

RESEARCH ARTICLE

Open Access

Genomic comparison of sporeforming bacilli isolated from milk

Andrea I Moreno Switt, Alexis D Andrus, Matthew L Ranieri, Renato H Orsi, Reid Ivy, Henk C den Bakker, Nicole H Martin, Martin Wiedmann and Kathryn J Boor*

Abstract

Background: Sporeformers in the order Bacillales are important contributors to spoilage of pasteurized milk. While only a few *Bacillus* and *Viridibacillus* strains can grow in milk at 6°C, the majority of *Paenibacillus* isolated from pasteurized fluid milk can grow under these conditions. To gain a better understanding of genomic features of these important spoilage organisms and to identify candidate genomic features that may facilitate cold growth in milk, we performed a comparative genomic analysis of selected dairy associated sporeformers representing isolates that can and cannot grow in milk at 6°C.

Results: The genomes for seven *Paenibacillus* spp., two *Bacillus* spp., and one *Viridibacillus* sp. isolates were sequenced. Across the genomes sequenced, we identified numerous genes encoding antimicrobial resistance mechanisms, bacteriocins, and pathways for synthesis of non-ribosomal peptide antibiotics. Phylogenetic analysis placed genomes representing *Bacillus*, *Paenibacillus* and *Viridibacillus* into three distinct well supported clades and further classified the *Paenibacillus* strains characterized here into three distinct clades, including (i) clade I, which contains one strain able to grow at 6°C in skim milk broth and one strain not able to grow under these conditions, (ii) clade II, which contains three strains able to grow at 6°C in skim milk broth, and (iii) clade III, which contains two strains unable to grow under these conditions. While all *Paenibacillus* genomes were found to include multiple copies of genes encoding β -galactosidases, clade II strains showed significantly higher numbers of genes encoding these enzymes as compared to clade III strains. Genome comparison of strains able to grow at 6°C and strains unable to grow at this temperature identified numerous genes encoding features that might facilitate the growth of *Paenibacillus* in milk at 6°C, including peptidases with cold-adapted features (flexibility and disorder regions in the protein structure) and cold-adaptation related proteins (DEAD-box helicases, chaperone DnaJ).

Conclusions: Through a comparative genomics approach we identified a number of genomic features that may relate to the ability of selected *Paenibacillus* strains to cause spoilage of refrigerated fluid milk. With additional experimental evidence, these data will facilitate identification of targets to detect and control Gram positive spore formers in fluid milk.

Background

Microbial spoilage of foods, particularly highly perishable foods (such as fluid milk produced by high temperature short time [HTST] pasteurization), is a considerable challenge that needs to be addressed to successfully supply food for an anticipated world population of 9 billion people by 2050 [1]. For example, while approximately 87 billion liters of milk are produced in the US annually [2,3], as much as 23 billion liters is

potentially lost every year due to microbial spoilage [4]. Whereas manufacturing-related spoilage (e.g. post processing contamination) can be eliminated, contamination with Gram-positive sporeformers, which are microorganisms that can survive pasteurization, is not easily eliminated [5-8]. Among these sporeformers, those that can grow at refrigeration temperatures are a specific concern with regard to food spoilage and spoilage of fluid milk in particular. Two bacterial genera, *Bacillus* and *Paenibacillus*, have been found to be the predominant aerobic sporeformers isolated from pasteurized milk [5,7-11]. *Bacillus* spp. and *Paenibacillus* spp. have also

* Correspondence: kjb4@cornell.edu
345 Stocking Hall, Department of Food Science, Cornell University, Ithaca, NY 14853, USA

been isolated from dairy farm environments and processing facilities [12], suggesting that these spoilage organisms can be introduced at various stages of the dairy production chain. While the predominant sporeformers isolated early in the shelf life of HTST pasteurized milk are typically *Bacillus* spp., later in the shelf life a shift in bacterial ecology has been reported with *Paenibacillus* spp. representing the majority of sporeformers isolated [7,8], suggesting that *Paenibacillus* spp. may, on a population basis, show a better ability to grow in milk stored at refrigeration temperatures. Molecular and phenotypic characterization of >1,200 sporeformers isolated from milk production not only supported that *Bacillus* spp. and *Paenibacillus* spp. represented the phylogenetic groups most common among aerobic sporeformers isolated from milk, but also identified a considerable number of *Viridibacillus* spp. isolates [5]. This study also showed that while most *Paenibacillus* spp. display β -galactosidase activity and were able to grow in milk at 6°C, only few *Bacillus* spp. and *Viridibacillus* spp. isolates show these phenotypic characteristics. It has been well established though that isolates representing specific *Bacillus* spp. (*B. weihenstephanensis*) and *Viridibacillus* spp. (*V. arenosi*) typically have the ability to grow in refrigerated milk [5]. While *Paenibacillus* spp. are thus the main aerobic sporeformer group of concern as a spoilage organism for the fluid milk industry, some *Bacillus* and *Viridibacillus* spp. may also cause spoilage problems in these types of products.

