Highlights:

- A new high throughput sequencing approach was developed to target *Bacillus cereus* group
- Milk contained a complex *B. cereus* group population, which is dominated by psychrotolerant strains divided in several sequence groups.
- Dynamics of the *B. cereus* group in milk was influenced by storage temperature, sampling period and dairy.

Title:

Application of a novel amplicon-based sequencing approach reveals the diversity of the *Bacillus cereus* group in stored raw and pasteurized milk

Author names and affiliations:

Davide Porcellato[a], Marina Aspholm[b], Siv Borghild Skeie[a], Hilde Mellegård[b]

[a] Faculty of Chemistry, Biotechnology and Food Science, the Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

[b] Department of Food Safety and Infection Biology, Faculty of Veterinary Medicine, the Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway

Key words: *Bacillus cereus*, sequencing, consumption milk

Corresponding author:

Davide Porcellato, P.b. 5003, 1432 Ås, Norway

Tel.: +4764965143 - Fax: +4764965901

E-mail address: davide.porcellato@nmbu.no

Abstract

Members of the *Bacillus cereus sensu lato* (*B. cereus* group) are spore-forming organisms commonly associated with spoilage of milk and dairy products. Previous studies have shown, by using 16S marker gene sequencing, that the genus *Bacillus* is part of the core microbiota of raw bovine milk and that some members of this genus is able to grow during sub-optimal storage (8 °C) of pasteurized consumption milk. Here, the composition of this genus in pasteurized consumption milk samples, collected from two dairies, over a one-year period and stored at 4 or 8 °C up to the end of shelf life is uncovered. Our results show that the *B. cereus* group is the dominant *Bacillus* group in stored consumption milk. By applying a new marker gene sequencing approach, several dominating phylogenetic clusters were identified within the *B. cereus* group populations from the milk samples. There was a higher phylogenetic diversity among bacteria from milk stored at 8 °C compared to milk stored at 4 °C. Sampling period and the dairy the samples were collected from, also significantly influenced the diversity, which shows that the *B. cereus* group population in consumption milk is heterogeneous and subjected to temporal and spatial changes. The new approach applied in this study will facilitate the identification of isolates within the *B. cereus* group, of which some are potential spoilage bacteria and pathogenic contaminants of milk and dairy products.

1 INTRODUCTION

Bacillus cereus sensu lato (informally termed the *Bacillus cereus* group) is a sub-group within the genus *Bacillus*, which currently contains the eight species *B. cereus* (*sensu stricto)*, *Bacillus thuringiensis*, *Bacillus anthracis*, *Bacillus weihenstephanensis*, *Bacillus mycoides*, *Bacillus pseudomycoides*, *Bacillus cytotoxicus* and *Bacillus toyonensis* (Patino-Navarrete and Sanchis, 2016). Recently, other species have also been suggested as members of this group (Jung et al., 2011; Liu et al., 2014; Miller et al., 2016). Members of this group show a great ecological diversity. Some species are harmful to humans, for example *B. anthracis*, the etiologic agent of anthrax, and *B. cereus*, which is involved in food poisoning (Granum and Lindbäck, 2013), while *B. thuringiensis* produces insecticidal toxins and is used commercially for crop protection (Aronson and Shai, 2001). *Bacillus cereus* group species also vary in their ability to grow at different temperatures, with the psychrotolerant *B. weihenstephanensis* and the thermotolerant *B. cytotoxicus* as extremes within this group (Guinebretiere et al., 2013; Lechner et al., 1998).

Bacillus spp. are frequently found in soil environments and on plants, and they easily spread to bovine raw milk. Members of the genus *Bacillus* have been defined as part of the core microbiota of raw milk (Christiansson et al., 1999; Kable et al., 2016; Magnusson et al., 2007). Contamination of milk with these bacteria may also occur during transport and at the dairy plant through the processing facilities (Eneroth et al., 2001; Flint et al., 1997; Svensson et al., 2000). The ability of *Bacillus* spp.to form endospores renders them a challenge to the dairy industry. Spores are able to survive heat treatment regimes commonly applied to consumption milk, such as high temperature-short time (HTST) pasteurization. The spores may then germinate into proliferative bacteria if environmental conditions allow it (Novak et al., 2005; Setlow,

2003). The species *B. weihenstephanensis* and psychrotolerant strains of *B. cereus* are frequently isolated from milk products and are able to grow and cause spoilage when the temperature is favorable (Bartoszewicz et al., 2008). For example, *B. weihenstephanensis* is able to grow at 7 °C or below, and can reach high numbers in milk products during storage (Francis et al., 1998). Since different *Bacillus* spp. are present in HTST pasteurized dairy consumption milk products, and various species might impact product quality and safety differently, detailed knowledge on the *Bacillus* spp. population in this products will benefit the food quality and safety assessment work of the dairy industry.

Recent progresses in sequencing technologies and bioinformatics tools have enabled characterization of the total microbiota directly from food matrices, including milk (Ercolini, 2013; Kable et al., 2016). However, in the resulting data sets the bacterial composition is often described at genus or family level. For *Bacillus* spp., some members of the *B. cereus* group have highly conserved 16S and 23S rRNA sequences. Therefore, accurate species characterization based on these sequences is challenging (Liu et al., 2015). For example, amplicon-based high throughput sequencing (HTS) approaches can only assign *Bacillus* sequences to the genus level. Therefore, it is of interest for both research and industrial purposes to develop culture-independent HTS methods for identification of subpopulations of the *Bacillus cereus* group directly from food products.

Here, we developed a novel culture-independent approach to follow the dynamics and composition of the *B. cereus* group population in samples of consumption milk. This approach achieved higher resolution compared to 16S rRNA sequencing and was applied to evaluate the composition of the *B. cereus* group population in pasteurized milk stored at different temperatures.

2 MATERIALS AND METHODS

2.1 Milk Samples and DNA extraction

A total of 184 bovine milk samples, which were part of a larger study on milk microbiota (Porcellato et al., 2017), were included in the present work. The milk samples were collected monthly (with the exception of one month) from two dairy plants in Norway over a period of 13 months. The milk samples were obtained from two dairy plants with different production volumes (A and B, respectively), and details on the sampling procedures are described previously (Porcellato et al., 2017). Samples from every sample month included 100 mL of raw milk from the dairy silo tanks and six cartons (1 L) of homogenized and pasteurized (72 °C for at least 15 sec) full fat milk. The milk samples were stored cold (maximum temperature of 4 °C) before and during transport to the laboratory, and were kept at 4 °C (3 milk cartons) or 8 °C (3 milk cartons and 100 mL samples of raw milk) until the end of carton shelf life (as indicated on the carton; 13-14 days after production). At the end of storage, three replicates from each milk carton and two replicates of each raw milk sample were analysed. Storage of raw milk samples at 8 °C for the same period of time as the milk cartons has not to our knowledge been previously reported. Bacterial pellet and DNA extraction was performed as described before (Porcellato et al., 2016a). The DNA was stored at -20 °C until further analysis.

