Bacterial diversity of the Colombian fermented milk "Suero Costeño" assessed by culturing and high-throughput sequencing and DGGE analysis of 16S rRNA gene amplicons

Abstract

"Suero Costeño" (SC) is a traditional soured cream elaborated from raw milk in the Northern-Caribbean cost of Colombia. The natural microbiota that characterizes this popular Colombian fermented milk is unknown, although several culturing studies have been previously attempted. In this work, the microbiota associated to SC from three manufacturers of two regions, "Planeta Rica" (Córdoba) and "Caucasia" (Antioquia), was analysed by means of culturing methods in combination with high-throughput sequencing and DGGE analysis of 16S rRNA gene amplicons. The bacterial ecosystem of SC samples was revealed to be composed by lactic acid bacteria belonging to the *Streptococcaceae* and *Lactobacillaceae* families, varying the proportions and genera among manufacturers and region of elaboration. Members of the *Lactobacillus acidophilus* group, *Lactococcus lactis*, *Streptococcus infantarius* and *Streptococcus salivarius* characterized this artisanal product. In comparison with culturing, the use of molecular in deep culture-independent techniques allows a more realistic picture of the overall bacterial communities residing SC. Besides the descriptive purpose, these approaches will facilitate the rational strategy to follow (culture media and growing conditions) for the isolation of indigenous strains for the standardization in the manufacture of SC.

Keywords: Suero Costeño; fermented milk; bacterial diversity; *Streptococcus infantarius*.

1. Introduction

Fermentation is one of the most traditional ways to transform raw foods worldwide, which confers to the matrix specials features, such as desirable sensorial characteristics, functional properties and prolonged shelf-life, given by the dynamics of the microbial community living there (Leroy and De Vuyst, 2004). "Suero Costeño" (SC) is an artisanal soured cream produced in the Northern-Caribbean cost of Colombia. It is widely produced by the rural population in the regions of Córdoba, Sucre and Bolívar, and represents a gastronomic heritage of Colombian cuisine (Simanca et al., 2010). Due to its specific organoleptic properties, similar to a moderate acid soured cream with a salty taste, SC is usually used as a dressing accompanying most meals. The artisanal product is usually obtained by the natural fermentation (between 12-24 h depending of the manufacturer's practices and the desired viscosity) at ambient temperature (~30ºC) of raw-cow milk, which is initiated through continuous reutilization of fermentation containers, such as calabash (for example "Totumo" seed *Crescentia cujete*) or plastic vessels. The high environmental humidity (greater than 74%), together with the high microbial load in raw milk, as well as the inoculation from the containers and the surrounding air, favours this natural fermentation. During fermentation a characteristic liquid-solid two phase system is formed, the whey is removed and the cream like thickness phase (precipitated casein curd) is mixed and salted. The final product is an acidic, soured and salty fermented-milk with a creamy consistency (Chaves-López et al., 2014).

Until now very little information is available about the microbiota inhabiting this particular Colombian fermented dairy product. Previous works, by culturing techniques, have reflected the poor hygienic-sanitary conditions in the elaboration due to the presence of elevated numbers of enterobacteria, coliforms and staphylococci (Chams et

al., 2012; Chaves-López et al., 2014; Grandos-Conde et al., 2012). Manufacture standardization, through the utilization of pasteurized milk and commercial dairy cultures, would reduce the sanitary problem but, unfortunately, this will lead to losses in the typical sensorial properties of SC. Cueto and co-workers (2007) have carried out preliminary culture analyses about the occurrence of indigenous lactic acid bacteria (LAB) in this product, with the identification of *Lactobacillus* species as being the predominant LAB members. However, a description of the whole bacterial composition present in the fermented product has not been undertaken to date. This should be the starting point to the search for and to isolate autochthonous LAB that could form part of a specific designed SC starter, helping to maintain the sanitary conditions and particular organoleptic properties of this product.

Culture-dependent methods must be used to isolate new strains with potential interest, but they are rather unreliable to investigate the overall diversity of the microbial communities present in a fermented food ecosystem. In the last decade, culture-independent molecular techniques, such as denaturing gradient gel electrophoresis (DGGE), real-time quantitative PCR (qPCR) or the construction and analysis of 16S rRNA gene libraries, have contributed to the description of the microbial ecology of dairy products (Aldrete-Tapia et al., 2014; Jany and Barbier, 2008; Liu et al., 2015; Quigley et al., 2011). With the development of high-throughput sequencing (HTS) technologies, a new era in microbial ecology applied to food systems has been initiated (Bokulich and Mills, 2012). Most of the investigations performed until now allowed the identification of the bacterial communities through the massive sequencing of 16S rRNA gene amplicons, which makes achievable in depth studies of complex communities (Ercolini, 2013; Mayo et al., 2014).

In this work, we have combined the application of culture-independent techniques (DGGE and HTS analysis of 16S rDNA amplicons) for the whole bacterial characterization of artisanal SC, together with culturing methods in order to confirm the presence of viable taxa, as well as to isolate representative strains with potential technological interest for the standardization in the elaboration of this product.

2. Material and methods

2.1. "Suero Costeño" fermented milk samples

Six artisanal SC productions, made by three manufacturers (named as P1, P2 and P3) from two Caribbean regions of Northern Colombia, "Planeta Rica" in Córdoba and "Caucasia" in Antioquia (Figure 1 and Suppl. 1), were studied during February-March 2015. In the case of "Caucasia" producer (P1), the milk was inoculated with a portion (30%) of a previous batch using the back-slopping technique, whereas the fermentation occurred spontaneously in the SC elaborated by producers P2 and P3 from "Planeta Rica". The pH of the milk before and after fermentation was measured and time for fermentation (12 h for P1, 16 h for P3 and 24 h for P2) was considered when the isoelectric point of the caseins was reached (around pH 4.6) causing the coagulation of milk. Two different batches (A and B) manufactured in different days were obtained from each producer. Samples of SC were collected from different fermentation containers (plastic vessels or "Totumo" seed) and transported to the laboratory, located at Medellin (about 286 km far), under refrigerated (4ºC) conditions.

