<sub>18</sub> microcolumns (section 2.4.2.) were dried by  
218 centrifugal evaporation (miVac Centrifugal Vacuum Concentrators, Gene Vac, Ipswich, UK)  
219 for 2h at 40°C. Dried peptides were redissolved in 100 µL H<sub>2</sub>O, 0.1 % TFA and centrifuged  
220 for 10 min at 8,000 × g. Ten µL were chromatographed on an ACQUITY UPLC system  
221 (Waters, Manchester, UK) using a C<sub>18</sub>-Kinetex column (150 x 4.6 mm, 2.6µ 100Å,  
222 Phenomenex, Torrance, USA), a linear gradient of ACN containing 0.1% FA (5-15% ACN  
223 over 30 min, 15-30% ACN over 60 min, 30-50% ACN over 10 min and 50-95% ACN over  
224 10 min) and a flow rate of 500 µL.min<sup>-1</sup>. The HPLC eluent was directly electrosprayed from  
225 the column end at an applied voltage of 3 kV, using a desolvation gas (N<sub>2</sub>) flow of 600 L.h<sup>-1</sup>,  
226 a nebulizer gas flow of 2.5 bar and desolvation temperature of 300°C, respectively. The  
227 chromatography device was coupled to SYNAPT-G2-Si mass spectrometer (Waters,  
228 Manchester, UK). MS analysis was performed in sensitivity, positive ions and data dependent  
229 analysis (DDA) modes and MS data were collected in the 100-2000 m/z range with a scan  
230 time of 0.2 s. A maximum of 15 precursor ions with an intensity threshold of 10,000 counts  
231 were selected for ion trap collision-induced dissociation (CID) fragmentation and subjected to  
232 a collision energy ramping from 8V to 9V for low mass and 40V to 90V for high mass. The  
233 MS/MS spectra were recorded on the 100 to 2,000 m/z range with a scan time of 0.1 s. Peaks  
234 were analyzed using Mass Lynx software (ver.4.1, Waters, Manchester, UK).

235 Database searches via PEAKS Studio 7.0 (Bioinformatics Solutions) were performed using  
236 the UniProt databases (May 15, 2017) restricted to *Bos taurus* organism. A mass tolerance of  
237 35 ppm and 3 missing cleavage sites and an MS/MS tolerance of 0.2 Da were allowed.  
238 Variable methionine oxidation was also considered. The relevance of protein and peptide  
239 identities was judged according to their identification generated by PEAKS Studio 7.0  
240 ( $p < 0.05$ ) and a false discovery rate  $< 1\%$ .

241

#### 242 **2.4.7. $\beta$ -casein heat map construction from MS analysis**

243 Heat maps were constructed from  $\beta$ -casein peptides. The frequency of appearance of each  
244 amino acid was calculated from all identified peptides. Apparition frequency was established  
245 from *in silico* approach (Lacroix & Li-Chan, 2012) using Microsoft Excel software (2013  
246 version) with the following formula:  $A = a/N$ , where  $A$  is the apparition frequency of an amino  
247 acid,  $a$  the amino acid occurrence and  $N$  the number of peptide sequences. According to  
248 calculated frequency, a color was associated to each amino acid from white (no apparition  
249 frequency) to dark grey (high apparition frequency).

250

#### 251 **2.4.8. Statistical analysis and principal component analysis (PCA)**

252 All statistical differences between strains were measured using one way ANOVA and  
253 Tukey's post-hoc test. Differences between means were considered significant when p value  
254  $< 0.05$ . In order to analyze all measured variables at the same time, a first principal  
255 component analysis (PCA) was performed on data corresponding to results from proteolytic  
256 activity on milk agar, growth determination, pH determination and estimation of peptide  
257 contents. A second PCA was conducted to compare data from MS analysis ( $\beta$ -casein heat  
258 maps). PCA results were represented both as a variables map (correlation between studied

259 variables) and as individuals map (similarity between individuals). All statistical analyses and  
260 PCA were performed on R software (FactorMineR package) (R core team, 2016, Vienna,  
261 Austria).

262

### 263 **3. Results**

#### 264 **3.1. Identification of *Lactobacillus* species in the collected dairy samples**

265 One hundred and seventy (170) strains of LAB were isolated from 67 dairy products  
266 (fermented milks and yogurts) collected in three different regions of Mongolia: Gobi desert,  
267 Hangai region, Altaï mountains. Twelve different bacterial species were represented among  
268 the isolates (**Figure 1A**). Three dominant species accounted for more than 75% of the  
269 isolates: *L. helveticus* (73 strains), *L. delbrueckii* (39 strains) and *L. kefiranofaciens* (29  
270 strains). There were however differences between the three regions: the dominant species in  
271 samples from Gobi area was *L. kefiranofaciens* (half of the isolates, retrieved in all types of  
272 samples including yogurt, airag, camel fermented milk) whereas this species was virtually  
273 absent in the Altaï samples (only one isolate). In the Altaï samples, the dominant species was  
274 *L. helveticus*, which for instance accounted for 76% of the isolates obtained from airag. In  
275 airag samples collected in the Gobi area, *L. helveticus* accounted for only 42% of all isolates.

276

#### 277 **3.2. Selection of highly proteolytic *L. helveticus* and *L. delbrueckii* strains**

278 A standardized skim milk agar-well diffusion test was used to evaluate the proteolytic activity  
279 of all 170 strains. Strains showing the highest proteolytic activity (clear zone diameter larger  
280 than 28 mm, **Table 1**) belonged to the species *L. helveticus* and *L. delbrueckii*. Fifteen strains  
281 (nine *L. helveticus*, six *L. delbrueckii*) were selected for the further steps of the screening

282 strategy presented in **Figure 1B**. Briefly, this strategy relies on 48h skim milk fermentations  
283 with the selected strains, in order to i) follow milk acidification, ii) harvest cells to monitor  
284 growth and perform PCR-based detection of peptidase genes and iii) purify and extract  
285 peptides for quantification, evaluation of their apparent molecular weight distribution and  
286 identification of sequences by peptidomics approach.

287

### 288 **3.3. Multiparametric analysis of fermented cow milk**

#### 289 **3.3.1. Correlation between strain growth and acidification capacity in UHT skim milk**

290 Skim milk was inoculated with the selected strains; bacterial growth and pH were measured  
291 after 48 hours incubation. Two clearly distinct growth behaviors were observed (**Table 1,**  
292 **upper and lower parts**): *L. helveticus* strains showed a much faster growth, reaching an  
293 OD<sub>600</sub> mean of  $3.03 \pm 0.322$  whereas *L. delbrueckii* strains only reached a mean of  $0.856 \pm$   
294  $0.234$ .

295 Acidification capacity of the selected strains was evaluated by pH determination after 48h of  
296 culture. The initial pH of NISM (control) was 6.5 and reached  $6.45 \pm 0.05$  after 48h of  
297 incubation (**Table 1**). *L. helveticus* and *L. delbrueckii* strains displayed two distinct  
298 acidification capabilities. In almost all cases *L. helveticus* strains led to a pH decrease of  
299 minimum 3 pH units, reaching a mean of  $3.43 \pm 0.08$ , whereas the pH of milk inoculated with  
300 *L. delbrueckii* decreased by a maximum of 2 pH units, reaching a mean of  $4.57 \pm 0.07$ . *L.*  
301 *helveticus* strains were therefore more efficient than *L. delbrueckii* strains for both bacterial  
302 growth and acidification capacity. As expected, all milk fermentations led to milk protein  
303 coagulation because of casein precipitation at pH lower than 4.6 (Raak, Rohm, & Jaros,  
304 2017).

305

### 306 3.3.2. Increase of peptide concentration in inoculated UHT skim milk

307 Peptides in fermented milk were extracted by protein precipitation and purified by C<sub>18</sub> solid  
308 phase extraction. Assessment of peptide concentrations using the FC reagent showed, for all  
309 tested strains, an enrichment after 48h fermentation (**Table 1**). Peptide concentration in NISM  
310 was 2.06 g.L<sup>-1</sup> ± 0.62 after 48h incubation while inoculated milks showed concentrations  
311 higher than 4 g.L<sup>-1</sup> in all cases but one. On average, *L. helveticus* strains generate more  
312 peptides than *L. delbrueckii* strains (mean 5.17 g.L<sup>-1</sup> ± 0.92 vs. 4.96 g.L<sup>-1</sup> ± 0.74). The two  
313 best strains, with a final peptide concentration higher than 6 g.L<sup>-1</sup>, are two *L. helveticus* (25c,  
314 45a).