2.2 Analysis of *Bacillus* spp. 16S rRNA gene sequences

To investigate the most abundant *Bacillus* species within the 16S sequence library from a previous study (Porcellato et al., 2017), a new in silico database was constructed. The database was constructed using the complete 16S rRNA gene

sequences (n = 28605) collected from the ribosomal database project (RDP) database (parameters: Genus = *Bacillus*, Strain = Both, Source = Isolate, Size > 1200, and Quality = Good; Cole *et al* 2014). A custom database was constructed by performing an *in silico* PCR, using the same primers against the *Bacillus* 16S rRNA (allowing 2 mismatches) as used for the 16S library preparation. The obtained PCR products were filtered by removing sequences containing unknown bases ("N") and by removing sequences which were not assigned to a *Bacillus* species. The *in silico* PCR products were then clustered at 99% sequence identity using the Usearch algorithm (Edgar, 2010) and the OTU-based approaches as described previously for the 16S rRNA library. For each *Bacillus* OTU detected at 99% sequence identity, the species names were used to assign the different species that belonged to the OTU. All sequences, which were previously assigned to the genus *Bacillus* in the previous study (Porcellato et al., 2017), were extracted and a new OTU table was constructed using the Usearch algorithm with 99% identity. Taxonomic assignment of new *Bacillus* OTUs was performed by searching each OTU representative sequence against the previously created database.

2.3 Primer design

To assign the *Bacillus cereus* group members to a sub-species level, marker genes were chosen from a previously described MLST scheme (Tourasse et al., 2006). Selection of the three most discriminating genes was performed after alignment of all type sequences for the 7 genes used in the MLST scheme (available online http://mlstoslo.uio.no/). The sequences were aligned using MAFFT and the three genes with the lowest % of identical sites were chosen. The three selected genes were *panC*, *pycA* and *glpT*. After alignment of each gene sequence, new primers were

designed to amplify PCR products of lengths between 300 and 400 bp. This length was chosen in order to obtain a good overlap of the sequences after merging the paired-end sequences acquired from the Miseq Illumina system (Illumina, San Diego, CA, USA). Three degenerate bases and 1 mismatch were allowed on both forward and reverse primers. Regions with high similarity between all the sequences were visualized using the software Geneious v 7.0 and primers were designed using the primer3 software (Untergasser et al., 2012). The selected primers were searched against the "nr" nucleotide database using the blast algorithm (Camacho et al., 2009). Primers were validated for specificity using isolates of different *Bacillus* species and other species from the in-house laboratory collection at the Norwegian University of Life Sciences (Table S1).

2.4 Illumina sequencing

Each marker gene was amplified using a Lightcycler 480 system (Roche). Each PCR reaction was run in a volume of 20 µl using 1X of Q5 reaction buffer (New England Biolabs inc., Pswich, MA, USA), 10 mM dNTPs, 2 µM of each of the forward and reverse primer, 1X of Evagreen dye (Biotium, Fremont, CA, USA), 0.02 U/µl of Q5 high-fidelity DNA polymerase (New England Biolabs inc.,) and 2 µl of DNA. The PCR conditions were as follows: one cycle of initial denaturation at 98 °C for 30 sec, followed by 35 cycles of 98 °C for 10 sec, 52 - 54 °C (annealing temperature according to Table 1) for 20 sec and 72 °C for 20 sec. The final extension was performed for 2 min at 72 °C. Five µl of each PCR product was cleaned using Illustra ExoProStar S (GE Healthcare Life Sciences, Oslo, Norway) according to the manufacturer´s protocol. The cleaned PCR product was used as template for a second indexing PCR using customized primers with unique 8 bp barcodes on the forward

and reverse primer. Each sample obtained the same combination of barcodes for all three genes. The PCR mix was similar to the first PCR mix except for the volume of template added (5 µl). The PCR reaction was carried out in a volume of 20 µl and the amplification conditions were as follows: Initial denaturation at 98 °C for 30 sec followed by 10 cycles of 98 °C for 10 sec, 55 °C for 20 sec and 72 °C for 20 sec. The final extension was performed at 72 °C for 2 min. Library normalization was performed using a SequalPrep™ Normalization plate (Thermo Fischer Scientific, Oslo, Norway) and quantified using a PerfeCTa NGS quantification kit (Quanta Biosciences, Beverly, MA, USA). The samples were diluted to a final concentration of 4 nM. Sequencing of the marker genes *panC*, *pycA* and *glpT* from the total DNA was performed on a Miseq Illumina platform (Illumina) using a 300 bp paired-end sequencing kit. The library was diluted to 8 pM as described by the Illumina sequencing protocol (Illumina) before sequencing.

2.5 Bioinformatics analysis

After sequencing, sequences were divided for each sample by the Miseq Reporter (Illumina). The obtained files were quality filtered (q30) using trimmomatic (Bolger et al., 2014) and merged using Qiime 1.9.0 (Caporaso et al., 2010). Sequences shorter than 250 bp were discarded. The remaining sequences were divided in three files (one for each gene) after alignment to a custom made database using Usearch v8 (Edgar, 2010). The databases consisted of a FASTA file with all the type sequences of the *panC, pyCa* and *glpT* genes obtained from the online database (http://mlstoslo.uio.no/). Each set of sequences were subjected to *in silico* PCR using a custom made R script allowing one mismatch for each of the two primers. Only sequences that obtained a correct *in silico* amplification, were kept for further

analysis. The sequences were then clustered using the Usearch OTU picking algorithm (Edgar, 2010) with 99 % sequence identity. The pipeline included chimera and singletons removal steps. OTUs with a relative abundance below 0.01 % were discarded. All sequences from the 16S library and the *Bacillus* marker genes are available after request.

2.6 Statistical analysis

The three OTU tables (one for each gene) were normalized using the metagenomeSeq package. Alpha indexes (richness and diversity) were calculated for each OTU table and compared using Student's t-test with 1,000 Monte Carlo simulations. Beta diversity was analyzed using the principal coordinate analysis (PCoA) of the Unifrac distance matrix (Lozupone and Knight, 2005). Multivariate dispersion between groups of samples was calculated with the R package Vegan and ANOVA test (Anderson et al., 2006; Oksanen et al., 2017). Permutational multivariate analysis of variance between groups was also calculated with the R package Vegan using the Unifrac distance matrix and the function "Adonis" (Oksanen et al., 2017). The centroid sequence of the 19 most abundant *panC* sequences were searched against with the online tool (https://www.tools.symprevius.org/bcereus/english.php) to assign the sequence group to a specific *Bacillus* group as described by Guinebretiere et al (2008). The same sequences were aligned using MAFFT (Katoh and Standley, 2013) and the alignment was used to construct a Neighbor-Joining phylogenetic tree (with 999 boostraping replicates) with the R package *phangorn* (Schliep, 2011).