2.2. Culturing

2.2.1. Sample analysis and microbial counts in different media

The refrigerated samples (10 g) were homogenized in 90 mL of 2% sodium citrate (Merck, Darmstadt, Germany) and serial decimal dilutions were made on peptone water solution (Merck). These dilutions were plated, in duplicate, on the surface

of the different agar-media as follows: i) “Plate Count Agar” (PCA, Merck) for total aerobic counts, ii) “Man, Rogosa, and Sharpe” (MRS, Merck) and M17 (Merck) with 1% lactose (M17L) for counts of LAB, in particular lactobacilli and lactococci respectively, and iii) “Baird Parker” (BP, Merck) for staphylococci counts. All plates were incubated in aerobic conditions at 32°C for 24-48 h. For yeast and molds counts, samples were plated on Oxytetracycline-Glucose-Yeast Extract (OGYE, Merck) and incubated in aerobic conditions at 25°C for 120 h. Additionally, the presence of *Listeria* spp. was determined on “Listeria Selective Oxford Agar Base” (Oxoid, Thermo Fisher Scientific Inc., Waltham, MA. USA) after incubation at 37ºC for 48 h. The most probable number (MPN) of total coliforms and *Escherichia coli* was also evaluated in “Fluorocult LMX” broth (Merck) using 3 tubes of media for three different dilutions of each sample of study. After incubation at 37ºC for 24 h the tubes that presented color alteration from the normal yellow to greenish blue were considered positive for total coliforms, whereas the tubes with fluorescence were considered positive for *E. coli*. The MPN was determined using the corresponding table (Swanson et al., 2001).

Finally, a portion (10 ± 0.5 g) of the fermented SC samples obtained from the three manufacturers (2 batches each) was lyophilised in a Labconco® (Kansas City, MO, USA) lyophilizer before being delivered to “Instituto de Productos Lácteos de Asturias - Consejo Superior de Investigaciones Científicas” (IPLA-CSIC), in Spain, for molecular analysis.

2.2.2. LAB isolation and identification

Colonies, representatives of all morphologies and sizes from the MRS and M17L counting plates, were randomly picked and consecutively streaked in plate twice in the same isolation medium. Then, a single colony was inoculated in the corresponding liquid medium and incubated for 24 h at 32°C. The grown cultures were preserved at -

80°C in each specific medium with 20% glycerol, and lyophilised in skim milk (reconstituted at 11%) to be sent to IPLA-CSIC in Spain.

In order to obtain DNA for bacterial identification, the isolates were cultured in liquid media (MRS or M17) and 1 mL of each overnight culture was processed using the GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich, Sigma Chemical Co., St. Louis, MO). Total DNA was employed as template to amplify by PCR a segment of the 16S rRNA gene. Amplifications were performed with the universal prokaryotic primers Y1 and Y2, which flanked a portion of approximately 350 bp including the variable regions V1 and V2 (Young et al., 1991), with the 5´ Prime Taq polymerase (VWR International, Radnor, PA. USA) and the following conditions: 95°C for 5 min, 35 cycles of 94°C for 30 s, 50°C for 45 s, 72°C for 2 min, and a final extension step at 72°C for 5 min. Additionally, for a proper identification of streptococci at species level, amplification and sequencing of *sodA* gene was done. PCR reactions for an internal fragment (438 bp) of the *sodA* gene, encoding the manganese-dependent superoxide dismutase, were accomplished with the degenerate primers *d1* and *d2* described elsewhere (Poyart et al., 2000) and the following conditions: 95°C for 3 min, 30 cycles of 95°C for 30 s, 42°C for 60 s, 72°C for 60 s, and a final extension step at 72°C for 7 min. Amplicons were sent for sequencing by cycle extension in an ABI DNA sequencer to the company Macrogen (Seoul, Rep. of Korea). Sequences were compared against databases of NCBI using the BLAST program (Camacho et al., 2009) and the Ribosomal Database Project (RDP) (Wang et al., 2007); those with a percentage of nucleotide identity equal or higher than 99% were assigned to species level.

2.2.3. Statistical analysis

Counts of the different microbial groups were statistically analysed using the IBM SPSS Statistics for Window Version 23.0 (IBM Corp., Armonk NY). One-way

ANOVA and the SNK (Student-Newman-Keuls) mean comparison test was used to determine differences among the three manufacturers ($p<0.05$).

2.3. Culture-independent bacterial analysis

2.3.1. DNA extraction of "Suero costeño" samples.

The six lyophilised samples of SC were reconstituted 1/10 in 2% sodium citrate. An additional 1/10 dilution was made in Maximum Recovery Diluent (MRD, Scharlab S.L., Spain) before the extraction of total DNA, which was performed using the QIAamp DNA stool kit (Qiagen, GmbH, Hilden, Germany) following an optimized protocol for isolation of total microbial DNA from milk (Delgado et al., 2013). DNA obtained from SC samples was resuspended in 50 µL of molecular-biology grade water (Sigma-Aldrich) and stored at -20°C until use. DNA concentration and quality was determined in an BioTek Epoch™ spectrophotometer system (Thermo Fisher Scientific).

2.3.2. PCR-denaturing gradient gel electrophoresis (DGGE) analysis.

Extracted DNA from the SC samples was used as template for amplification by PCR of the variable region V3 of the bacterial 16S rRNA gene, using the universal primers F357-GC and R518 as reported by Muyzer and collaborators (1993). PCR conditions were as follows: 95°C for 2 min, 35 cycles of 95°C for 30 s, 56°C for 40 s, 72°C for 1 min, and a final step at 72°C for 5 min.

Isolates from SC samples of the species *Lactobacillus plantarum*, *Enterococcus faecium*, *Streptococcus infantarius* and *Lactococcus lactis* were used to construct a DGGE ladder. Genomic DNA from these strains was obtained with the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich) and subsequently employed in amplifications involving the DGGE primers, as above mentioned. Amplicons were

purified using GenElute PCR Clean-Up columns (Sigma-Aldrich) and mixed in equal amounts to obtain the DGGE ladder.

DGGE was performed as previously described using a DCode system (Bio-Rad, Richmond, CA, USA), employing 8% polyacrylamide gels with a urea-formamide denaturing range of 40-60% (Leite et al., 2012). Electrophoresis was performed at 75 V for 16 h at 60°C. Bands were visualized under UV light after staining with ethidium bromide, and DGGE gels were captured with a GBox system (Syngene, Cambridge, UK) equipped with the GeneSys image acquisition software (Syngene). Representative bands were excised from the acrylamide gels and their DNA eluted overnight in 50 µL of sterile molecular grade water at 4°C. DNA was then re-amplified using the same primer pair without the GC-clamp. Purified amplicons were sequenced, as indicated above, and the identity of the bands (at the lowest possible taxonomic range, i.e. species or genus level) was established by comparison with the previously mentioned databases.