315 The size of the peptides generated during fermentation was analyzed by SEC. The  
316 chromatogram presented (**Figure 1C**) depicts the peptide apparent molecular weight  
317 distribution. Roughly, the SEC profiles showed two distinct regions: a region of high  
318 molecular weight proteins/peptides (HMWPs: fermentation products with an apparent  
319 molecular weight higher than 1,700 Da) and a region with low molecular weight peptides  
320 (LMWPs: products with an apparent molecular weight lower than 1,700 Da). Profiles were  
321 analyzed by integration of the two distinct molecular weight ranges (HMWPs, > 1700 Da and  
322 LMWPs, < 1700 Da) and were expressed (**Table 1**) as % of total area or as % of control  
323 (NISM). Proportions for NISM were 85.62% ± 2.72 and 14.39% ± 2.72 of HMWPs and  
324 LMWPs respectively. As expected, inoculation and fermentation of milk with lactic acid  
325 bacteria lead to an increase of LMWPs, and a decrease of HMWPs compared to NISM  
326 (**Figure 1C**), reflecting the proteolytic activity of the strains against protein and large  
327 peptides. *L. helveticus* and *L. delbrueckii* strains displayed two distinct peptide production  
328 capabilities: the increase of LMWPs proportion compared to NISM was, on average, 1.5 fold  
329 for *L. delbrueckii* and 4.2 fold for *L. helveticus* (**Table 1, last column**). The highest activity  
330 was observed for strain 25c (% of LMWPs increased 6.7 times compared to NISM).

331

### 332 **3.3.3. Strain growth and acidification capacities correlate with peptide production**

333 A first PCA was applied to evaluate correlations between the following parameters:  
334 proteolysis diameter on milk agar, cell growth, acidification and peptide content evaluation  
335 (FC reagent-based peptide quantification, total area of SEC profiles, %HMWPs,  
336 %HMWPs/NISM, %LMWPs and %LMWPs/NISM). **Figure 2A** shows the obtained variable  
337 map. The principal components Dim1 and Dim2 explained 60.82% and 23.56% of the total  
338 variance, respectively. A positive correlation was observed between acidification and HMWP  
339 proportion (%HMWPs). Moreover strain growth was correlated with a set of variables linked  
340 to peptide production (FC reagent-based peptide quantification, %LMWPs and  
341 %LMWPs/NISM) and with proteolysis diameter on milk agar plates (**Table 1**). Acidification  
342 and HMWP content were negatively correlated with variables such as growth, LMWP  
343 proportion and skim milk agar test results. Regarding the individuals map (**Figure 2B**), all  
344 tested strains were clearly separated from the NISM control. *L. delbrueckii* strains formed an  
345 homogeneous group, clustered together on the negative sides of Dim1 and Dim2 whereas *L.*  
346 *helveticus* strains were scattered along the positive sides of Dim1 and Dim2.

347

### 348 **3.4. CEP gene distribution**

349 The distribution of CEP genes was analyzed by PCR in the selected strains and in a control *L.*  
350 *helveticus* strain (CNRZ32) known to possess all four *prtH* genes (*prtH1* to *prtH4*). Results  
351 are shown in **Table 2**. The *prtB* gene was detected in all *L. delbrueckii* strains but not in *L.*  
352 *helveticus* strains. The *prtH1* gene was detected in 6 *L. helveticus* strains (16b, 25b, 25c, 49d,  
353 58b and 60b) and, as expected, in CNRZ32 control strain. Interestingly, all *L. helveticus*  
354 strains presenting the *prtH1* gene were isolated from mare fermented milk. *prtH2* and *prtH3*

355 genes were detected in 3 *L. helveticus* strains (45a, 46b and 47d) and in the CNRZ32 control  
356 strain. The *prtH4* gene was detected in the CNRZ32 control strain but was not found in any of  
357 the Mongolian strains.

358

### 359 **3.5. Peptide heterogeneity in fermented milk**

360 Peptides were extracted from fermented milk by C<sub>18</sub> SPE and analyzed by peptidomics  
361 approach. Analysis of the control sample (NISM) allowed defining the base level as 596  
362 peptides belonging to 22 proteins. Nevertheless, after filtering compared to the cow milk  
363 protein library ([http://www.dallaslab.org/resources\\_CowMilkProteinDatabase.fasta](http://www.dallaslab.org/resources_CowMilkProteinDatabase.fasta)), 12 cow  
364 milk proteins ( $\alpha_{S1}$ -casein,  $\beta$ -casein,  $\alpha_{S2}$ -casein, GLYCAM1,  $\kappa$ -casein, BTN1A1,  $\beta$ -  
365 lactoglobulin, FGFBP, lactoperoxidase, osteopontin, perilipin, TRIP6) were finally identified  
366 in the NISM control sample. 32% of all peptides were from  $\beta$ -casein (data not shown). In  
367 comparison, between 150 and 722 peptides were identified in milks fermented with the  
368 selected strains. In these fermented milks,  $\beta$ -casein peptides accounted for 44 to 67% of all  
369 sequences identified.

370 Because of the large proportion of  $\beta$ -casein peptides identified in milk cultures, and the  
371 natural high abundance of  $\beta$ -casein in milk, we decided to focus the analysis on this protein.

372 **Figure 3** shows the heat maps obtained for two strains with different CEP profiles, 60b  
373 (*prtH1*) and 47d (*prtH2/prtH3*). In such type of graphical representation, all peptides  
374 identified are repositioned (multiple horizontal grey bars in the middle of Figure 3) on the  $\beta$ -  
375 casein backbone, clearly illustrating the  $\beta$ -casein proteolysis patterns. Very few peptides came  
376 from the  $\beta$ -casein N-terminal region (signal peptide excluded); the vast majority of them came  
377 from a part of  $\beta$ -casein covering amino acids 56 to 224. The repartition of peptides on the  $\beta$ -  
378 casein backbone defines a specific proteolysis signature. For the two strains shown as

379 examples (60b, 47d) these signatures were clearly different: the preferential areas of cleavage  
380 (dark grey regions on the heat maps) were not situated in the same regions of the protein.

381

### 382 **3.6. $\beta$ -casein heat maps are correlated with CEP gene distribution**

383 In order to compare the  $\beta$ -casein proteolysis signatures obtained for each of the 15 strains, a  
384 second PCA was performed which included the heat map data. **Figure 4A** represents the  
385 individuals map of this PCA. The principal components Dim1 and Dim2 explained 43.11%  
386 and 31.43% of the total variance, respectively. The CEP gene profile of the strains is reported  
387 on the figure. Interestingly, the PCA clearly grouped the strains according to their CEP gene  
388 profile. As expected, NISM condition (no strain = no *prt* gene) stands apart from the other  
389 points and only depends on Dim2. *L. delbrueckii* strains, which possess the *prtB* gene, were  
390 grouped on the positive sides of Dim1 and Dim2. *L. helveticus* strains were scattered along  
391 Dim1 and on the negative side of Dim2. Interestingly, strains harboring *prtH1* gene were  
392 separated from strains harboring *prtH2* and *prtH3* genes. *prtH1* strains showed a higher  
393 dispersion, one half on the negative side of Dim1 (strains 16b, 25b and 60b) and the second  
394 half on its positive side (strain 25c, 49d and 58b). This dispersion may reflect the presence of  
395 different *prtH1* gene variants in our isolates, which cannot be distinguished by our PCR  
396 approach.

397  $\beta$ -casein is composed (before physiological maturation) of 209 amino acids. The variables  
398 map of the PCA (**Figure 4B**) was constructed using 209 variables corresponding to the  
399 occurrence frequency of these 209 amino acids. Interestingly, the most implicated variables  
400 on Dim1 and Dim2 (defined by  $\cos^2 > 0.85$ ) were not corresponding to amino acids randomly  
401 distributed in the protein. Instead, different specific regions on the  $\beta$ -casein sequence were  
402 identified. This observation allowed to distinguish the  $\beta$ -casein areas that were differentially

403 hydrolyzed by the strains. Dim2 corresponded to the N<sub>t</sub> region of β-casein (hydrophilic  
404 region), whereas Dim1 corresponded to different areas spread along β-casein until the C<sub>t</sub>  
405 region (hydrophobic parts of the protein) (**Figure 4C**). *In fine*, this representation illustrates  
406 the protein regions preferentially hydrolyzed during fermentation.