3 Results

3.1 16s rRNA library evaluation

To explore the *Bacillus* population during storage of consumption milk, we used 1) data from a previously published 16S rRNA library (Porcellato et al., 2017) and 2) a new marker gene sequencing approach. First, we extracted all reads from the 16S rRNA library assigned to the genus *Bacillus* (n= 6326686) with a pairwise identity of 97%. These reads were then clustered in OTUs based on 99% identity (to account for sequencing errors). From 186 samples of milk, six *Bacillus* OTUs with a relative abundance higher than 0.001 % were obtained. The centroid sequence for each OTU cluster was used to identify the *Bacillus* group, which the sequences belong to. This was done by searching the sequences against the in-house *Bacillus* 16S rRNA database. One sequence group (OTU 1), which matched species of the *Bacillus cereus* group (Table S2), represented over 99.6 % of the total reads and was the most abundant OTU in all milk samples. This OTU dominated the *Bacillus* population (> 95%) in all samples, with a few exceptions. Two of the three replicates of the milk cartons stored at 8 °C from dairy B collected in April, were dominated by OTU 5 (species included in this OTU are described in Table S2). In these samples, both duplicates of the same carton contained 81 and 71 % of *Bacillus* OTU 5. The same OTU was also detected in two replicate samples from the March sampling from dairy A and showed a relatively abundance of 11.8 and 11.5 %.

3.2 New marker gene sequencing

The primers designed in this study targeted the *Bacillus cereus* group genes *panC, pycA* and *glpT* and generated PCR products of 304, 347 and 378 bp in length, respectively (Table 1). These three marker genes were used to further characterize the

Bacillus cereus group from raw milk samples from dairy silo tanks stored at 8 °C, and samples of carton milk stored at either 4 or 8 °C. However, no visible amplification products were received within 35 PCR cycles for some samples due to the low amount of *Bacillus*. An increased number of PCR cycles was not applied due to the potential formation of primer dimers, which could interfere with the sequencing quality and results.

Due to the low level of *Bacillus* in the other samples, only samples from milk cartons stored at 8 °C until end of shelf-life were used to evaluate the performance of the three primer sets in discriminating between species within the *Bacillus cereus* group population. A total of 285190, 175810 and 195668 good quality sequences were obtained for the *panC*, *glpT* and *pycA* genes, respectively. After grouping the sequences based on > 99% identity and excluding rare sequences (OTUs with relative abundance < 0.01 %), there were 186, 126 and 127 sequence groups for *panC*, *glpT* and *pycA*, respectively. The highest sequence group richness and highest diversity among the milk samples was found for the *panC* gene (Fig. 1). Both the feature richness and the diversity were significantly higher for *panC* compared to the other two genes (Fig. 1). The lower Shannon index detected for *glpT* and *pycA* compared to *panC* (Fig. 1) indicated that the communities described by these two genes were dominated by a few sequence groups of high abundance. This was also indicated by the relative abundance analysis (Fig. S1). Another advantage of using the *panC* gene was that it allowed for identification of *Bacillus cereus* sequence groups, according to Guinebretière et al. (2008), among the different phylogenetic groups (Fig. 5).

3.3 Evaluation of the *panC* library

The number of milk samples that generated positive amplification (within 35 cycles of PCR), were 78 and 66 for dairy A and B, respectively. Principal coordinate analysis of the weighted Unifrac distance showed that the *Bacillus* spp. composition was diverse between the three factors considered (temperature, dairy and month). Using multivariate analysis of variance, all the three factors significantly contributed to the diversity between the different samples (month: adonis p value 0.001, dairy: adonis p value 0.02, temperature adonis p value 0.004). The summer and autumn months (June, Jul, Aug, Sep, and Oct) were characterized by similar *B. cereus* group composition, as did also the winter months (Nov, Jan and Feb). Samples from March, April and May showed the largest variation in composition (Fig. 2). ANOVA of the beta-dispersion or homogeneity of group dispersions was not significant ($p > 0.05$) for all three factors tested, and therefore the dispersion within each group of samples was homogeneous. However, samples of carton milk stored at 8 °C had a higher distribution of the distance to centroid compared to samples of carton milk stored at 4 °C and samples of raw milk stored at 8 °C (Fig. 2). This indicates that the highest storage temperature increased the diversity of the *B. cereus* group composition compared to the lower storage temperature and compared to the raw milk samples stored at 8 °C.

The two most abundant *panC* sequence groups detected in milk from both dairies were panC_1 and panC_3. These two sequence groups accounted for 39 and 21 % of the total number of reads, respectively, and were detected in all samples. The findings from the multivariate analysis of variance were confirmed by the relative abundance analysis. The *Bacillus cereus* group population in milk cartons was influenced by the month of sampling, the dairy and the temperature of storage. In particular, a larger diversity among the *panC* sequences was found in samples from

March, April and May compared to other times of the year (Fig. 3). During these months some *panC* sequences groups, which were not detected or detected in low abundance during other months, dominated the *Bacillus* population (e.g. panC_28, panC_41, panC_7). Furthermore, the *Bacillus cereus* group population in samples from milk cartons was significantly different between the two storage temperatures (Adonis p value <0.001). The higher storage temperature increased the relative abundance of some *panC* OTUs (Fig. 4), as well as the percentage of *Bacillus cereus* group OTUs quantified over the entire sample microbiota (Fig. 3). In contrast, storage at 4 °C resulted in a more uniform distribution of the relative abundance of the *B. cereus* population throughout the year (Fig. 4). In addition to the higher *Bacillus* population diversity in cartons stored at 8 °C compared to cartons stored at 4 °C, the higher storage temperature increased the variation between replicate cartons. When stored at 8 °C, replicate cartons from the same sampling month, contained different sequence groups at 3 and 9 occasions for dairy A and B, respectively (Fig. 3 and Fig. S1). Particularly, for dairy B, samples from June 2015 and from October 2015 to May 2016 contained different sequence groups in at least one of the three replicate cartons stored at 8 °C. For some of these months, namely January, March and April, all the three replicates contained different compositions of *B. cereus* sequence groups. For dairy A, differences in sequence groups between replicate cartons were observed only in June and August 2015 and in March 2016 (Fig. S1).