2.3.3. High-throughput 16S rDNA sequencing and analysis

Partial 16S rRNA gene sequences were amplified from the extracted DNA of the SC samples using the primer´s pair Probio_Uni and /Probio_Rev, which target the variable region V3 of the bacterial 16S rRNA gene, as previously described (Milani et al., 2013). Samples were submitted to 2 x 250 bp paired-end sequencing by means of an Illumina MiSeq System (Illumina, San Diego, CA, USA) at the DNA sequencing facility of GenProbio S.R.L. (Parma, Italy). Sequence reads were filtered by the Illumina software to remove low quality sequences. All Illumina quality-approved, trimmed and filtered sequences (average read length 194 bp) were processed using a custom script based on the QIIME software suite (Caporaso et al., 2010). After paired-end reads joining using the fastq-join software (Aronesty, 2013), the quality control phase retained sequences with a mean sequence quality score >20 and a length between

140 and 400 bp. Sequences with homopolymer regions >7 bp and those with mismatched primers were omitted. In order to calculate downstream diversity measures, 16S rDNA sequences were clustered into operational taxonomic units (OTUs), defined at ≥ 97 % sequence similarity using the UCLUST tool (Edgar, 2010). All sequences were classified to the lowest possible taxonomic rank using QIIME and the SILVA database as reference (Quast et al., 2013). Similarity of the bacterial communities between samples was calculated by the unweighted UniFrac method (Lozupone and Knight, 2005).

3. Results and discussion

3.1. Culturing analysis and LAB isolation

The microbial characterization of SC sour cream from two regions of North Colombia was initially investigated by culturing. The pH of the samples ranged from 3.8 (from P1) to 4.4 (from P2). Figure 2 shows the mean counts, and the corresponding statistical analysis, of the different microbial groups analysed in the SC samples from three different manufacturers. The overall bacterial load of the final products was significantly lower in SC samples from P1 of the "Caucasia" region. The total cultivatable aerobic counts in PCA, as well as LAB numbers in MRS and M17L, were statistically lower, i.e., from two to three logarithmic units, than those obtained from samples of the other two producers. It is worth mentioning that producer P1 used a back-slopping technique and the time of fermentation for the elaboration of SC was shorter (12 h) than in P2 and P3 from "Planeta Rica" (Figure 1). Back-slopping technique involves inoculating milk with a portion of a previous successful fermentation to the new batch, thus seeding the milk. The success of such fermentations depends on the particular blend of microorganisms present in the previous batch (Parente and

225 Cogan, 2004). The typical manufacture of SC from producers of "Planeta Rica" region
226 is non-inoculated; this means that a natural fermentation occurs by leaving the raw milk
227 at ambient temperature and fermentation occurred through the activity of the
228 autochthonous microorganisms present in milk and the ambient surrounding.
229 Accordingly, numbers of staphylococci were significantly higher in SC from this region
230 (~10^4 CFU/mL), in comparison to producer P1 (~10^2 CFU/mL), and presence of
231 coliforms and *Listeria* spp. was also noticed at a high level (Table S1). We assume that
232 this could be due to the fact that in these cases fermentation occurred naturally by
233 indigenous microorganisms present in milk or contaminants from manufacturing tools
234 and environment, meanwhile in P1 a 30% re-inoculum is used which might seeding the
235 milk and compete and overpass the microbiota present. The activity of successful
236 favourable dairy LAB types produced lactic acid which improves safety by inhibiting
237 spoilage and pathogenic microorganisms. Staphylococci, coliforms and pathogens
238 (*Staphylococcus aureus, Bacillus cereus, Listeria monocytogenes*) are occasionally
239 found in other natural fermented milks in warm countries (Gonfa et al., 2001), stressing
240 the need for improving the microbial quality and safety of these artisanal dairy products.
241 The methods employed in our case only revealed the presence of *Listeria* spp. by
242 culturing. Muñoz et al. (2011) reported the presence of *Listeria monocytogenes* in this
243 type of products in retail outlets and open markets in Bogotá, with a higher prevalence
244 in fresh cheese made with raw milk. In the case of BP media, staphylococci counts
245 obtained in our study represent minor population if comparing with total counts in PCA
246 and MRS/M17L with numbers varying from 10^6-10^8 CFU/mL (Figure 1). The highest
247 counts were retrieved in SC samples from P2 (up to 10^8 CFU/mL), in which a natural
248 seed ("Totumo") container was utilized and, thus, it could ultimately contribute to the
249 contamination of the milk substrate. The highest counts in OGYE were obtained in P3

products from “Planeta Rica” and according to some authors, yeasts can have an effect on fermentation of SC (Cueto et al., 2007). Then, we have identified 5 different types of isolates growing in OGYE plates by amplification and sequencing the flanking internal transcribed spacer (ITS) region using primers ITS4 and ITS5 (White et al., 1990). With this approach, presence of *Candida parapsilosis*, *Pichia kudriavzevii* and *Saccharomyces cerevisiae* were revealed in these samples (data not showed).

For the identification of LAB recovered in MRS and M17L plates, between 10 and 15 representative colonies of each SC samples were picked, purified by sub-culturing and typed by the Repetitive extragenic palindromic (REP)-PCR method, using primer BoxA2R (data not showed). On the basis on this approach 20 isolates were selected for PCR amplification and sequencing. The identification was initially obtained using the hypervariable regions V1-V2 from the 16S rRNA gene and, in some particular cases, further confirmation with the *sodA* gene was needed. Only four different species were identified, including *Lactobacillus plantarum* (99 % identity) in P1, *Enterococcus faecium* (100 % identity) in P2 and P3, *Lactococcus lactis* (99 % identity) in P1 and P3 and *Streptococcus infantarius* in P2 and P3. This last species was impossible to be distinguished from *Streptococcus lutetiensis* and *Streptococcus equinus* by sequencing only the 16S rDNA. The use of *sodA* gene allowed undoubtedly the allocation of these isolates to the *S. infantarius* species (100 % identity). The partial sequences obtained were deposited in the GenBank nucleotide sequence database under accession numbers KX904823 to KX904826. The *Streptococcus bovis*/*S. equinus* complex (SBSEC) is a group of human and animal derived streptococci that include commensals and food-associated types, as well as opportunistic pathogens. This complex consists of the species *S. infantarius, S. lutetiensis, Streptococcus gallolyticus, Streptococcus alactolyticus* and *S. equinus*, the latter still frequently referred to as *S. bovis* (Jans et al.,

2015). Meanwhile some members of this complex (*S. bovis*) have been linked with infectious diseases, such as bacteremia and endocarditis (Herrera et al., 2009) or even colonic cancer development (*S. gallolyticus*) (Abdulamir et al., 2011), *S. infantarius* has been associated with traditional dairy and plant-based food fermentations. In fact, *S. infantarius* is the predominant LAB in traditionally fermented milk products from Africa, reaching levels of 10^8 CFU /mL with a wide distribution and presumably long history of use (Jans et al., 2013; Jans et al., 2017). In Asia, potential *S. infantarius* branch members were detected as predominant species in fermented "Dahi" milk (Rashid et al., 2009). To our knowledge, this is the first time that *S. infantarius* isolates from artisanal milk fermentations in South American are reported. Although less known than other LAB members in dairy foods, the ability of *S. infantarius* to ferment milk is supported by the presence of the *gal-lac* operon directed to lactose metabolism which allows the adaptation of this bacterium to the milk environment (Jans et al., 2013; Jans et al., 2017).