407

## 408 **4. Discussion**

### 409 **4.1. Strain identification and screening for proteolytic activity**

410 170 strains were isolated from various Mongolian dairy products, and identified by  
411 amplification and 16S rDNA sequencing. Strain identification showed a predominance of 3  
412 *Lactobacillus* species: *L. helveticus*, *L. delbrueckii* and *L. kefiranofaciens*. This observation is  
413 in agreement with previous reports. Several authors have reported that *L. helveticus* was the  
414 dominant LAB in mare fermented milk (airag) (Miyamoto et al., 2015; Watanabe et al.,  
415 2008). Yu et al. (2011) obtained 668 isolates from different fermented milk products and  
416 reported a majority of *L. delbrueckii* and *L. helveticus* strains. *L. kefiranofaciens* was also a  
417 species predominantly recovered from Mongolian dairy products (Watanabe et al., 2008).  
418 However, our findings show that the prevalence of these species depends on the geographical  
419 area: *L. kefiranofaciens*, predominant in dairy samples collected in the Gobi desert, was not  
420 found in samples from the Altai mountains. *L. helveticus* was more represented in the Altai  
421 samples than in samples collected in the two other regions.

422 The proteolytic activity of the strains was screened on milk agar. *L. helveticus* strains showed  
423 the highest proteolytic activity of all tested strains. Actually, *L. helveticus* is considered as the  
424 species displaying the highest proteolytic activity among all *Lactobacillus* (Griffiths & Tellez,  
425 2013). The other species showing a significant proteolytic activity in our study was *L.*  
426 *delbrueckii*. Again this is in agreement with previous reports; *L. delbrueckii* is used as

427 fermentation starter for initiating proteolysis in yogurt production (Pailin, Kang, Schmidt, &  
428 Fung, 2001). *L. kefiranofaciens* strains in our collection only displayed moderate proteolytic  
429 activities (data not shown).

430 Dairy products collected in the Altaï region are therefore characterized by a more proteolytic  
431 microbiota (mostly *L. helveticus*) than products from other areas. These products may  
432 therefore present specific properties and potential health benefits (more bioactive peptides).  
433 On the other hand, products collected in the Gobi desert are hosting a large variety of *L.*  
434 *kefiranofaciens* strains, known to produce kefiran, an exopolysaccharide with beneficial  
435 effects on immune system function and cardiovascular health (Maeda, Zhu, Suzuki, Suzuki &  
436 Kitamura, 2004). Studies investigating the link between consumption of these products and  
437 protection of the local populations against specific pathologies could be undertaken.

438

#### 439 **4.2. Differences in fermentation capacities between *L. helveticus* and *L. delbrueckii*** 440 **strains**

441 The fermentation capacity of *L. helveticus* and *L. delbrueckii* strains was investigated by  
442 measuring growth, milk acidification and peptide production 48 hours after UHT skim milk  
443 inoculation. Overall, *L. helveticus* strains reached a higher cell density after fermentation than  
444 *L. delbrueckii* strains. As expected, this high bacterial density resulted in a strong acidification  
445 of the medium, higher than for samples inoculated with *L. delbrueckii*, which is in agreement  
446 with previous reports (Gandhi & Shah, 2014).

447 A PCA was used to determine the correlations between all analyzed parameters. PCA are  
448 useful to synthesize the information carried by a large number of quantitative variables  
449 (Lacou, Lê, Pezennec, & Gagnaire, 2015). Bezerra et al. (2016), for instance, have used a  
450 PCA to study the proteolysis occurring in goat “coalho” cheese. In our study, the variables

451 map showed that high cellular growth and high acidification capacities were linked to a more  
452 abundant peptide production. Such correlation makes sense, especially in the case of *L.*  
453 *helveticus* species, where the efficient proteolytic system fulfill their nitrogen requirement  
454 leading to their high growth in pure milk. LAB is a heterogeneous microorganism group with  
455 multiple auxotrophies for amino acids and the capacity to rapidly degrade milk proteins, and  
456 thereby release these amino acids, is a major asset (Courtin, Monnet, & Rul, 2002; Griffiths &  
457 Tellez, 2013; Savijoki et al., 2006).

458

#### 459 **4.3. CEP distribution in the fifteen selected strains**

460 All *L. delbrueckii* strains were found to possess the *prtB* gene, which is so far the only CEP  
461 described for this species (Hou et al., 2015). *L. helveticus* strains showed more heterogeneity,  
462 being divided into two groups: 6 strains with *prtH1* gene, and 3 strains with the combination  
463 *prtH2/prtH3*. The CEP profile does not seem to be related to the geographical origin of the  
464 strains: among the six *prtH1* strains, one was from Gobi, two from the Hangai region, and  
465 three from Altai. Rather, the CEP profiles of the strains seem to be related to the type of  
466 sample from which they were isolated: all *prtH1* strains were isolated from airag samples  
467 (horse milk) whereas the three strains harboring *prtH2/prtH3* were isolated from other sources  
468 (yak and camel milks). The presence of *prtH1* in *L. helveticus* strains isolated from airag had  
469 already been reported in another study (Miyamoto et al., 2015). However, the combination of  
470 the two CEP genes *prtH2* and *prtH3* is unusual. Investigating CEP genes distribution in 50 *L.*  
471 *helveticus* strains of European and North American origin, Broadbent et al. (2011) reported  
472 only three strains with the same profile. This seems to confirm previous observations that  
473 Mongolian *L. helveticus* strains differ from Western strains in terms of CEP gene profiles  
474 (Miyamoto et al., 2015).

475 *In fine*, the *L. helveticus* species is characterized by a high intraspecific diversity of CEP  
476 profiles. Fermentation of milk with different *L. helveticus* strains may therefore lead to an  
477 important heterogeneity of peptides (Sadat-Mekmene, Genay, Atlan, Lortal, & Gagnaire,  
478 2011a).

479

#### 480 **4.4. Correlation between CEP profiles and $\beta$ -casein cleavage patterns**

481  $\beta$ -casein is the main substrate of *Lactobacillus* proteases (Jensen, Vogensen, & Ardö, 2009).  
482 It is thought that different CEPs can have very different cleavage specificities (Oberg,  
483 Broadbent, Strickland, & McMahon, 2002). However it is somewhat difficult to compare the  
484 results of separate publications: depending on the study, caseins are indeed used either as  
485 fragments, as purified forms, or as micelles in milk, and this of course may influence the  
486 ability of proteases to access specific cleavage sites in the protein (Sadat-Mekmene et al.,  
487 2011a). Oberg et al. (2002), who assessed the proteolytic activity of 14 *L. delbrueckii* strains  
488 based on the cleavage of an  $\alpha_{s1}$ -casein fragment, reported several distinct digestion patterns.  
489 In contrast, in our study, the  $\beta$ -casein cleavage patterns of all *L. delbrueckii* strains revealed  
490 some common cleavage sites. Our strains may be genetically closer to each other; but our  
491 observation may also result from the fact that we worked with native milk  $\beta$ -casein (micelle  
492 configuration), and not with a fragment of this protein.

493 The PCA separated *L. delbrueckii* strains from *L. helveticus* strains, indicating dissimilarities  
494 on  $\beta$ -casein hydrolysis patterns. Sadat-Mekmene et al. (2011b) obtained similar results when  
495 incubating purified  $\beta$ -casein with either *L. helveticus* CP790 or *L. delbrueckii* CRL581 strains  
496 (Sadat-Mekmene et al., 2011b). It is notable however that three of our *L. helveticus* strains  
497 (25c, 49d, 58b, all *prtH1*) were located very close to the *prtB* strains (*L. delbrueckii*). The  
498 three other *prtH1* strains (16b, 25b, 60b) were located close to the *prtH2/prtH3* strains

499 **(Figure 4A)**. Therefore, it appears that *prtB* and *prtH2/prtH3* genotypes are associated with  
500 clearly distinct proteolytic activities; the situation for *prtH1* is however more complex, which  
501 may be related to the existence of different alleles for this gene, not differentiated by our PCR  
502 (Genay et al., 2009). Interestingly, while PrtH2 and PrtH3 share 58% of amino acid strict  
503 identity, PrtH1 only shares 32 and 34% of amino acid strict identities with PrtH2 and PrtH3,  
504 respectively (Broadbent et al., 2011). It is remarkable that despite this high genetic variation,  
505 some *prtH1* strains still display a proteolytic activity which appears very similar to the  
506 activity of *prtH2/prtH3* strains.

507 As reported in section 3.6, the Dim2 of PCA depicted the hydrophilic region of  $\beta$ -casein while  
508 Dim1 depicted the hydrophobic region of  $\beta$ -casein **(Figure 4)**. The selection of the most  
509 implicated variables from PCA allowed to highlight the main important differences between  
510 conditions. The NISM condition was used to define the basal proteolytic pattern due to  
511 endogenous milk peptidases such as plasmin and cathepsin D (Dallas et al., 2016; Sadat-  
512 Mekmene et al., 2011b). As previously reported for human milk (Khaldi et al., 2014), our  
513 results indicate that these endogenous milk peptidases hydrolyzed the hydrophilic part (N<sub>t</sub>  
514 region) of  $\beta$ -casein during the our experimental conditions (37°C and 48h): the point  
515 corresponding to NISM was indeed placed by the PCA right and high on the Dim2 axis  
516 **(Figure 4A)**.