Representative sequences for each of all *panC* sequence groups were identified and assigned to phylogenetic groups according to Guinebretiere *et al*. (2008). Among the 19 most abundant *panC* sequence groups, two phylogenetic groups were identified. These two groups were separated into two clusters based on their sequences (Fig. 5) and were assigned to the group VI (14 OTUs) and Group II (5

OTUs). While the first group dominated most of the samples (86% of the total number of reads), Group II (5 % of the total number of reads) was found in high abundance in a few samples of milk stored at 8 °C. In particular, the *panC* group sequences panC_28 and panC_41 were detected in all April milk carton replicates from dairy A stored at 8 °C until end of shelf-life, and in one replicate of carton milk collected from dairy B in May 2016 and stored under the same conditions.

4 Discussion

Utilization of the 16S rRNA gene for microbial community studies have previously revealed that the genus *Bacillus* is part of the core microbiota present in raw and processed milk (Kable et al., 2016; Porcellato et al., 2017). To describe the composition and dynamics of *Bacillus* spp. in samples of raw milk and consumption milk stored at different temperatures, two different approaches were used in the present study. First, the 16S rRNA library from Porcellato et al. (2017) was reanalyzed to target only sequences assigned to the genus *Bacillus*, and second, a new HTS sequencing approach was applied to target marker genes for the *Bacillus cereus* group.

The re-analysis of the 16S rRNA library revealed that the *Bacillus cereus* group was predominant in the milk samples. To further identify the composition of the *B. cereus* group in the milk samples, a culture-independent approach, based on three marker genes, was applied in the present study. This approach divided the *B. cereus* group population in several phylogenetically–related subgroups and provided an indication of the *B. cereus* group composition without the bias of a culturing step where the temperature, the media and the incubation conditions might influence the outcome. The *panC* gene, demonstrated the best discriminative power, and produced

a more comprehensive description of the phylogenetic diversity in the samples, compared to the *glpT* and *pycA* genes.

The two dominant phylogenetic groups, detected from the *panC* sequences, include, according to Guinebretiere *et al*. (2008), psychrotrophic strains of the genus *Bacillus*. Group VI, which include *B. weihenstephanensis, B. mycoides* and *B. thuringiensis* strains, was most abundant. This group contains strains able to grow at low temperature (5 °C) and can be separated from strains in group II (the second most abundant phylogenetic group in our study) by the presence of a *cspA* gene signature (Francis et al.,1998). Group II includes both mesophilic and psychrotolerant strains of *B. cereus* and *B. thuringiensis* (Guinebretiere 2008).

The two storage temperatures used in this study represented optimal (4 °C) and sub-optimal (8 °C) storage conditions for consumption milk. In milk stored at 4 °C, the two dominant sequence groups were detected over the entire year at both dairies. In our previous work, we found that the level and relative abundance of *Bacillus* in samples from carton milk did not change between day 1 after production to the end of shelf-life when stored at 4 °C (Porcellato et al., 2017). In contrast, storage of milk cartons at 8 °C changed the microbial composition and gave significantly higher plate count levels of presumptive *B. cereus* (>log 6 CFU / mL, Porcellato et al. (2017). We therefore hypothesized that a change in the *B. cereus* group population is very unlikely to happen at 4 °C.

In the present work, we found that the two most abundant *panC* sequence groups (panC_1 and panC_3) did not always dominate the *B. cereus* group population after storage at 8 °C. In samples stored at 8 °C, the dominant sequence groups differed between the two dairies involved, the sampling months and also between the replicates within each sampling. Dairy B had more sampling months with a *B. cereus*

group composition that differed between the two storage temperatures, compared to dairy A (9 months vs 3 months, respectively). Seasonal differences in the *B. cereus* group population were also found; during the spring months (March, April and May) the samples were dominated by sequence groups that were only present in low abundance or even not detected during the rest of the year. Variations in the milk microbiota over time and between samples have also been described previously (Doyle et al., 2017; Kable et al., 2016). Our results show that compositional differences between milk samples also apply for the genus *Bacillus*. This is supported by previous culture-dependent studies where different *Bacillus cereus* group isolates have been collected from milk produced in different countries and at different times of the year (Coorevits et al., 2008; Schmidt et al., 2012). Another important consideration about the *B. cereus* group population, which grew at 8 °C, is the differences observed between replicate samples. Each replicate was obtained from a single milk carton and all cartons were collected at the same day from the same production line and production batch at each dairy plant. This suggests that some, but not all the cartons, were contaminated by a low abundance *B. cereus* sequence groups, capable of growing at 8 °C. These sequence groups could already be present in the raw milk and survive the pasteurization process, or they could have re-contaminated the milk during or after processing. In a previous study, *Bacillus* spores were detected in all parts of biofilms and in particular in those parts that were in contact with the growth medium (Faille et al., 2014). During milk processing, spores and vegetative cells from biofilms can end up in the milk at a low concentration and contaminate the final product.

In conclusion, we present a new approach to target the *Bacillus cereus* group in consumption milk. This new method allowed us to obtain a more detailed

understanding of the structure of the *B. cereus* group population that contaminate milk and how it is influenced by the storage temperature. The *B. cereus* population in milk was composed of a mix of psychrotolerant strains divided in several sequence groups. The observed variation in sequence groups between months and dairies, suggest that the composition of the *B. cereus* group population in milk is dynamic and influenced by several factors along the value chain. More detailed studies on the composition of the *B. cereus* population along the milk value chain, from raw milk to the end product, are necessary to obtain a more in-depth understanding on the pathways for contamination. Such investigations might consider applying the method described in this paper to achieve a higher resolution on characterization of the *B. cereus* group population.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Ahmed Abdelghani, Marte Monshaugen (Norwegian University of Life Sciences, NMBU) and the Norwegian dairy cooperative TINE SA for their assistance during sampling and DNA isolation. This work received financial support from the Norwegian Foundation for Research Levy on Agricultural Products (FFL) and the Norwegian Agricultural Agreement Research Fund (JA) (grant no 244149/E50). Department of Chemistry, Biotechnology and Food Science (NMBU), Center for Food Safety (NMBU), and TINE SA also contributed economically to this project.

REFERENCE

Anderson, M.J., Ellingsen, K.E., McArdle, B.H., 2006. Multivariate dispersion as a measure of beta diversity. Ecol. Lett. 9, 683-693.

Aronson, A.I., Shai, Y., 2001. Why Bacillus thuringiensis insecticidal toxins are so effective: unique features of their mode of action. Fems Microbiol. Lett. 195, 1-8.

Bartoszewicz, M., Hansen, B.M., Swiecicka, I., 2008. The members of the Bacillus cereus group are commonly present contaminants of fresh and heat-treated milk. Food Microbiol. 25, 588-596.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST plus : architecture and applications. Bmc Bioinformatics 10.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Tumbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335-336.