Respect to the other cultivatable members identified in the SC, *L. lactis* is described as dominant in other natural fermented milks (Alegría et al., 2010; Gonfa et al., 2001). On the other hand, *Enterococcus* species may be present in relevant numbers in fermented milk products and have been reported in the production of "Kumis" elaborated in Colombia (Chaves-López et al., 2011). In dairy products manufactured from raw milk it is also common to identify species of mesophilic lactobacilli, such as *L. plantarum* and *Lactobacillus casei*. In warm geographical areas other lactobacilli, such as *Lactobacillus helveticus*, *Lactobacillus fermentum* or *Lactobacillus acidophilus*, are also habitually detected (Alegría et al., 2010; Gonfa et al. 2001). None of these last species were isolated in the SC samples analysed, although its presence was detected by DNA molecular techniques as will be described below.

3.2. Culture-independent analysis

3.2.1. PCR-DGGE analysis.

The bacterial community structure and dominant populations in the SC samples was further investigated by culture-independent methods. Initially, PCR-DGGE with universal primers for the variable region V3 of the 16S rRNA gene was the approach used. The diversity observed with this technique was rather simple with few predominant bands and higher number of faint bands (Figure 3). Fingerprints of the bacterial communities were different among manufacturers but with minimal sample-to-sample variations in P1 and P3. Clearer differences were detected in the DGGE patterns obtained from the two SC batches of producer P2. Identification of bands was performed by matching with the species present in the DGGE ladder and further confirmed by band excision from the gels, amplification and sequencing. This double strategy allowed allocating most of the more intense bands to four different groups of microorganisms. After sequence comparison against the GenBank and RDP databases, identification at species level (≥ 99% similarity) was obtained only for *Streptococcus salivarius* (band 2, Figure 3) presents in SC samples from P1 and P2, and *L. lactis* (band 4, Figure 3) revealed in samples form the three manufacturers, but with a more intense band in sample P2A. Some bands present in the two batches from P3, and with a lower intensity in P2A, matched in the databases with sequences of the species *S. infantarius*/*S. equinus*/*S. lutetiensis* (band 3, Figure 3). As mentioned above, due to the small size in length of the amplicon sequenced (~130 bp) and the similarities displayed in the 16S rDNA sequences by the SBSEC *Streptococcus* group (Hinse et al., 2011), it was not possible to distinguish among these species. However, the migratory distance in the gels, in comparison with the position of *S. infantarius* in the ladder (band c, Figure 3), and the isolation of this species by culturing, suggests that band 3 corresponds with

this particular species. The same problem, to reach optimal species level identification, happened for bands present in SC samples from P1 and P3, whose sequences matches with closest relative of *Lactobacillus crispatus*/*Lb. helveticus*/*Lb. acidophilus* (band 1, Figure 3).

3.2.2. High-throughput 16S rDNA sequencing analysis.

To obtain a more complete picture of the bacterial communities present in the Colombian fermented milk SC, DNA samples were subjected to HTS of PCR-generated 16S rRNA gene amplicons. A total of 469,307 raw reads were obtained by paired-end Illumina sequence technology. Of these, a mean of 76,853 high-quality partial 16S rDNA sequences of the variable region V3 per sample were retrieved (Table S2). All raw data obtained with this technique were deposited in the Sequence Read Archive (SRA) of the NCBI (http://www.ncbi.nlm.nih.gov) under accession numbers SRR4342055 to SRR4342060.

The bacterial diversity estimated by the Shannon index (H) at 97 % similarity level revealed similar values for the SC samples analysed (H ~2) except for sample P2B which showed a reduced diversity (H =0.92) in comparison with the others.

Sequences were classified, using QIIME and SILVA database, to the lowest possible taxonomic rank, i.e. genus level. Sequences were assigned to four different phyla, including *Firmicutes* (which retrieved more than 90 % of sequences), *Proteobacteria* (with a percentage of assigned sequences ~8 % and presenting differences among samples) and, in minor proportion, *Actinobacteria* and *Bacteroidetes* with less than 1% of assigned sequences in all samples. Bacterial composition of the different SC samples studied at the family levels is represented in Table 1. Two different families, *Streptococcaceae* and *Lactobacillaceae*, were predominantly found. The first one was most abundant in non-inoculated SC samples, those from producers of

350 "Planeta Rica" region (P2 and P3), while members of *Lactobacillaceae* clearly
351 dominate in samples from P1 of "Caucasia" region who used the back-slopping
352 technique. At lower relative abundance, sequences belonging to other taxa of the
353 *Lactobacillales* order, such as *Leuconostocaceae* (only in samples from P2) and
354 *Enterococcaceae*, were detected. *Staphylococcaceae* and even *Bifidobacteriaceae* were
355 detected in some samples with relative abundances lower than 1%. Similarly, three
356 families of Gram-negative bacteria of the phylum *Proteobacteria* appeared at low
357 percentages; these comprised *Acetobacteraceae* family (in samples from P3 producer),
358 *Aeromonadaceae* (in P1) and *Enterobacteriaceae* (present in all samples).
359 Enterobacteria were also reported previously in SC fermentation (Cueto et al., 2007)
360 and sequences belonging to different genera of *Proteobacteria* (such as *Acinetobacter*
361 and *Pseudomonas*) were observed, by HTS techniques, to be predominant in milk in
362 some cases resulting from environmental contamination (Mayo et al., 2014). The
363 taxonomic composition at genus level is displayed in Figure 4, where clear differences
364 in the bacterial profiles between the two batches from P2 were observed, as previously
365 revealed the PCR-DGGE technique. The genus *Lactobacillus* (85 % of sequences
366 assigned), and to minor extend *Aeromonas* (10 %), predominated in the samples from
367 P1; whereas, *Streptococcus* was the dominant genus in samples from P2 and P3, with
368 higher abundances in the producer that use the "Totumo" seed container. In samples
369 from P3, *Lactobacillus*, together with *Streptococcus*, predominated in the SC, followed
370 by members of *Lactococcus* and *Acetobacter* genera. The presence of the latter,
371 belonging to the acetic bacteria group, was not disclosed by the other techniques used in
372 this study, however, has been described in other artisanal fermented and dairy products
373 such as kefir (Leite et al., 2012) contributing to the organoleptic properties of the final
374 product. Sequences of other bacterial members that are not usually associated (or

375 detected by conventional techniques) with artisanal fermented dairy products, such as
376 *Bifidobacterium*, were also revealed (Figure 4). The occurrence of this genus was also
377 detected in some traditional cheeses and kefir grains after the application of 16S rDNA
378 HTS techniques (Alegría et al., 2012; Leite et al., 2012).