Considerable efforts have been undertaken to analyze and characterize the genomes of *Bacillus cereus* sensu lato strains, which include human pathogenic species (e.g. *B. anthracis* [13,14]), and foodborne pathogens (e.g. *B. cereus* [14,15]). Conversely, our current knowledge of *Paenibacillus* spp. is rather limited. Recently, a number of *Paenibacillus* genomes, representing strains isolated from different environments (e.g. soil, honeybees, fermented food), have been sequenced and their genomes have been released ([16-25]); however, only a few studies have reported genomic characterization and phylogeny of the sequenced genomes [26-31]. Species belonging to the *Paenibacillus* genus have been isolated from diverse environments including soil, water, milk, insect larvae, and humans [30,32]. Some genomic characteristics associated with the adaptation and survival of *Paenibacillus* in different environments include genes encoding nitrogen fixation enzymes, or genes encoding antimicrobial compounds and bacteriocin production [29,32]. To date, genomic characteristics of *Paenibacillus* strains associated with spoilage of fluid milk remain unknown. Here we analyzed the genomes of isolates representing common *rpoB* allelic types of *Bacillus*, *Paenibacillus* and *Viridibacillus* spp. that were found to contaminate pasteurized milk [5]. The aim was to explore genomic

characteristics linked with phenotypic characteristics related to milk spoilage, such as the ability to grow in milk at 6°C and proteolysis of milk proteins, such as casein.

Results and discussion

The genomic comparisons and supporting phenotypic experiments reported here focused on (i) developing an improved understanding of the genomics of sporeforming food spoilage organisms in the order Bacillales and (ii) identifying specific genomic features that may allow specific members of this order to grow in milk at refrigeration temperatures. As all isolates sequenced here represent species and genetic types that are frequently isolated from commercially processed dairy products, we also hypothesized that genomic analyses may identify novel generally recognized as safe (GRAS) compounds and enzymes that are predicted to be produced by these organisms and that may be useful for food applications (e.g. novel bacteriocins). This hypothesis is supported by previous studies that have identified, through non-genomics classical screening approaches, GRAS bacteriocins from *Paenibacillus* isolates [33].

Phylogenomic analysis places *Bacillus*, *Paenibacillus* and *Viridibacillus* into three highly divergent clades, including three distinct *Paenibacillus* clades

The genomes of ten sporeforming Gram positive bacteria isolated from pasteurized milk and dairy farm environments (feed mix, see Table 1) were sequenced; these isolates represented *Paenibacillus* spp. (n = 7), *Bacillus* spp. (n = 2) and *Viridibacillus* sp. (n = 1) (Table 1). Genome sequencing was conducted with the Illumina HiSeq; the draft genomes reported here represented 47 contigs for *Viridibacillus* sp., 106 to 122 contigs for *Bacillus* spp., and 66 to 161 contigs for *Paenibacillus* spp. isolates (see Table 1 for detailed genome assembly statistics). Due to the fact that the genomes were not closed, it is possible that specific genes or orthologs could be misclassified as absent in a given genome; this should not affect the overall enrichment analyses reported below though and absence of specific genes was confirmed in selected target genomes through read mapping (see Methods). Genome sizes (as inferred from the total assembly size) for the two *Bacillus* strains sequenced here were approximately 5.6 Mb, while the genome sizes for the *Paenibacillus* spp. isolates ranged from approx. 6.4 to 7.6 Mb (Table 1). The *Viridibacillus* strain sequenced represented the smallest genome (4.4 Mb). These genome sizes are consistent with previous data, which also found smaller genome sizes for *Bacillus* isolates (ranging from 3.7 Mb for *B. pumilus* [GenBank: CP000813] [34] to 5.7 Mb for *B. pseudomycoloides* [GenBank: CM000745]) as compared to *Paenibacillus* spp. isolates; previously reported genome sizes for *Paenibacillus* spp. isolates ranged from 5.3 Mb (*P. polymyxa*; GenBank: CP000154 [35]) to 8.6 Mb