Christiansson, A., Bertilsson, J., Svensson, B., 1999. Bacillus cereus spores in raw milk: Factors affecting the contamination of milk during the grazing period. J. Dairy Sci. 82, 305-314.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B.L., McGarrell, D.M., Sun, Y.N., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database

Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633-D642.

Coorevits, A., De Jonghe, V., Vandroemme, J., Reekmans, R., Heyrman, J., Messens, W., De Vos, P., Heyndrickx, M., 2008. Comparative analysis of the diversity of aerobic spore-forming bacteria in raw milk from organic and conventional dairy farms. Syst. Appl. Microbiol. 31, 126-140.

Doyle, C.J., Gleeson, D., O'Toole, P.W., Cotter, P.D., 2017. Impacts of Seasonal Housing and Teat Preparation on Raw Milk Microbiota: a High-Throughput Sequencing Study. Appl. Environ. Microb. 83.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460-2461.

Eneroth, A., Svensson, B., Molin, G., Christiansson, A., 2001. Contamination of pasteurized milk by Bacillus cereus in the filling machine. J. Dairy Res. 68, 189-196.

Ercolini, D., 2013. High-Throughput Sequencing and Metagenomics: Moving Forward in the Culture-Independent Analysis of Food Microbial Ecology. Appl. Environ. Microb. 79, 3148-3155.

Faille, C., Benezech, T., Midelet-Bourdin, G., Lequette, Y., Clarisse, M., Ronse, G., Ronse, A., Slomianny, C., 2014. Sporulation of Bacillus spp. within biofilms: A potential source of contamination in food processing environments. Food Microbiol. 40, 64-74.

Flint, S.H., Bremer, P.J., Brooks, J.D., 1997. Biofilms in dairy manufacturing plant - Description, current concerns and methods of control. Biofouling 11, 81-97.

Francis, K.P., Mayr, R., von Stetten, F., Stewart, G.S., Scherer, S., 1998. Discrimination of psychrotrophic and mesophilic strains of the Bacillus cereus group

by PCR targeting of major cold shock protein genes. Appl Environ Microbiol 64, 3525-3529.

Granum, P.E., Lindbäck, T., 2013. *Bacillus cereus*, in: Doyle, M.P., Buchanan, R.L. (Eds.), Food Microbiology: Fundamentals and Frontiers, 4th ed. ASM Press, Washington, D.C. , pp. 491-502.

Guinebretiere, M.H., Auger, S., Galleron, N., Contzen, M., De Sarrau, B., De Buyser, M.L., Lamberet, G., Fagerlund, A., Granum, P.E., Lereclus, D., De Vos, P., Nguyen-The, C., Sorokin, A., 2013. Bacillus cytotoxicus sp. nov. is a novel thermotolerant species of the Bacillus cereus Group occasionally associated with food poisoning. Int J Syst Evol Microbiol 63, 31-40.

Guinebretiere, M.H., Thompson, F.L., Sorokin, A., Normand, P., Dawyndt, P., Ehling-Schulz, M., Svensson, B., Sanchis, V., Nguyen-The, C., Heyndrickx, M., De Vos, P., 2008. Ecological diversification in the Bacillus cereus Group. Environ Microbiol 10, 851-865.

Jung, M.Y., Kim, J.S., Paek, W.K., Lim, J., Lee, H., Il Kim, P., Ma, J.Y., Kim, W., Chang, Y.H., 2011. Bacillus manliponensis sp nov., a new member of the Bacillus cereus group isolated from foreshore tidal flat sediment. J Microbiol 49, 1027-1032.

Kable, M.E., Srisengfa, Y., Laird, M., Zaragoza, J., McLeod, J., Heidenreich, J., Marco, M.L., 2016. The Core and Seasonal Microbiota of Raw Bovine Milk in Tanker Trucks and the Impact of Transfer to a Milk Processing Facility. Mbio 7.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772-780.

Lechner, S., Mayr, R., Francis, K.P., Pruss, B.M., Kaplan, T., Wiessner-Gunkel, E., Stewart, G.S., Scherer, S., 1998. Bacillus weihenstephanensis sp. nov. is a new psychrotolerant species of the Bacillus cereus group. Int. J. Sys. Bact. 48 Pt 4, 1373-1382.

Liu, B., Liu, G.H., Hu, G.P., Cetin, S., Lin, N.Q., Tang, J.Y., Tang, W.Q., Lin, Y.Z., 2014. Bacillus bingmayongensis sp nov., isolated from the pit soil of Emperor Qin's Terra-cotta warriors in China. Anton Leeuw Int J G 105, 501-510.

Liu, Y., Lai, Q.L., Goker, M., Meier-Kolthoff, J.P., Wang, M., Sun, Y.M., Wang, L., Shao, Z.Z., 2015. Genomic insights into the taxonomic status of the Bacillus cereus group. Sci. Rep. Uk 5.

Lozupone, C., Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microb. 71, 8228-8235.

Magnusson, M., Christiansson, A., Svensson, B., 2007. Bacillus cereus spores during housing of dairy cows: Factors affecting contamination of raw milk. J. Dairy Sci. 90, 2745-2754.

Miller, R.A., Beno, S.M., Kent, D.J., Carroll, L.M., Martin, N.H., Boor, K.J., Kovac, J., 2016. Bacillus wiedmannii sp nov., a psychrotolerant and cytotoxic Bacillus cereus group species isolated from dairy foods and dairy environments. Int. J. Syst. Evol. Micr. 66, 4744-4753.

Novak, J.S., Call, J., Tomasula, P., Luchansky, J.B., 2005. An assessment of pasteurization treatment of water, media, and milk with respect to Bacillus spores. J. Food Protect 68, 751-757.

Oksanen, J., Blanchet, F.G., Friendly, F., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H.,

Szoecs, E., Wagner, H., 2017. vegan: Community Ecology Package. R package version 2.4-2. https://CRAN.R-project.org/package=vegan.

Patino-Navarrete, R., Sanchis, V., 2016. Evolutionary processes and environmental factors underlying the genetic diversity and lifestyles of Bacillus cereus group bacteria. Res. Microbiol.

Porcellato, D., Narvhus, J., Skeie, S.B., 2016a. Detection and quantification of Bacillus cereus group in milk by droplet digital PCR. J. Microbiol. Meth. 127, 1-6.

Porcellato D., Aspholm M., Skeie S.B., Monshaugen M., Brendehaug J., Mellegård H. 2017. Microbial diversity of consumption milk during processing and storage. Int. J. Food Microboil.. Submitted.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592-593.

Schmidt, V.S.J., Kaufmann, V., Kulozik, U., Scherer, S., Wenning, M., 2012. Microbial biodiversity, quality and shelf life of microfiltered and pasteurized extended shelf life (ESL) milk from Germany, Austria and Switzerland. Int. J. Food Microbiol. 154, 1-9.