379 In an attempt to assign the most abundant OTUs at the species level, manual
380 sequence comparisons of representative reads were performed against the GenBank
381 database. This approach allowed confirming i) the presence of *S. salivarius* at high
382 abundance in samples from P2, ii) the existence of closest relatives to the *Lb.*
383 *acidophilus* group in both P1 and P3 elaborations and iii) OTUs matching SBSEC
384 members predominated in P3 (data not shown).

385 Finally, the Unifrac method was used to compare the bacterial communities
386 among samples based on their phylogenetic relationship. Principal coordinates analysis
387 (PCoA) confirmed the similarities between batches of the same producer, and the
388 differences among the bacterial composition in SC elaborations depending on the
389 producer (Figure 5). The samples from P1 of "Caucasia", who uses a re-inoculum for
390 the fermentation, were located at a considerable distance apart from the two producers
391 of the "Planeta Rica" region. The bacterial communities from the samples elaborated
392 from these two producers, even if they used different recipients and produced the SC
393 through natural fermentation, were more closely related. The use of back-slopping
394 technique in the elaboration of SC in "Caucasia" region possibly contributes to
395 constraint the bacterial growth and determines the final populations of the product
396 which, as observed by culture-independent PCR-based techniques, was enriched in
397 sequences representatives of the *Lb. acidophilus* complex. Members of this group were
398 not recovered from the SC samples, at least with the culture media and growing
399 conditions used in this work. Additionally, regarding the detection in SC of *S.*

salivarius, which was neither recovered by culturing in this study, its presence has been previously reported in raw camel milk and fermented dairy products from Africa (Jans et al. 2017). In fact, members of the genus *Streptococcus* (including *S. thermophilus* and *S. salivarius*) are regularly present in dairy products worldwide. Although, *S. salivarius* is a human commensal closely related to *S. themophilus*, only the species *S. thermophilus* is currently approved for use in dairy fermentations by the European Food Safety Authority (EFSA, 2016). Standardization in the elaboration of SC could be reached with the use of pasteurized milk and controlled fermentations by means of adapted strains specifically isolated from this artisanal product. Our study revealed that the characteristic microbiota of SC fermentation is a result of a combination of factors, including the warm temperature conditions in the Caribbean region, as well as the traditional containers used (calabashes as Totumo") and handling techniques.

4. Conclusions

Altogether our data suggest that the bacterial ecosystem in the Colombian fermented sour-cream "Suero Costeño" depends not only of the origin and quality of the milk, but also on the practices of the manufacturers, such as the type and material of the recipients regularly used for the fermentation of the milk. Although limited by the small number of samples/manufacturers analysed, it seems that re-inoculation of the milk with a portion of the previous batch is the main driver determining the fermentation and bacterial composition of "Suero Costeño". Major contamination and lack of reproducibility between batches, such as those encountered with the use of "Totumo" as vessel for fermentation, difficult the search and isolation of adapted representative strains typical from this product. Even so, autochthonous *Streptococcus infantarius* isolates (previously associated to warm climates) from this traditional milk fermentation were obtained.

Acknowledgments

This study was supported by Spanish Ministry of Economy and Competitiveness and the European Union FEDER funds (project reference AGL2015-64901-R). The cooperation project COOPA20057, between IPLA-CSIC (Spain) and Antioquia University (Colombia), funded by CSIC, allowed several short stays of K.E. Motato at IPLA-CSIC. S. Deldado is supported by a Postdoctoral research contract associated to a project from the Spanish "Plan Estatal de I+D+ i" research plan (project reference BIO2014-55019-JIN).

FIGURE CAPTIONS

Figure 1.- Diagram outline of "Suero Costeño" elaboration.

Overview of sampling, producers and regions for the manufacture of "Suero Costeño" Colombian fermented milk analysed in this study.

Figure 2.- Mean microbial counts of two batches from "Suero Costeño" from different producers.

P1, P2 and P3 represent the three producers. Counts are expressed as Log_{10} of colony forming units per millilitre (CFU /mL). Abbreviations for media are as follow: PCA for mesophilic bacteria, MRS for lactobacilli, M17L for lactococci, BP for staphylococci and OGYE for yeast and moulds. Data were analysed by means of one-way ANOVA and the SNK (Student-Newman-Keuls) mean comparison test. Values that in the same medium do not share a common letter are statistically different ($p<0.05$).

Figure 3.- Denaturing gradient gel electrophoresis (DGGE) showing the bacterial diversity present in "Suero Costeño" samples.

M: DGGE ladder comprises the species: a) *Lactobacillus plantarum*, b) *Enterococcus faecium*, c) *Streptococcus infantarius* and d) *Lactococcus lactis*. Bands identified by gel excision and PCR sequencing are indicated by numbers; 1) *Lactobacillus crispatus*/*Lactobacillus helveticus*/*Lactobacillus acidophilus*, 2) *Streptococcus salivarius*, 3) *S. infantarius*/*Streptococcus equinus/ Streptococcus lutetiensis* and 4) *Lactococcus lactis*. The two different batches of the three producers (P1, P2 and P3) are referred as A and B.

Figure 4.- **Bacterial composition at the genus level of the "Suero Costeño" samples**.

Composition is represented as the relative abundance. Only genera contributing >0.5 % of the total abundance in at least one sample are showed. P1, P2 and P3 refer to the three different producers, meanwhile letters A and B refer to the two batches analysed.

Figure 5.- **Clustering of operational taxonomic units (OTUs) at a 97 % similarity level of "Suero Costeño" samples**.

Principal Coordinate Analysis (PCoA) plot; percentages shown in the axes represent the proportion of dissimilarities. P1, P2 and P3 refer to the three different producers, meanwhile letters A and B refer to the two batches analysed

Table 1.- Family composition (in relative abundance) of different batches and producers of "Suero Costeño" fermented milk as revealed by high-throughput 16S rDNA sequencing analysis (only families with a percentage > 0.3% are showed). Shadow numbers represent the relative abundances higher than 1%.