Table 1 Strains and genome description of strains used in this study

Strain	Species ¹	<i>rpoB</i> allelic type (frequency of AT as reported by Ivy et al. (2012)) ²	Source (day of shelf life)	Total assembly size (number of contigs) ³	N50	Mol% (G + C) ⁴	GenBank accession number
FSL R5-860	<i>Bacillus</i> sp. ⁵	158 (10.6%)	Pasteurized 2% milk (7 days)	5.6 Mb (106)	337,248	35.0	ASPZ00000000
FSL H7-687	<i>Bacillus weihenstephanensis</i>	3 (1.5%)	Pasteurized 2% milk (12 days)	5.6 Mb (112)	405,076	35.2	ASPY00000000
FSL R5-213	<i>Viridibacillus arenosi</i>	17(1.8%)	Pasteurized 2% milk (7 days)	4.4 Mb (47)	661,909	35.3	ASQA00000000
FSL H8-237	<i>Paenibacillus odorifer</i>	15 (9.5%)	Pasteurized 2% milk (21 days)	7.3 Mb (128)	256,380	44.0	ASPV00000000
FSL R7-277	<i>Paenibacillus</i> sp. ⁶	45 (0.2%)	Pasteurized 2% milk (7 days)	7.6 Mb (122)	257,268	52.5	ASPX00000000
FSL R7-269	<i>Paenibacillus</i> sp. ⁶	163 (0.7%)	Pasteurized 2% milk (7 days)	7.5 Mb (161)	252,603	51.8	ASPS00000000
FSL R5-192	<i>Paenibacillus amylolyticus</i>	23 (2.7%)	Pasteurized 2% milk (7 days)	7.0 Mb (80)	383,457	45.8	ASPR00000000
FSL H7-689	<i>Paenibacillus amylolyticus</i>	23 (2.7%)	Pasteurized 2% milk (12 days)	6.8 Mb (66)	525,446	45.9	ASPU00000000
FSL R5-808	<i>Paenibacillus glucanolyticus</i>	159 (0.5%)	Pasteurized 2% milk (1 day)	6.4 Mb (89)	316,677	48.8	ASPT00000000
FSL H8-457	<i>Paenibacillus lautus</i>	117 (0.1%)	Feed mix (N/A)	7.0 Mb (69)	696,553	51.2	ASPW00000000

¹Putative species were inferred based on 16S rRNA phylogenies (using 16S rRNA sequences extracted from the draft genome sequences) and confirmed using phylogenetic analysis of 31 protein sequences (Figure 1).

²Percentage of isolates with a given allelic type (AT) from a total of 1,288 isolates tested.

³Only contigs >199 bp are included.

⁴Inferred from genome sequence.

⁵FSL R5-860 was designated as *Bacillus* sp. as this strain clusters with *B. cereus* in a 16S rRNA phylogeny, but clusters with *B. weihenstephanensis* in the 31 protein phylogeny (Figure 1); while this is consistent with previous reports that *B. cereus* is polyphyletic, final species assignment of this isolate will require sequencing and analysis of further *Bacillus* isolates.

⁶Isolates FSL R7-277 and R7-269 were classified as *Paenibacillus* sp. as they could not be classified to species based on 16S rRNA sequences; in a 16S rRNA phylogeny both of these isolates clustered with *P. caespitis* (GenBank accession no. AM745263), but were not assigned this species name as this species has not yet been validly published.

(*P. mucilaginosus*; GenBank: CP002869 [25]). It has been previously suggested [18] that this difference in genome sizes may indicate a role of widespread horizontal gene transfer or gene duplication in the evolution of different *Paenibacillus* species.