Setlow, P., 2003. Spore germination. Curr. Opin. Microbiol. 6, 550-556.

Svensson, B., Eneroth, A., Brendehaug, J., Molin, G., Christiansson, A., 2000. Involvement of a pasteurizer in the contamination of milk by Bacillus cereus in a commercial dairy plant. J. Dairy Res. 67, 455-460.

Tourasse, N.J., Helgason, E., Okstad, O.A., Hegna, I.K., Kolsto, A.B., 2006. The Bacillus cereus group: novel aspects of population structure and genome dynamics. J. Appl. Microbiol. 101, 579-593.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res. 40.

Table 1. List of primers designed in this study.

Target gene	Gene name	Primer forwar (5’ - 3’)	Primer Reverse (5’ - 3’)	Annealing temperature (°C)	Length of PCR fragment (bp)
panC	Pantothenate synthetase	panC_F TCCWGCAGAACARACRACAA	panC_R GAGGAGCYTCYTCACGCTCW	52	304
glpT	Glycerol-3-phosphate transporter	glpT_F TGCGGMTGGATGAGYGA	gplT_R AAGTWAGRGCAAGGAAKA	54	378
pycA	Pyruvate carboxylase	pycA_F CTAYGCWCCRTTTGAAAGTG	pycA_R TTTTTCGGGAAACCACCRTA	52	347

Fig. 1. Alpha diversity indexes obtained from milk samples stored at 8 °C for 13 days and grouped by gene. A) Richness measured using the Chao1 index. B) Alpha diversity measure with the Shannon index. *) Significantly different genes obtained using t-test with Monte Carlo simulation.

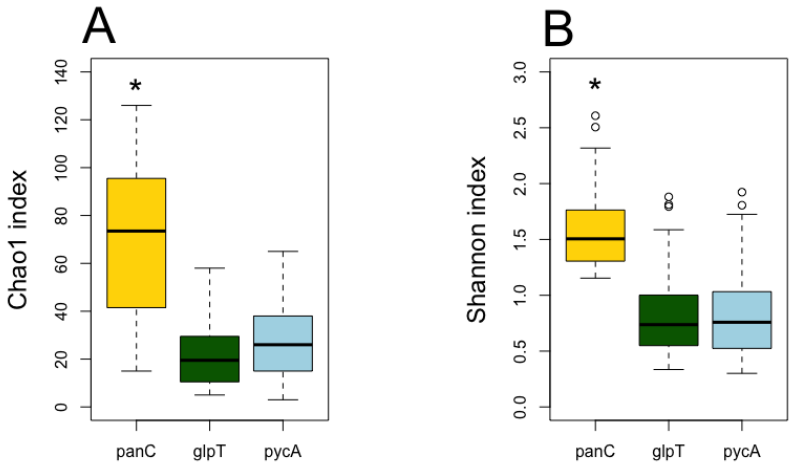
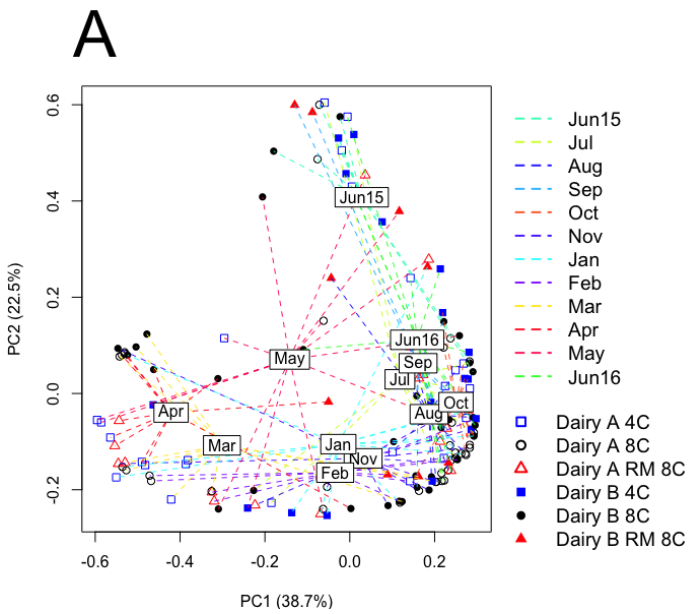


Fig. 2. A) Principal coordinate analysis of the weighted Unifrac distance between the *Bacillus cereus* group populations. B) Boxplot of the beta-dispersion grouped by dairy and temperature. 4C: milk samples from cartons stored at 4 °C for 13 days, 8C: milk samples from cartons stored at 8 °C for 13 days, RM: raw milk collected from the silo tank and stored at 8 °C for 13 days.



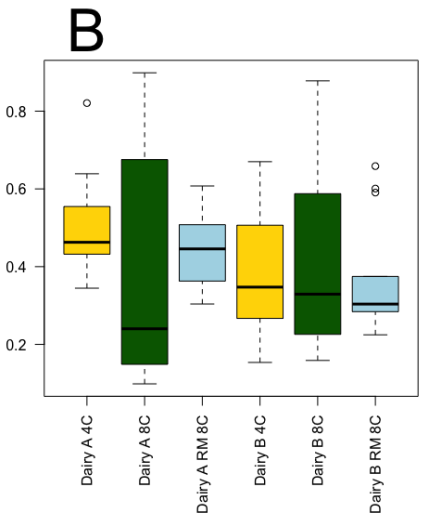


Fig. 3. Relative abundance of the *panC* sequence groups obtained for milk samples with positive amplification from A) dairy A, B) dairy B. The group number next to the sequence group name indicates the affiliation to phylogenetic group according to Guinebretiere et al. (2008). 4: milk samples from cartons stored at 4 °C for 13 days, 8: milk samples from cartons stored at 8 °C for 13 days, R: raw milk collected from the silo tank and stored at 8 °C for 13 days.