Families	**Relative abundance (%)**					
	P1A[1]	**P1B[1]**	**P2A[1]**	**P2B[1]**	**P3A[1]**	**P3B[1]**
Streptococcaceae	0.09	1.40	87.97	92.49	65.63	50.88
Lactobacillaceae	89.91	83.51	0.30	3.80	24.99	41.67
Leuconostocaceae	0.00	0.00	2.12	2.86	0.00	0.00
Enterococcaceae	0.03	0.09	0.35	0.06	0.29	0.21
Staphylococcaceae	0.02	0.06	0.70	0.01	0.03	0.05
Bifidobacteriaceae	0.05	0.04	0.60	0.00	0.00	0.00
Acetobacteraceae	0.00	0.01	0.00	0.00	5.12	5.97
Aeromonadaceae	8.64	11.44	0.71	0.30	0.00	0.00
Enterobacteriaceae	0.50	2.58	6.89	0.21	2.80	0.11

[1]: P1A (producer 1, batch A); P1B (producer 1, batch B); P2A (producer 2, batch A); P2B (producer 2, batch B); P3A (producer 3, batch A); P3B (producer 3, batch B)

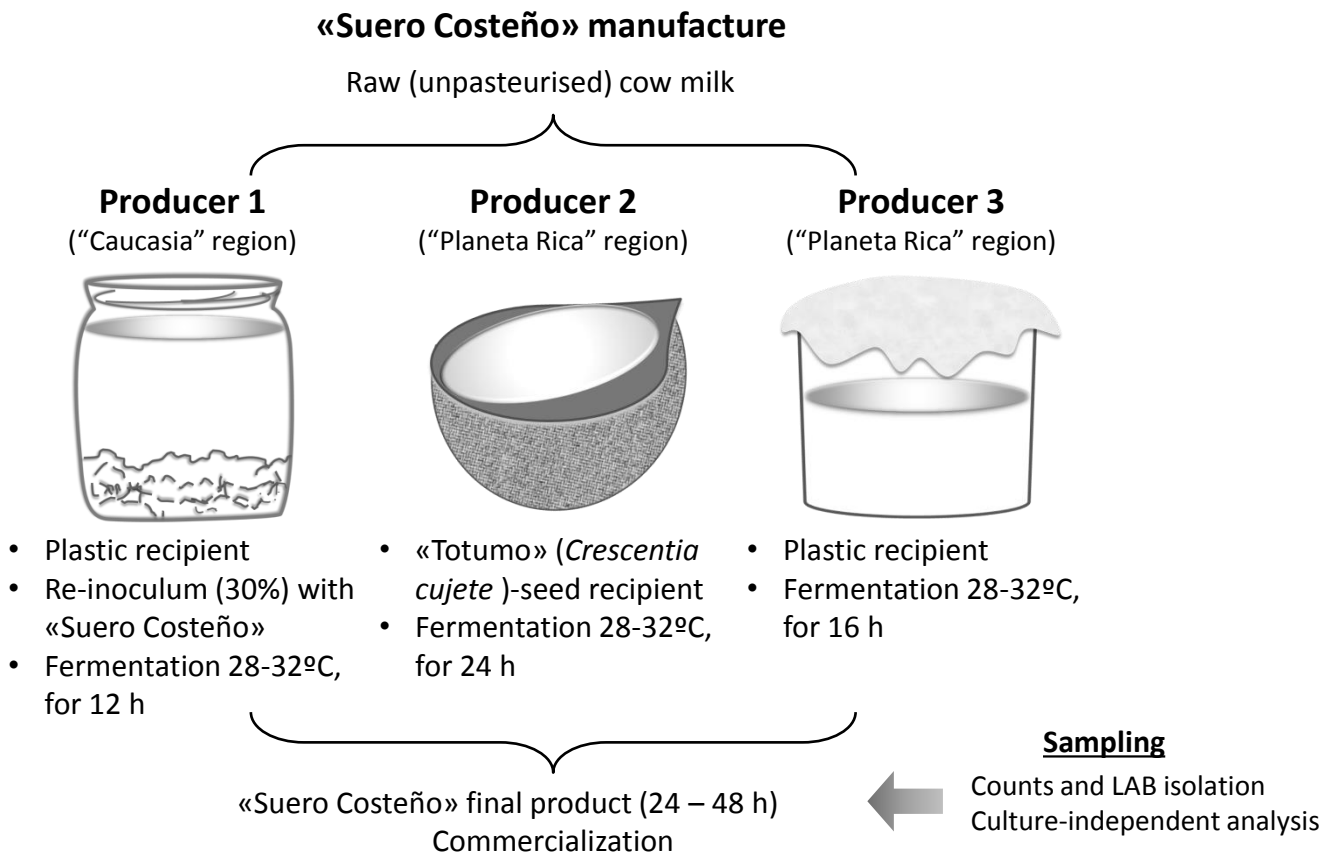
«Suero Costeño» manufacture
Raw (unpasteurised) cow milk
Producer 1
("Caucasia" region)
Producer 2
("Planeta Rica" region)
Producer 3
("Planeta Rica" region)
• Plastic recipient
• Re-inoculum (30%) with «Suero Costeño»
• Fermentation 28-32ºC, for 12 h
• «Totumo» (Crescentia cujete)-seed recipient
• Fermentation 28-32ºC, for 24 h
• Plastic recipient
• Fermentation 28-32ºC, for 16 h
«Suero Costeño» final product (24 – 48 h)
Commercialization
Sampling
Counts and LAB isolation
Culture-independent analysis

Figure 1

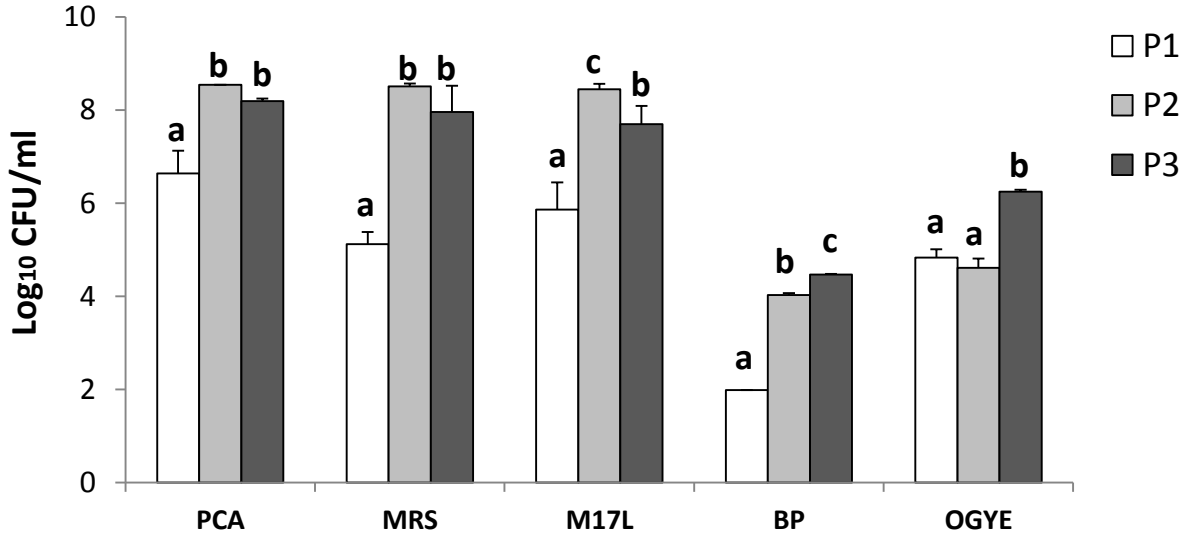
10
8
6
4
2
0
Log10 CFU/ml
a
b
b
a
b
b
a
c
b
a
b
c
a
a
b
PCA
MRS
M17L
BP
OGYE
P1
P2
P3