517 All the *L. delbrueckii* strains were closely grouped together, reflecting a high homogeneity in  
518 terms of proteolytic capacities; they were placed high on Dim1 and close to 0 on the Dim2  
519 axis, suggesting that hydrophobic areas along  $\beta$ -casein were preferentially hydrolyzed by  
520 these strains. In contrast, strains of *L. helveticus* were scattered along the Dim1 axis,  
521 reflecting a much higher heterogeneity; strains harboring *prtH2* and *prtH3* were located far in  
522 the negative side of Dim1, showing a preferential hydrolysis of hydrophobic  $\beta$ -casein regions.

523 Our PCA therefore revealed the existence of different proteolytic activities, resulting in  
524 different cleavage patterns of  $\beta$ -casein. The capacity to cleave  $\beta$ -casein in its more  
525 hydrophobic regions, which are more difficult to access when the protein is under micellar  
526 form (Dumas, Brignon, Grosclaude, & Mercier, 1972), is a significant species advantage.

527

## 528 **5. Conclusion**

529 In this study, we used an original approach to screen a collection of 170 *Lactobacillus* strains  
530 for their ability to produce peptides derived from milk proteins. All strains had been isolated  
531 from dairy products collected in three different regions of Mongolia. Fifteen strains were first  
532 selected by milk agar-well diffusion test; bacterial growth, milk acidification and production  
533 of peptides were then evaluated after 48h fermentation of skim milk. The CEP gene profile of  
534 each strain was also determined. A multiparametric analysis demonstrated that growth and  
535 milk acidification capacity correlated closely with peptide production. The CEP gene  
536 distribution was different in Mongolian *L. helveticus* strains than in European and North  
537 American strains, and PCA revealed that presence of specific CEPs was linked to specific  $\beta$ -  
538 casein hydrolysis patterns. The tested strains display, at various levels, the capacity to cleave  
539  $\beta$ -casein on specific sites buried inside the micellar form of caseins (hydrophobic regions).  
540 This multiparametric analysis allowed to select strains harboring i) distinct proteolytic  
541 patterns, ii) different CEP gene profiles, iii) high growth rate in milk and iv) high acidification  
542 capacity. The selected strains can be used as new fermentation starters, able to generate milk  
543 peptides with potential bioactivities. The analysis of such bioactivities, using *in silico*  
544 approach by comparison with milk bioactive peptide database (Nielsen et al., 2017), is  
545 ongoing.

546

## 547 **Acknowledgements**

548 This research was supported by a grant from BPI France/Métropole Européenne de  
549 Lille/Région Hauts de France (DOS0020745) and by a grant from Association Nationale  
550 Recherche et Technologie (ANRT). This work was supported by the Hauts-de-France region  
551 funding through the ALIBIOTECH research program. The HPLC-MS experiments were  
552 performed on the REALCAT platform funded by a French governmental subsidy managed by  
553 the French National Research Agency (ANR) within the frame of the “Future  
554 Investments” program (ANR-11-EQPX-0037). The Hauts-de-France region and the FEDER,  
555 the Ecole Centrale de Lille and the Centrale Initiatives Foundation are also warmly  
556 acknowledged for their financial contributions to the acquisition of REALCAT platform  
557 equipment. We are grateful to Prof. Anil Kulkarni (University of Texas Medical School,  
558 Houston, USA), Dr Anarkhuu Bold-Erdene and Dr Amarbat Baatarum (National Cancer  
559 Center, Ulaanbaatar, Mongolia) for their support during the initiation of the project. The in-  
560 house tool used for heat maps and peptide mappings was designed by Yohan Guegaden. We  
561 thank Aurélie Mateos for helpful discussions.

562

## 563 **Conflict of interest**

564 CR, MF and MV are affiliated with VF Bioscience SAS, France, a company commercializing  
565 lactic acid bacteria strains for use as probiotics or dairy starters.

566

## 567 **References**

568 Abraham, A. G., Antoni, G. L., & Anon, M. C. (1993). Proteolytic activity of *Lactobacillus*  
569 *bulgaricus* grown in milk. *Journal of Dairy Science*, 76(1900), 1498–1505.

570 [https://doi.org/10.3168/jds.S0022-0302\(93\)77481-0](https://doi.org/10.3168/jds.S0022-0302(93)77481-0)

571 Alcántara, C., Bäuerl, C., Revilla-Guarinos, A., Pérez-Martínez, G., Monedero, V., & Zúñiga,  
572 M. (2016). Peptide and amino acid metabolism is controlled by an OmpR-family  
573 response regulator in *Lactobacillus casei*. *Molecular Microbiology*, *100*(1), 25–41.  
574 <https://doi.org/10.1111/mmi.13299>

575 Bezerra, T. K. A., De Araujo, A. R. R., Do Nascimento, E. S., De Matos Paz, J. E., Gadelha,  
576 C. A., Gadelha, T. S., ... Madruga, M. S. (2016). Proteolysis in goat “coalho” cheese  
577 supplemented with probiotic lactic acid bacteria. *Food Chemistry*, *196*, 359–366.  
578 <https://doi.org/10.1016/j.foodchem.2015.09.066>

579 Broadbent, J. R., Cai, H., Larsen, R. L., Hughes, J. E., Welker, D. L., De Carvalho, V. G., ...  
580 Steele, J. L. (2011). Genetic diversity in proteolytic enzymes and amino acid metabolism  
581 among *Lactobacillus helveticus* strains. *Journal of Dairy Science*, *94*(9), 4313–4328.  
582 <https://doi.org/10.3168/jds.2010-4068>

583 Courtin, P., Monnet, V., & Rul, F. (2002). Cell-wall proteinases PrtS and PrtB have a  
584 different role in *Streptococcus thermophilus*/*Lactobacillus bulgaricus* mixed cultures in  
585 milk. *Microbiology*, *148*(11), 3413–3421.

586 Dallas, D. C., Citerne, F., Tian, T., Silva, V. L. M., Kalanetra, K. M., Frese, S. A., ... Barile,  
587 D. (2016). Peptidomic analysis reveals proteolytic activity of kefir microorganisms on  
588 bovine milk proteins. *Food Chemistry*, *197*, 273–284.  
589 <https://doi.org/10.1016/j.foodchem.2015.10.116>

590 De Man, J. C., Rogosa, M., & Sharpe, M. E. (1960). A medium used for the cultivation of  
591 *Lactobacilli*. *J Appl Bacteriol*, *23*(1), 130–135.

592 Dumas, B. R., Brignon, G., Grosclaude, F., & Mercier, J.-C. (1972). Structure primaire de la

593 caseine bovine: Sequence complete. *European Journal of Biochemistry*, 25(3), 505–514.  
594 <https://doi.org/10.1111/j.1432-1033.1972.tb01722.x>

595 Gandhi, A., & Shah, N. P. (2014). Cell growth and proteolytic activity of *Lactobacillus*  
596 *acidophilus*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii* ssp. *bulgaricus* , and  
597 *Streptococcus thermophilus* in milk as affected by supplementation with peptide  
598 fractions. *International Journal of Food Sciences and Nutrition*, 65(8), 937–941.  
599 <https://doi.org/10.3109/09637486.2014.945154>

600 Genay, M., Sadat, L., Gagnaire, V., & Lortal, S. (2009). PrtH2, Not prtH, Is the ubiquitous  
601 cell wall proteinase gene in *Lactobacillus helveticus*. *Applied and Environmental*  
602 *Microbiology*, 75(10), 3238–3249. <https://doi.org/10.1128/AEM.02395-08>

603 Giacometti, J., & Buretić-Tomljanović, A. (2017). Peptidomics as a tool for characterizing  
604 bioactive milk peptides. *Food Chemistry*, 230, 91–98.  
605 <https://doi.org/10.1016/j.foodchem.2017.03.016>

606 Griffiths, M. W., & Tellez, a. M. (2013). *Lactobacillus helveticus*: The proteolytic system.  
607 *Frontiers in Microbiology*, 4(MAR), 1–9. <https://doi.org/10.3389/fmicb.2013.00030>

608 Hou, J.-C., Liu, F., Ren, D.-X., Han, W.-W., & Du, Y.-O. (2015). Effect of culturing  
609 conditions on the expression of key enzymes in the proteolytic system of *Lactobacillus*  
610 *bulgaricus*. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 16(4), 317–326.  
611 <https://doi.org/10.1631/jzus.B1400230>