Phylogenetic analysis of the genomes sequenced here, along with 52 genomes representing selected previously sequenced Firmicutes, including five additional *Paenibacillus* and nine additional *Bacillus* (Figure 1 and Additional file 1), was performed using amino acid (aa) sequences encoded by 31 core genes previously reported to represent suitable phylogenetic markers for bacteria [36]. This analysis places *Bacillus*, *Paenibacillus* and *Viridibacillus* into three distinct well supported clades, which is consistent with reclassification, in 1993, of species now designated as *Paenibacillus* into a genus distinct from the genus *Bacillus* [37]. As no genome sequence for the genus *Viridibacillus* has previously been reported, the analyses reported here also provided an opportunity to clarify the phylogenetic position of a representative of this genus. Based on the 31 protein sequence phylogeny, the *V. arenosi* strain sequenced here falls

into a distinct clade that shares a common ancestor with a number of non-sporeforming Firmicutes (Figure 1). This not only confirms that *Viridibacillus* is distinct from *Bacillus* and *Paenibacillus*, but also suggests a single emergence of the most recent common ancestor of non-sporeforming Firmicutes from a sporeforming ancestor.

Our phylogenetic analysis of 31 protein sequences (Figure 1) also identified three distinct well supported clades among the *Paenibacillus* genomes sequenced here, including (i) clade I, which contains the two *P. amylolyticus* strains (FSL H7-689 and FSL R5-192), (ii) clade II, which contains the two *Paenibacillus* spp. strains (FSL R7-277 and FSL R7-269) and one *P. odorifer* (FSL H8-237) strain, and (iii) clade III, which contains two *P. lautus* strains (one sequenced in here [FSL H8-457] and one previously sequenced [Y4.12MC10]) and one *P. glucanolyticus* strain (FSL R5-808) (Figure 1). These three clades are well supported (100% bootstrap); however, the number of genomes characterized here is small and additional studies will be necessary to further validate the three clades identified here. Considerable