Figure 2

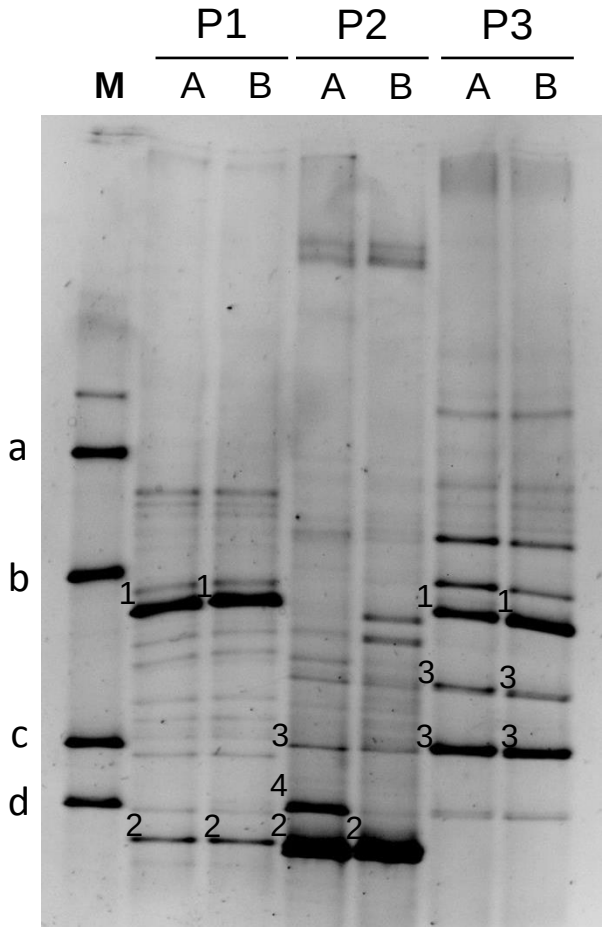

Figure 3

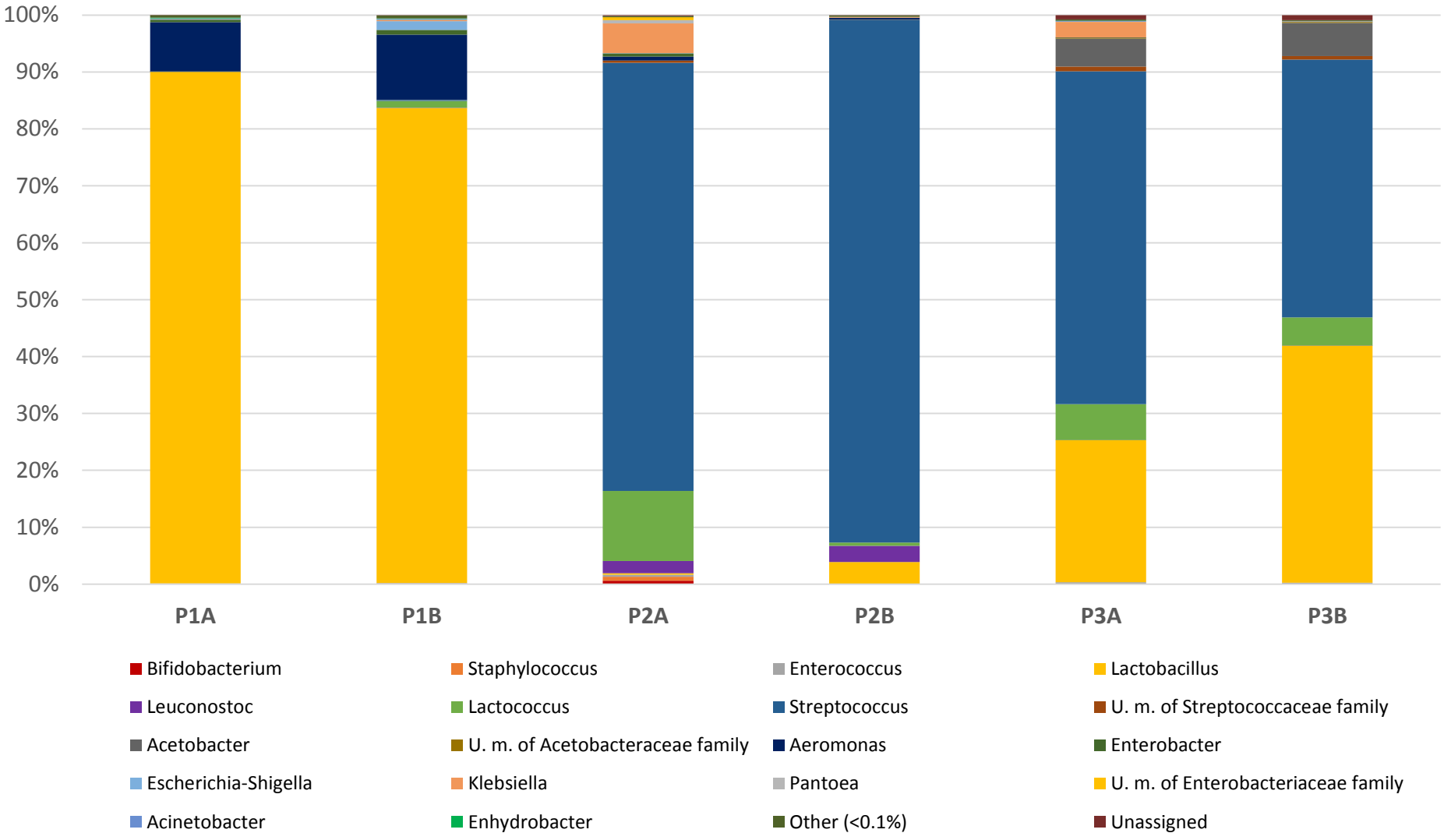
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
P1A
P1B
P2A
P2B
P3A
P3B
Bifidobacterium
Staphylococcus
Enterococcus
Lactobacillus
Leuconostoc
Lactococcus
Streptococcus
U. m. of Streptococcaceae family
Acetobacter
U. m. of Acetobacteraceae family
Aeromonas
Enterobacter
Escherichia-Shigella
Klebsiella
Pantoea
U. m. of Enterobacteriaceae family
Acinetobacter
Enhydrobacter
Other (<0.1%)
Unassigned