612 Jensen, M. P., Vogensen, F. K., & Ardö, Y. (2009). Variation in aminopeptidase and  
613 aminotransferase activities of six cheese related *Lactobacillus helveticus* strains.  
614 *International Dairy Journal*, 19, 661–668. <https://doi.org/10.1016/j.idairyj.2009.09.007>

615 Khaldi, N., Vijayakumar, V., Dallas, D. C., Guerrero, A., Wickramasinghe, S., Smilowitz, J.

616 T., ... German, J. B. (2014). Predicting the important enzymes in human breast milk  
617 digestion. *Journal of Agricultural and Food Chemistry*, 62(29), 7225–7232.  
618 <https://doi.org/10.1021/jf405601e>

619 Korhonen, H., & Pihlanto, A. (2006). Bioactive peptides: Production and functionality.  
620 *International Dairy Journal*, 16(9), 945–960.  
621 <https://doi.org/10.1016/j.idairyj.2005.10.012>

622 Lacou, L., Lê, S., Pezennec, S., & Gagnaire, V. (2015). An in silico approach to highlight  
623 relationships between a techno-functional property of a dairy matrix and a peptide  
624 profile. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 475(1), 44–  
625 54. <https://doi.org/10.1016/j.colsurfa.2014.10.052>

626 Lacroix, I. M. E., & Li-Chan, E. C. Y. (2012). Evaluation of the potential of dietary proteins  
627 as precursors of dipeptidyl peptidase (DPP)-IV inhibitors by an in silico approach.  
628 *Journal of Functional Foods*, 4(2), 403–422. <https://doi.org/10.1016/j.jff.2012.01.008>

629 Lowry, O. H., Rosebrough, N. J., Farr, L. A., & Randall, R. J. (1951). Protein measurement  
630 with the folin phenol reagent. *The Journal of Biological Chemistry*, 193(1), 265–275.  
631 [https://doi.org/10.1016/0304-3894\(92\)87011-4](https://doi.org/10.1016/0304-3894(92)87011-4)

632 McSweeney, P. L. H., & Fox, P. F. (1997). Chemical methods for the characterization of  
633 proteolysis in cheese during ripening. *Le Lait*, 77(1), 41–76.  
634 <https://doi.org/10.1051/lait:199713>

635 Miyamoto, M., Ueno, H. M., Watanabe, M., Tatsuma, Y., Seto, Y., Miyamoto, T., &  
636 Nakajima, H. (2015). Distinctive proteolytic activity of cell envelope proteinase of  
637 *Lactobacillus helveticus* isolated from airag, a traditional Mongolian fermented mare's  
638 milk. *International Journal of Food Microbiology*, 197, 65–71.  
639 <https://doi.org/10.1016/j.ijfoodmicro.2014.12.012>

640 Narva, M., Halleen, J., Väänänen, K., & Korpela, R. (2004). Effects of *Lactobacillus*  
641 *helveticus* fermented milk on bone cells in vitro. *Life Sciences*, *75*, 1727–1734.  
642 <https://doi.org/10.1016/j.lfs.2004.04.011>

643 Nielsen, S. D., Beverly, R. L., Qu, Y., & Dallas, D. C. (2017). Milk bioactive peptide  
644 database: A comprehensive database of milk protein-derived bioactive peptides and  
645 novel visualization. *Food Chemistry*, *232*, 673–682.  
646 <https://doi.org/10.1016/j.foodchem.2017.04.056>

647 Nongonierma, A. B., & FitzGerald, R. J. (2015). The scientific evidence for the role of milk  
648 protein-derived bioactive peptides in humans: A Review. *Journal of Functional Foods*,  
649 *17*, 640–656. <https://doi.org/10.1016/j.jff.2015.06.021>

650 Oberg, C. J., Broadbent, J. R., Strickland, M., & McMahon, D. J. (2002). Diversity in  
651 specificity of the extracellular proteinases in *Lactobacillus helveticus* and *Lactobacillus*  
652 *delbrueckii* subsp. *bulgaricus*. *Letters in Applied Microbiology*, *34*(6), 455–460.  
653 <https://doi.org/10.1046/j.1472-765X.2002.01123.x>

654 Pailin, T., Kang, D. H., Schmidt, K., & Fung, D. Y. C. (2001). Detection of extracellular  
655 bound proteinase in EPS-producing lactic acid bacteria cultures on skim milk agar.  
656 *Letters in Applied Microbiology*, *33*(1), 45–49. [https://doi.org/10.1046/j.1472-](https://doi.org/10.1046/j.1472-765X.2001.00954.x)  
657 [765X.2001.00954.x](https://doi.org/10.1046/j.1472-765X.2001.00954.x)

658 Park, Y. W., & Nam, M. S. (2015). Bioactive Peptides in Milk and Dairy Products: A  
659 Review. *Korean Journal for Food Science of Animal Resources*, *35*(6), 831–840.  
660 <https://doi.org/10.5851/kosfa.2015.35.6.831>

661 Raak, N., Rohm, H., & Jaros, D. (2017). Enzymatic Cross-Linking of Casein Facilitates Gel  
662 Structure Weakening Induced by Overacidification. *Food Biophysics*, *12*(2), 1–8.  
663 <https://doi.org/10.1007/s11483-017-9483-6>

664 Sadat-Mekmene, L., Jardin, J., Corre, C., Mollé, D., Richoux, R., Delage, M. M., ...  
665 Gagnaire, V. (2011). Simultaneous presence of PrtH and PrtH2 proteinases in  
666 *Lactobacillus helveticus* strains improves breakdown of the pure  $\alpha$ s1-casein. *Applied and*  
667 *Environmental Microbiology*, 77, 179–186. <https://doi.org/10.1128/AEM.01466-10>

668 Sadat-Mekmene, Leila, Genay, M., Atlan, D., Lortal, S., & Gagnaire, V. (2011). Original  
669 features of cell-envelope proteinases of *Lactobacillus helveticus*. A review. *International*  
670 *Journal of Food Microbiology*, 146(1), 1–13.  
671 <https://doi.org/10.1016/j.ijfoodmicro.2011.01.039>

672 Savijoki, K., Ingmer, H., & Varmanen, P. (2006). Proteolytic systems of lactic acid bacteria.  
673 *Applied Microbiology and Biotechnology*, 71(4), 394–406.  
674 <https://doi.org/10.1007/s00253-006-0427-1>

675 Shuangquan, Tsuda, H., & Miyamoto, T. (2008). Angiotensin I-converting enzyme inhibitory  
676 peptides in skim milk fermented with *Lactobacillus helveticus* 130B4 from camel milk in  
677 Inner Mongolia, China. *Journal of Science of Food and Agriculture*, 88(2), 2688–2692.  
678 <https://doi.org/10.1002/jsfa>

679 Watanabe, K., Fujimoto, J., Sasamoto, M., Dugersuren, J., Tumursuh, T., & Demberel, S.  
680 (2008). Diversity of lactic acid bacteria and yeasts in Airag and Tarag, traditional  
681 fermented milk products of Mongolia. *World Journal of Microbiology and*  
682 *Biotechnology*, 24(8), 1313–1325. <https://doi.org/10.1007/s11274-007-9604-3>

683 Wu, S., Qi, W., Li, T., Lu, D., Su, R., & He, Z. (2013). Simultaneous production of multi-  
684 functional peptides by pancreatic hydrolysis of bovine casein in an enzymatic membrane  
685 reactor via combinational chromatography. *Food Chemistry*, 141(3), 2944–2951.  
686 <https://doi.org/10.1016/j.foodchem.2013.05.050>

687 Yu, J., Wang, W. H., Menghe, B. L. G., Jiri, M. T., Wang, H. M., Liu, W. J., ... Zhang, H. P.

688 (2011). Diversity of lactic acid bacteria associated with traditional fermented dairy  
689 products in Mongolia. *Journal of Dairy Science*, 94(7), 3229–3241.  
690 <https://doi.org/10.3168/jds.2010-3727>