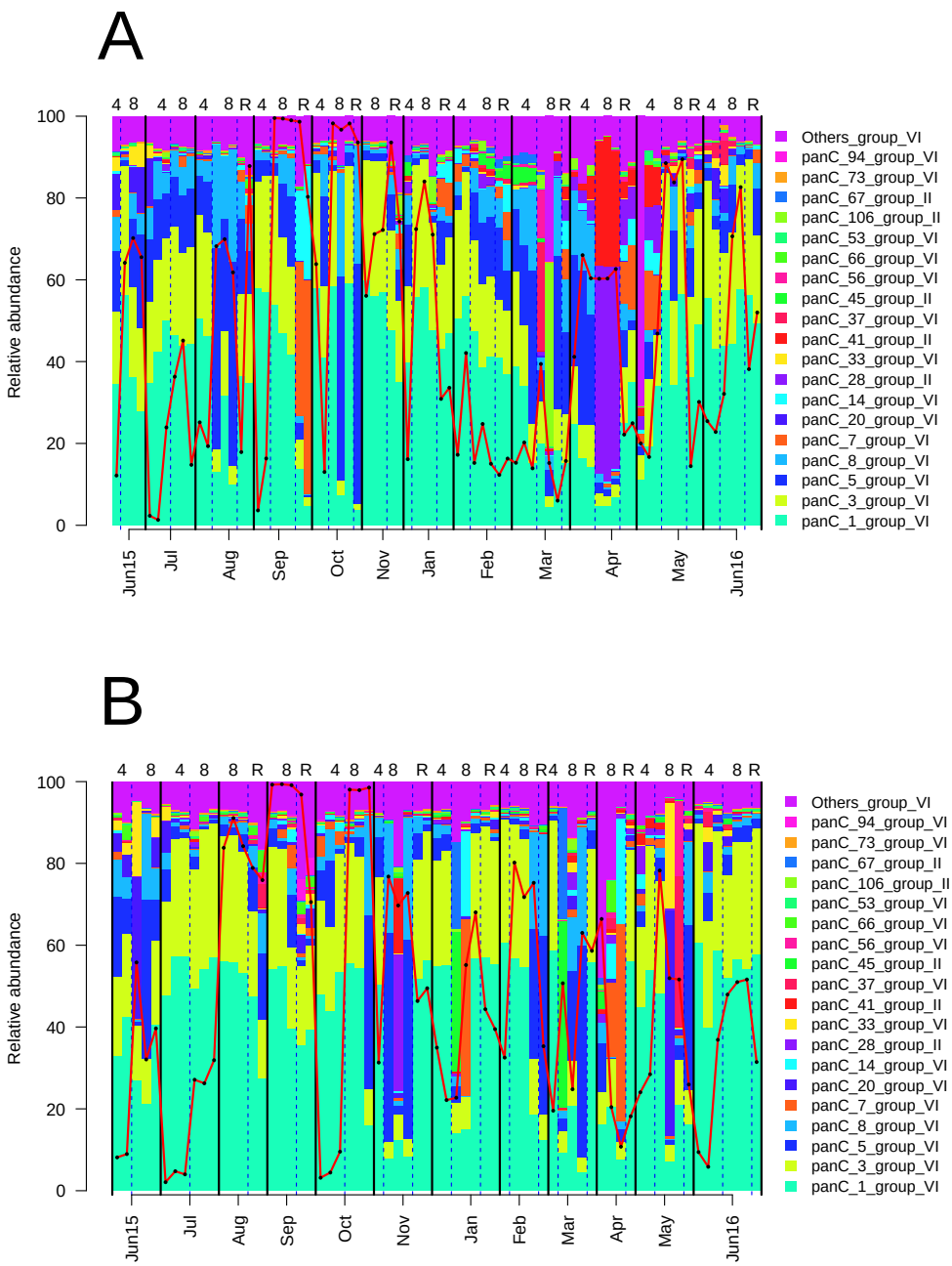
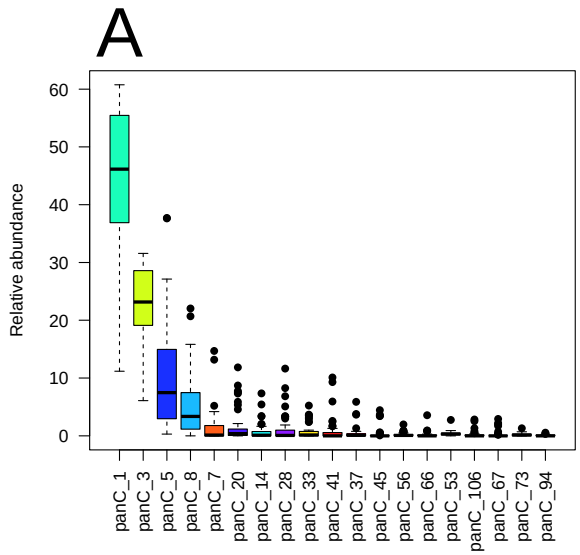


Fig. 4: Relative abundance distribution of the 19 most abundant *panC* OTUs for milk cartons stored at A) 4 °C and B) 8 °C.



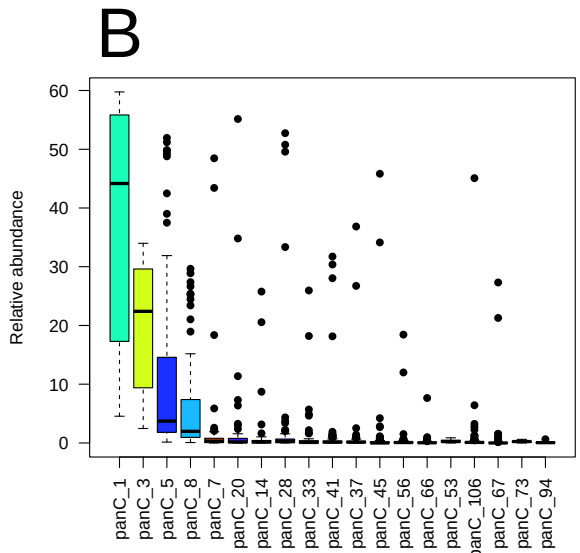


Fig. 5. Neighbor-joining phylogenetic tree obtained from the most abundant *panC* OTU sequences. Bootstrap values based on 999 replications are shown at the nodes of the tree. The cluster on the top (14 sequence groups) belongs to the phylogenetic group VI, while the cluster on the bottom (5 sequence groups) belongs to the phylogenetic group II according to Guinebretiere et al. (2008).