Figure 4

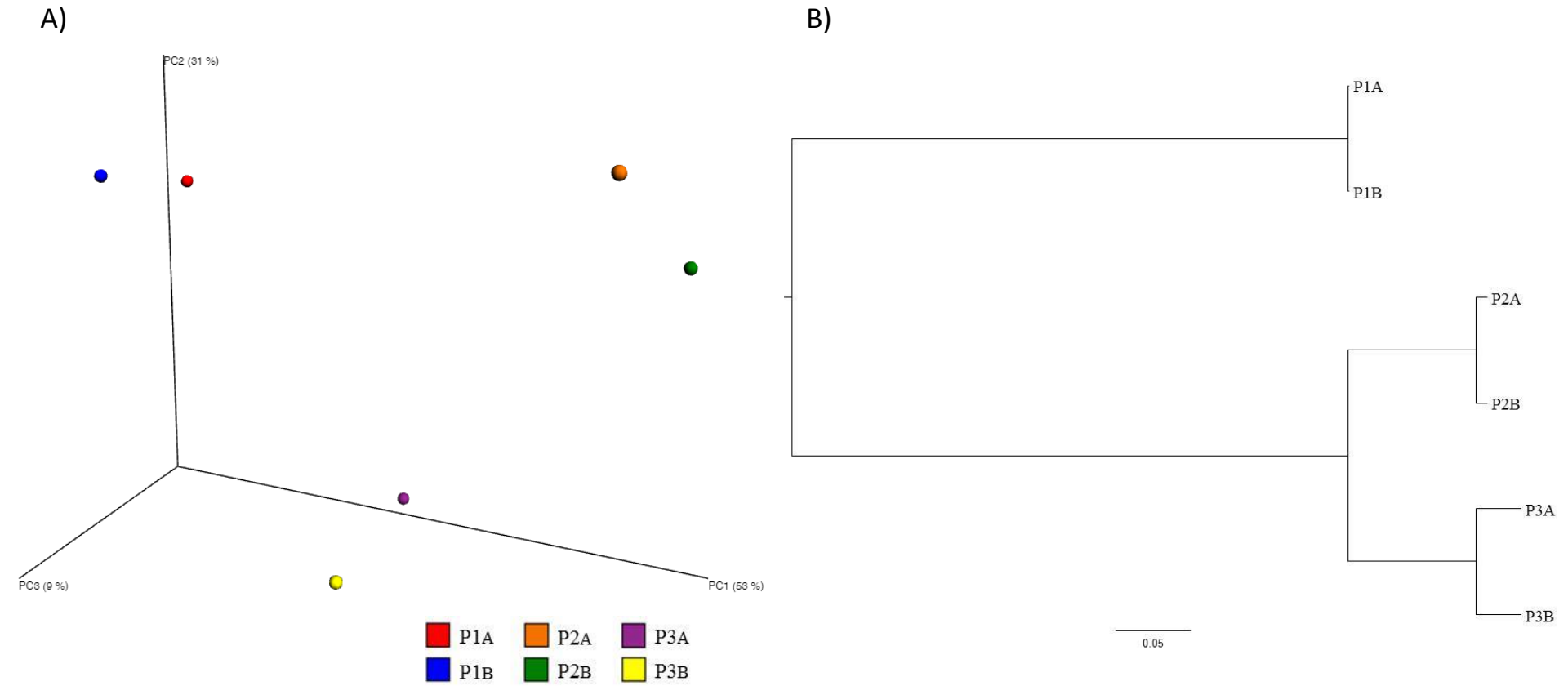
A)
PC2 (31 %)
PC3 (9 %)
PC1 (53 %)
P1A
P2A
P3A
P1B
P2B
P3B
B)
P1A
P1B
P2A
P2B
P3A
P3B
0.05

Figure 5

Fig. S1. Representation of the route followed from the laboratory located in Medellin to the two Colombian regions ("Caucasia" in Antioquia and "Planeta Rica" in Córdoba) where "Suero Costeño" was sampled. The illustration includes a small map of Colombia in which the departments of Antioquia and Córdoba are both marked in brown colour. The departments of Sucre and Bolívar are situated towards the northeast in the Caribbean coast of Colombia.

Table S1.

	MPN		Presence/absence
Producer	**Coliforms**	***E. coli***	***Listeria* spp.**
P1	23/23	2/2	0/1
P2	1101/1101	1101/1101	1/1
P3	1101/1101	1101/1101	1/1

Table S2.

Band	Sequence length	Closed relatives	% Sequence similarity	Sequence	E-value
1	130	*Lactobacillus acidophilus* *Lactobacillus crispatus* *Lactobacillus helveticus*	100%	GAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTGGTGAAGAAGGATA GAGGTAGTAACTGGCCTTTATTTGACGGTAATCAACCAGAAAGTCACGGCTAAC TACGTGCCAGCAGCCGCGGTAATA	3e-60
2	120	*Streptococcus salivarius*	100%	GTTTTCGGATCGTAAAGCTCTGTTGTAAGTCAAGAA CGAGTGTGAGAGTGGAAAGTTCACACTGTGACGGTAGCTTACCAG AAAGGGACGGCTAACTACGTGCCAGCAGCCGCGGTAATA	1e-54
3	129	*Streptococcus infantarius* *Streptococcus alactolyticus*	100%	TTTCCACTCTCACACACATTCTTCTCTTACAACAGAGCTTTACGATCCGAAAA CCTTCTTCACTCACGCGGCGTTGCTCGGTCAGGGTTGCCC CCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAA	4e-59
4	133	*Lactococcus lactis*	100%	CGTGAGTGAAGAAGGTTTTCGGATCGTAAAACTCTGTTGGTAGAGAA GAACGTTGGTGAGAGTGGAAAGCTCATCATGTGACGGTAACTACCCAGAAA GGGACGGCTAACTACGTGCCAGCAGCCGCGGTAAT	6e-58

Table S3.

Sample	Number of reads	Number of reads removed because of:						Final reads number	Average reads length
		Outside bounds (100-400)	Ambiguous bases	Mean quality <25	Homopolymer runs >7bp	Primer mismatch >1	Reverse primers not found		
P1A	71,722	0	0	1	1	1,047	11	70,662	193.8
P1B	78,422	0	0	0	0	1,231	32	77,159	193.8
P2A	95,975	0	0	0	0	1,982	47	93,946	194.7
P2B	74,005	0	0	0	0	1,241	32	72,732	194.9
P3A	74,484	0	0	0	0	1,216	14	73,254	193.3
P3B	74,699	0	0	0	2	1,315	18	73,364	192.9