691

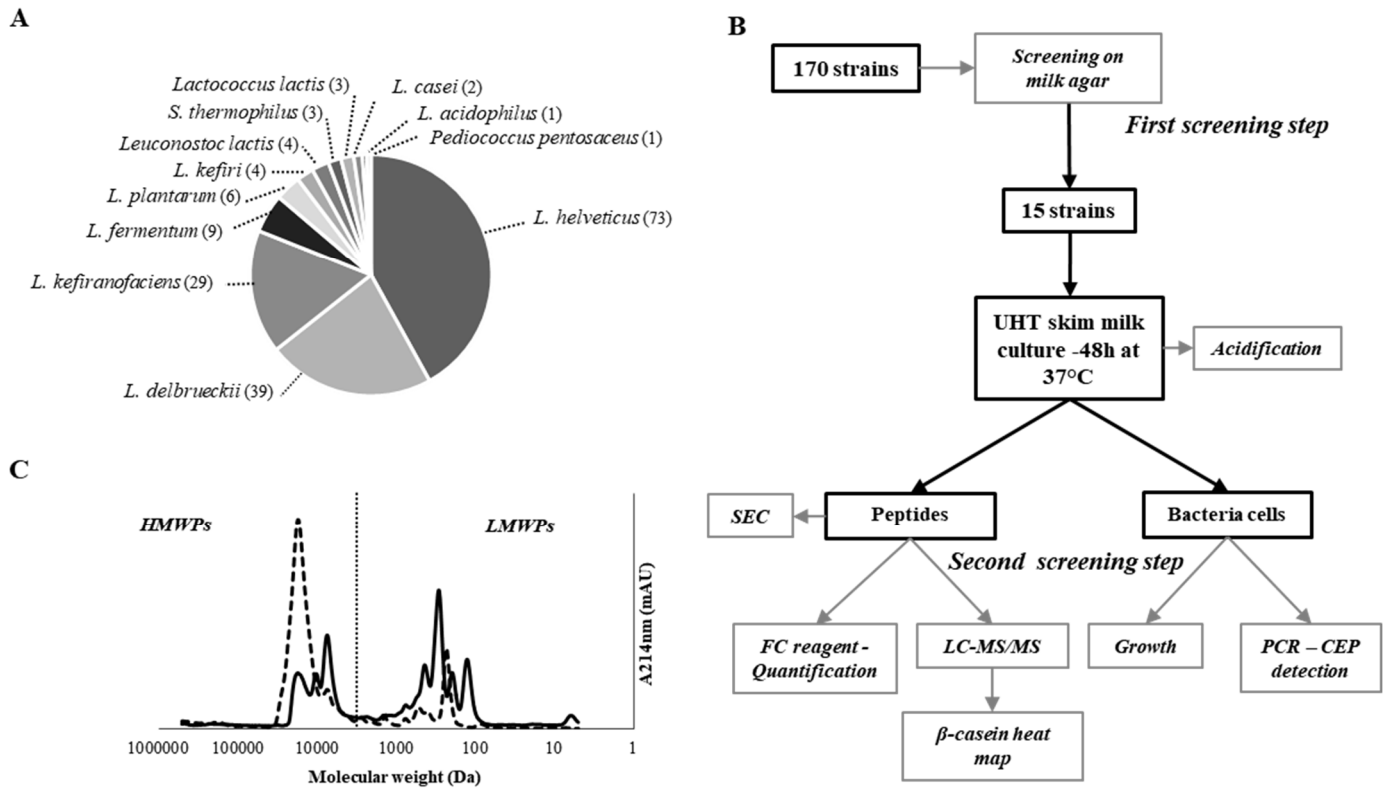
## 692 **Highlights**

- 693 • Mongolian strains show specific skim milk fermentation capacities
- 694 • Cell envelope associated peptidase genes distribution is different in Mongolian  
695 *Lactobacillus helveticus* strains than in Western strains
- 696 • Peptide production is closely correlated with growth and acidification patterns
- 697 • Patterns of  $\beta$ -casein hydrolysis are correlated with cell envelope associated peptidase  
698 genes profiles

699

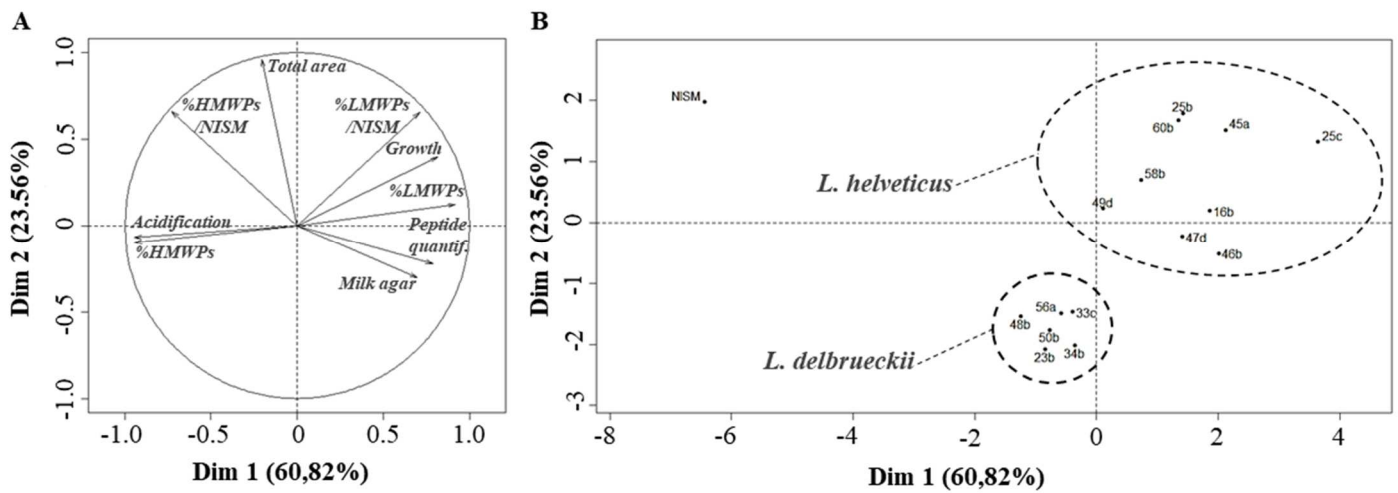
700

**Figure 1 :**



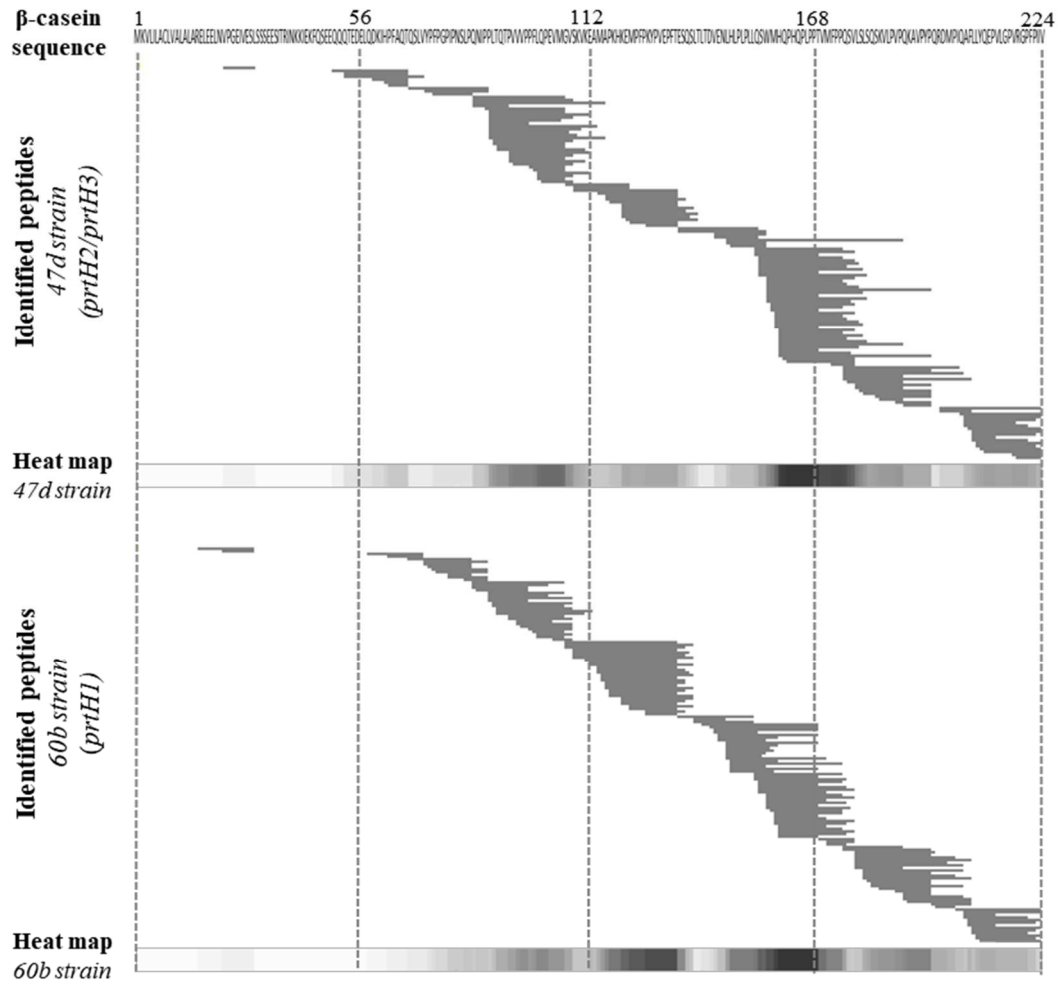
**Fig. 1.** (A) Relative abundance of bacterial species in the Mongolian collection identified by 16S rRNA sequencing (*L.* : *Lactobacillus*). (B) Schematic diagram of the screening strategy. A first phase of screening on milk agar was used to select 15 strains. The selected strains were then studied on UHT milk cultures. Different parameters were considered (growth, acidification, CEP gene detection, peptide identification, quantification and molecular weight profiling). (C) SEC profile of peptides extracted from non-inoculated milk (dash line) and an inoculated one (solid line) on a Superdex Peptide 10/300 GL column.