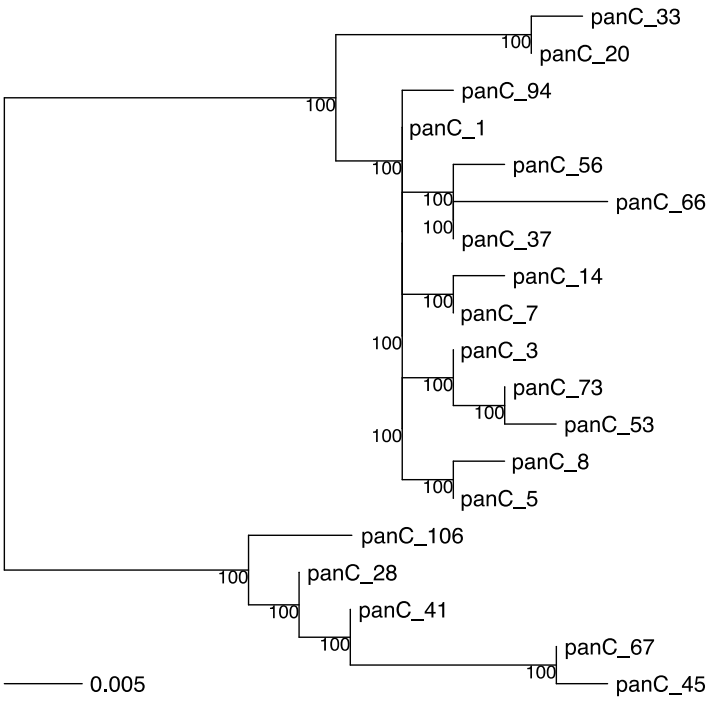
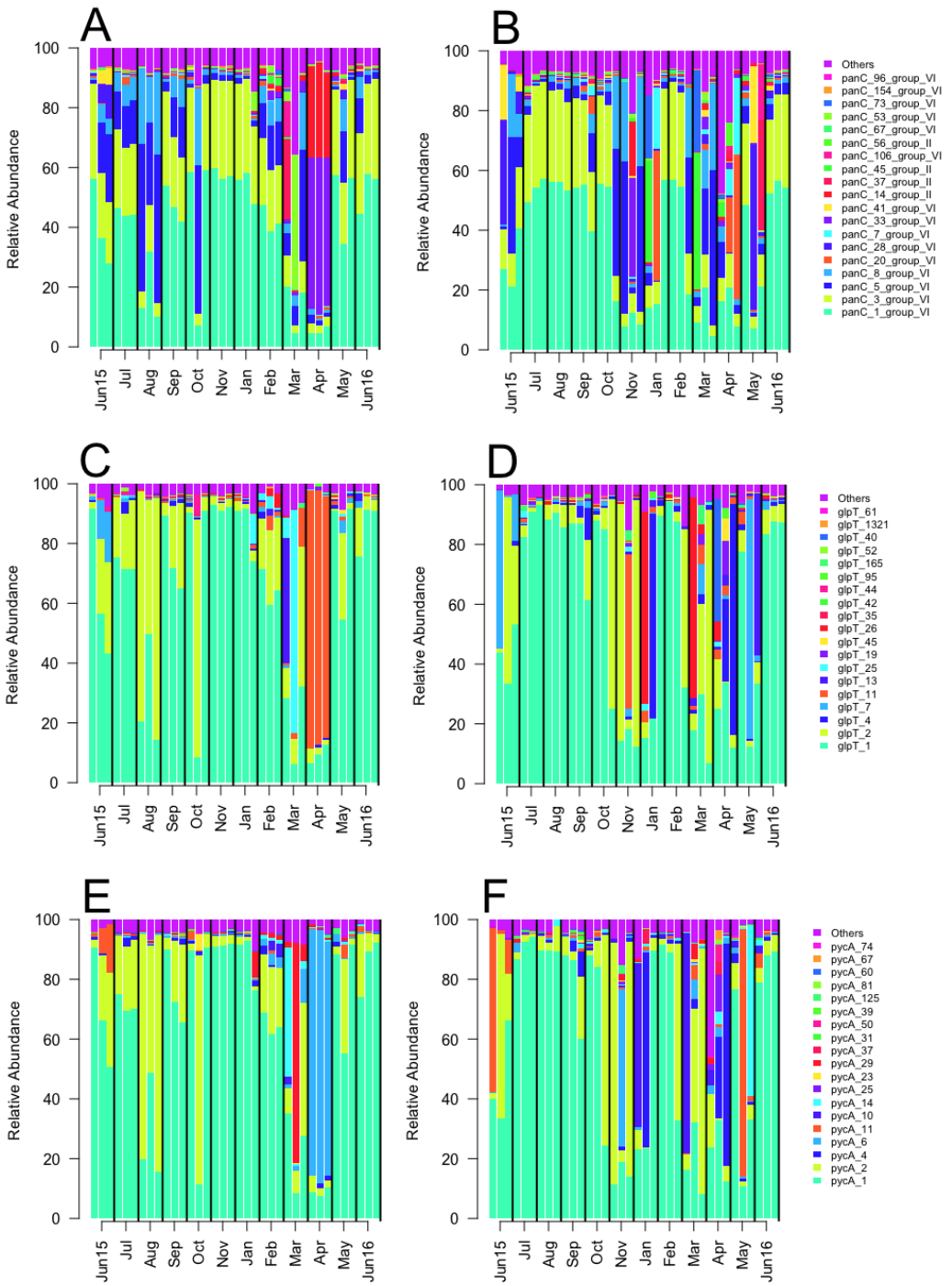


Fig. S1. Relative abundance of the *panC* (A and B), *glpT* (C and D) and *pycA* (E and F) sequence groups obtained from milk samples from dairy A and (A, C and E) and from dairy B (B, D and F). The milk samples were stored at 8 °C for 13 days before analysis. The group number next to the *panC* sequence group name indicates the affiliation to phylogenetic groups according to Guinebretiere et al. (2008)



TableS1. Bacterial species used to check for inclusivity and exclusivity of the primers designed for *panC*, *glpT* and *pycA*.

Species	***panC***	***glpT***	***pycA***
Bacillus cereus ATCC 14579	+	+	+
Bacillus mycoides T158	+	+	+
Bacillus mycoides T143	+	+	+
Bacillus smithinii T189	-	-	-
Bacillus licheniformis T56	-	-	-
Bacillus amyloliguenfaciens T390	-	-	-
Bacillus pumilus T77	-	-	-
Kocuria rhizophila T10	-	-	-
Streptococcus thermophilus T5	-	-	-
Aneurinibacillus thermoaerophilus T216	-	-	-
Oxybacillus flavithermus T223	-	-	-
Brevibacillus thermoruber T303	-	-	-
Staphylococcus warneri T325	-	-	-

Table S2. *Bacillus* species included in each OTUs after clustering the in silico PCR product with 99% identity. *Bacillus* sequences were obtained from the RDP database (Cole et al., 2014).

OTU_1	**OTU_5**	**OTU_10**	**OTU_3**	**OTU_4**	**OTU_15**
anthracis	*butanolivorans*	*coagulans*	*aerius*	*aryabhattai*	*infantis*
bombysepticus	*frigoritolerans*		*licheniformis*	*flexus*	
brevis	*litoralis*		*sonorensis*	*horikoshii*	
cereus	*macroides*			*megaterium*	
gaemokensis	*muralis*			*meqaterium*	
mycoides	*niabensis*				
pseudomycoides	*psychrosaccharolyticus*				
salmalaya	*simplex*				
samanii					
thuringiensis					
toyonensis					
weihenstephanensis					