**Figure 2 :**



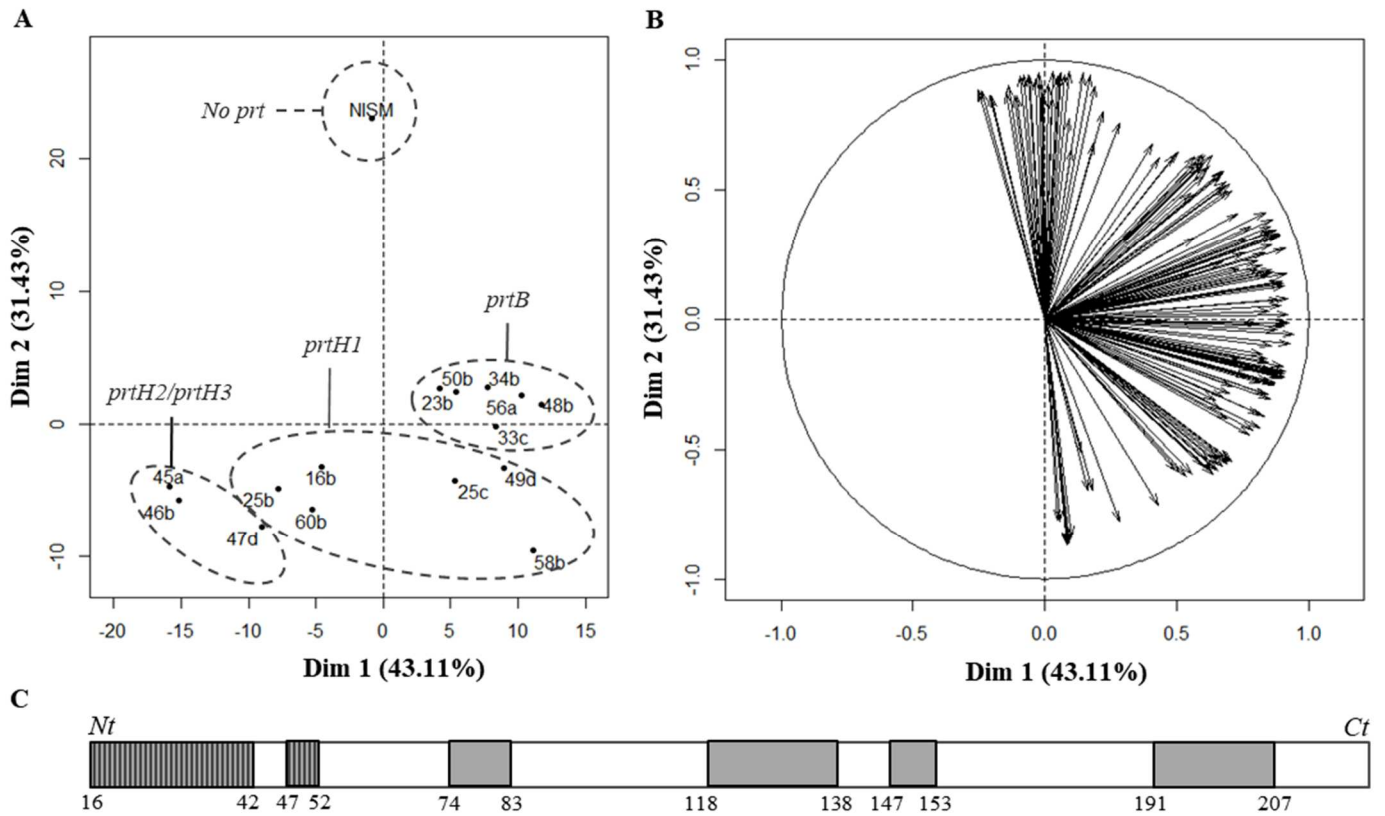
**Fig. 2.** Principal component analysis of variables measured after UHT skim milk fermentation for each of the 15 strains. Variables were divided into two principal components (PC1 and PC2). **(A)** Variable factor map representing correlations between parameters. **(B)** Individual factor map representing the repartition of the selected strains and NISM according to PC1 and PC2.

**Figure 3:**



**Fig. 3.**  $\beta$ -casein heat maps construction from identified peptides by LC-MS/MS. Examples of *L. helveticus* strains 47d (harboring prtH2/prtH3 CEP genes) and 60b (harboring prtH1 CEP gene). Heat maps under the identified peptides indicate the density of the amino acids. A color was associated to each amino acid from white (no apparition frequency) to dark grey (high apparition frequency).

**Figure 4:**



**Fig. 4.** Principal component analysis of  $\beta$ -casein heat maps constructed from identified peptides. Variables were divided into two principal components (PC1 and PC2). (A) Individual factor map representing the repartition of each selected strains according to PC1 and PC2. Results from CEP gene detection by PCR were reported on the map (circles). (B) Variable factor map representing the most implicated variables (defined by  $\cos^2 > 0.85$ ) on the PCA. Among all variables (209 amino acids), 51 and 31 were the most implicated in PC1 and PC2 respectively. (C) Schematic representation of  $\beta$ -casein with highlighted areas representing the most implicated variables for PC1 (51 amino acids, grey) and for PC2 (31 amino acids, grey striped).

**Table 1.** Results of the strain screenings

Strain	Species	Ø milk agar (mm)	Growth (OD)	Acidification (pH)	FC reagent (g.L <sup>-1</sup> )	Total area (mAU)	%HMWPs	%LMWPs	%HMWPs/NISM	%LMWPs/NISM
NISM	×	×	×	6.45 ± 0.05 <sup>e</sup>	2.06 ± 0.62 <sup>a</sup>	347.51 ± 42.8 <sup>a</sup>	85.62 ± 2.72 <sup>g</sup>	14.39 ± 2.72 <sup>a</sup>	100 <sup>d</sup>	100 <sup>a</sup>
23b	<i>L. delbrueckii</i>	33 ± 3 <sup>ac</sup>	0.739 ± 0.01 <sup>ac</sup>	4.52 ± 0.02 <sup>c</sup>	4.86 ± 0.21 <sup>ad</sup>	101.48 ± 31.8 <sup>a</sup>	61.37 ± 3.44 <sup>df</sup>	40.06 ± 5.13 <sup>bc</sup>	19.74 ± 5.66 <sup>ab</sup>	123.32 ± 52.81 <sup>a</sup>
33c	<i>L. delbrueckii</i>	32 ± 2 <sup>ac</sup>	0.825 ± 0.126 <sup>ac</sup>	4.64 ± 0.02 <sup>c</sup>	5.54 ± 0.61 <sup>bcd</sup>	141.65 ± 27.2 <sup>a</sup>	54.83 ± 1.28 <sup>cde</sup>	45.98 ± 2.43 <sup>bc</sup>	24.69 ± 4.3 <sup>ab</sup>	194.64 ± 47.3 <sup>a</sup>
34b	<i>L. delbrueckii</i>	39 ± 3.5 <sup>c</sup>	0.987 ± 0.03 <sup>bc</sup>	4.52 ± 0.02 <sup>c</sup>	5.32 ± 0.21 <sup>ad</sup>	115.96 ± 4.35 <sup>a</sup>	59.07 ± 3.17 <sup>df</sup>	41.08 ± 3.35 <sup>bc</sup>	21.79 ± 0.37 <sup>ab</sup>	141.83 ± 16.75 <sup>a</sup>
48b	<i>L. delbrueckii</i>	30 ± 0 <sup>ac</sup>	0.473 ± 0.118 <sup>ab</sup>	4.68 ± 0.01 <sup>c</sup>	3.59 ± 0.19 <sup>ad</sup>	123.56 ± 12.3 <sup>a</sup>	57.26 ± 1.76 <sup>de</sup>	44.14 ± 2.52 <sup>bc</sup>	22.50 ± 1.65 <sup>ab</sup>	162.65 ± 24.69 <sup>a</sup>
50b	<i>L. delbrueckii</i>	30 ± 1.5 <sup>ab</sup>	0.973 ± 0.199 <sup>bc</sup>	4.51 ± 0 <sup>c</sup>	4.89 ± 0.3 <sup>bcd</sup>	114.70 ± 12.5 <sup>a</sup>	58.12 ± 6.22 <sup>de</sup>	40.90 ± 4.26 <sup>bc</sup>	21.12 ± 1.81 <sup>ab</sup>	140.37 ± 28.22 <sup>a</sup>
56a	<i>L. delbrueckii</i>	28 ± 1.1 <sup>a</sup>	1.14 ± 0.124 <sup>c</sup>	4.52 ± 0.01 <sup>c</sup>	5.58 ± 0.29 <sup>bcd</sup>	131.81 ± 35.8 <sup>a</sup>	58.35 ± 3.93 <sup>df</sup>	42.32 ± 5.03 <sup>bc</sup>	24.21 ± 4.98 <sup>ab</sup>	169.25 ± 65.55 <sup>a</sup>
<b>Average</b>	<b><i>L. delbrueckii</i></b>	<b>32 ± 3.8</b>	<b>0.856 ± 0.234</b>	<b>4.57 ± 0.07</b>	<b>4.96 ± 0.74</b>	<b>121.53 ± 14.1</b>	<b>58.17 ± 2.1</b>	<b>42.4 ± 2.3</b>	<b>22.34 ± 1.9</b>	<b>155.34 ± 25.39</b>
16b	<i>L. helveticus</i>	31 ± 1.5 <sup>ab</sup>	2.79 ± 0.03 <sup>de</sup>	3.56 ± 0.05 <sup>b</sup>	4.49 ± 0.02 <sup>ac</sup>	184.11 ± 46.7 <sup>a</sup>	34.14 ± 1.98 <sup>ab</sup>	73.24 ± 2.06 <sup>de</sup>	19.89 ± 4.18 <sup>ab</sup>	400.81 ± 103.87 <sup>ab</sup>
25b	<i>L. helveticus</i>	32 ± 1 <sup>ac</sup>	3.18 ± 0.194 <sup>df</sup>	3.48 ± 0.1 <sup>ab</sup>	4.17 ± 0.68 <sup>ad</sup>	155.52 ± 4.92 <sup>a</sup>	38.83 ± 2.2 <sup>ab</sup>	64.62 ± 1.89 <sup>de</sup>	15.39 ± 0.6 <sup>a</sup>	503.42 ± 297.66 <sup>ab</sup>
25c	<i>L. helveticus</i>	32 ± 2 <sup>ac</sup>	2.62 ± 0.138 <sup>d</sup>	3.45 ± 0.17 <sup>ab</sup>	6.78 ± 1.63 <sup>d</sup>	268.7 ± 121.8 <sup>a</sup>	20.13 ± 3.6 <sup>a</sup>	85.7 ± 7.47 <sup>c</sup>	16.91 ± 6.77 <sup>a</sup>	678.54 ± 309.89 <sup>b</sup>
45a	<i>L. helveticus</i>	32 ± 4 <sup>ac</sup>	3.57 ± 0.12 <sup>f</sup>	3.32 ± 0.04 <sup>a</sup>	6.16 ± 1.32 <sup>bcd</sup>	289.95 ± 52 <sup>a</sup>	40 ± 3.74 <sup>ad</sup>	60.1 ± 3.77 <sup>cd</sup>	37.35 ± 11.95 <sup>ac</sup>	514.7 ± 86.67 <sup>ab</sup>
46b	<i>L. helveticus</i>	35 ± 2.6 <sup>ac</sup>	3.03 ± 0.02 <sup>df</sup>	3.39 ± 0.1 <sup>ab</sup>	4.44 ± 0.17 <sup>ad</sup>	139.24 ± 6.51 <sup>a</sup>	37 ± 3.1 <sup>abc</sup>	74.2 ± 3.67 <sup>de</sup>	16.37 ± 0.74 <sup>a</sup>	307.41 ± 27.56 <sup>ab</sup>
47d	<i>L. helveticus</i>	32 ± 1 <sup>ac</sup>	3.12 ± 0.326 <sup>df</sup>	3.34 ± 0.04 <sup>ab</sup>	5.01 ± 0.8 <sup>ad</sup>	170.7 ± 52.9 <sup>a</sup>	44.73 ± 12.24 <sup>bd</sup>	58.91 ± 16.94 <sup>cd</sup>	23.93 ± 7.31 <sup>ab</sup>	297.41 ± 114.8 <sup>ab</sup>
49d	<i>L. helveticus</i>	35 ± 4.3 <sup>ac</sup>	2.72 ± 0.05 <sup>d</sup>	3.43 ± 0.01 <sup>ab</sup>	4.46 ± 0.46 <sup>ad</sup>	225.08 ± 118.3 <sup>a</sup>	58.56 ± 14.51 <sup>ef</sup>	42.19 ± 14.63 <sup>b</sup>	43.87 ± 29.24 <sup>bc</sup>	263.68 ± 77.95 <sup>ab</sup>
58b	<i>L. helveticus</i>	37 ± 1.7 <sup>ab</sup>	3.403 ± 0.394 <sup>ef</sup>	3.44 ± 0.01 <sup>ab</sup>	5.95 ± 1.36 <sup>cd</sup>	263.89 ± 21.7 <sup>a</sup>	59.24 ± 8.93 <sup>df</sup>	40.79 ± 8.95 <sup>bc</sup>	49.48 ± 4.74 <sup>c</sup>	322.98 ± 93.51 <sup>ab</sup>
60b	<i>L. helveticus</i>	35 ± 1 <sup>ac</sup>	2.81 ± 0.257 <sup>de</sup>	3.49 ± 0.11 <sup>ab</sup>	5.09 ± 0.91 <sup>ad</sup>	303.15 ± 165.9 <sup>a</sup>	46.4 ± 6.48 <sup>bd</sup>	59.41 ± 8.02 <sup>cd</sup>	29.10 ± 7.85 <sup>ac</sup>	526.87 ± 274.97 <sup>ab</sup>
<b>Average</b>	<b><i>L. helveticus</i></b>	<b>33.4 ± 2.1</b>	<b>3.03 ± 0.322</b>	<b>3.43 ± 0.08</b>	<b>5.17 ± 0.92</b>	<b>222.26 ± 61.7</b>	<b>42.11 ± 12.1</b>	<b>62.13 ± 14.6</b>	<b>28.03 ± 12.8</b>	<b>423.98 ± 139.63</b>

Diameters obtained on skim milk agar during the first phase of screening for the 15 selected strains. Results from the second phase of screening on the selected strains obtained after UHT skim milk fermentation at 37°C for 48h: growth, acidification, peptide quantification with FC reagent and integration results from chromatogram obtained by SEC ((high molecular weight proteins/peptides (HMWPs); low molecular weight peptides (LMWPs)). NISM: Non-Inoculated UHT Skim Milk. \*Statistical analysis using one way ANOVA and Tukey's post-hoc test. Mean values in the same column with different uppercase subscripts differ significantly (p < 0.05).

**Table 2.** CEP gene distribution

<b>Strain</b>	<b>Specie</b>	<b><i>prtH1</i></b>	<b><i>prtH2</i></b>	<b><i>prtH3</i></b>	<b><i>prtH4</i></b>	<b><i>prtB</i></b>	<b>Isolated from</b>
23b	<i>L. delbrueckii</i>					+	Cow milk
33c	<i>L. delbrueckii</i>					+	Cow milk
34b	<i>L. delbrueckii</i>					+	Mare milk
48b	<i>L. delbrueckii</i>					+	Cow milk
50b	<i>L. delbrueckii</i>					+	Yak milk
56a	<i>L. delbrueckii</i>					+	Yak milk
CNRZ32	<i>L. helveticus</i>	+	+	+	+		CIRM (INRA)
16b	<i>L. helveticus</i>	+					Mare milk
25b	<i>L. helveticus</i>	+					Mare milk
25c	<i>L. helveticus</i>	+					Mare milk
49d	<i>L. helveticus</i>	+					Mare milk
58b	<i>L. helveticus</i>	+					Mare milk
60b	<i>L. helveticus</i>	+					Mare milk
45a	<i>L. helveticus</i>		+	+			Yak milk
46b	<i>L. helveticus</i>		+	+			Yak+Cow milk
47d	<i>L. helveticus</i>		+	+			Camel milk

Distribution of CEP genes detected by PCR (*prtH1*, *prtH2*, *prtH3*, *prtH4* and *prtB*) in the 15 selected strains and in the reference strain CNRZ32. Five primer set were used for the PCR.