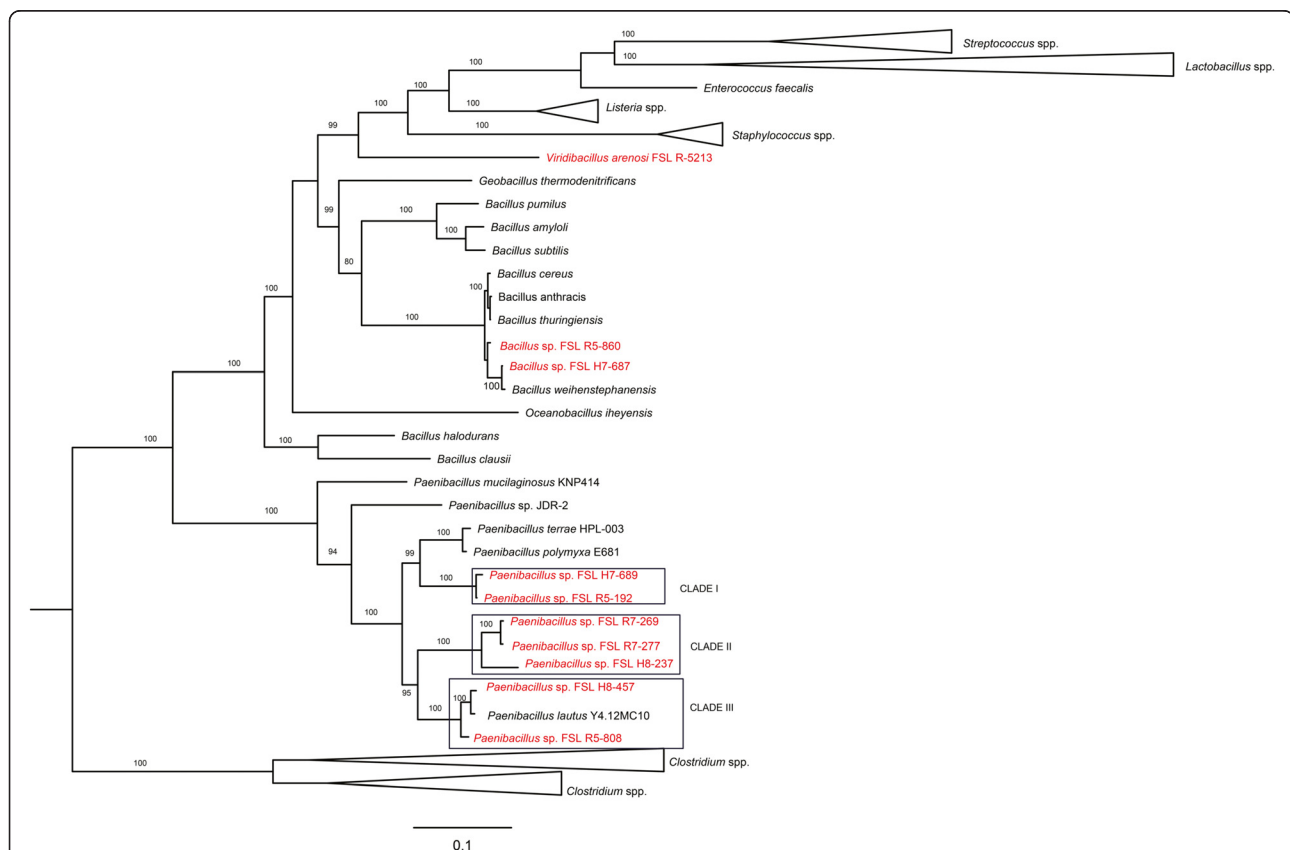


Figure 1 Maximum likelihood phylogeny from a concatenated amino acid sequence of 31 core genes. Genomes sequenced here as well as data for selected *Bacillus*, *Paenibacillus*, and additional Firmicutes were used to obtain the gene sequences for the phylogenetic analysis. The clade containing *Clostridium* spp. was selected to root the tree. For better visualization of the genera of interest (*Bacillus*, *Paenibacillus*, and *Viridibacillus*), branches that contained genomes of other genera were collapsed. Numbers on the branches represent bootstrap values calculated with 1000 replicates. Bootstrap values <80 are not included. Strains sequenced in this study are shown in red. The three *Paenibacillus* clades are labeled.

genome divergence among the analyzed *Paenibacillus* genomes is also supported by an overall genome alignment (see Additional file 2). Analysis of the average nucleotide identity based on the BLAST algorithm (ANIb) scores for *Paenibacillus* isolates sequenced here confirmed that isolates FSL H7-689 and FSL R5-192 (both classified as *P. amylolyticus* based on 16S rRNA) represent closely related genomes (ANIb = 97.39%, over 92.18% of the genomes; Additional file 3), which also supports their classification into the same species (as an ANI of 95% has been proposed as species cut-off [38]). On the other hand the genomes of strains FSL R7-269 and FSL R7-277 (both classified as *Paenibacillus* spp., according to 16S rRNA (see Table 1)) show an ANIb of 92.54% (over 72.88% of genome; Additional file 3), suggesting that these two strains represent separate species despite a 1223/1225 (> 94.99%) nucleotide 16S rRNA identity. This finding is consistent with previous reports that 16S rRNA sequence divergence statistics show less resolution for bacterial species designation than the ANIb values [38].

***Paenibacillus* and *Bacillus* genomes carry a diversity of antimicrobial resistance genes**

Characterization of phenotypic antimicrobial resistance patterns among the strains sequenced here (Table 2) found resistance to at least one antibiotic in all but two strains. The most common resistance pattern (lincomycin-penicillin) was found in two *Bacillus* and two *Paenibacillus* isolates (Table 2); in addition, one strain showed resistance to lincomycin-streptomycin, two strains showed resistant to streptomycin only, and one strain showed resistance to lincomycin only (Table 2). While this is, to our knowledge, the first report of antimicrobial resistant *Bacillus* and *Paenibacillus* isolates from milk,

previous studies have reported antimicrobial resistance in disease associated *Paenibacillus* isolates, including (i) resistance to metronidazole [30], tetracycline [39], and glycopeptides [40] in *P. larvae*, which is associated with disease in honeybees and (ii) multi-drug resistance (to glycopeptide, beta-lactams, aminoglycosides, macrolides and lincosomides) in a *Paenibacillus* isolated from fertile soil in India [41]. Similarly, antimicrobial resistant *Bacillus* isolates have previously been isolated from a variety of clinical [42-44] and non-clinical [45] sources.

Analyses of the sequenced genomes also identified putative antimicrobial resistance genes, in both pan-susceptible and resistant *Bacillus* and *Paenibacillus* isolates (Additional file 4). Most genomes contained more than 20 putative antimicrobial resistance genes (including genes putatively encoding heavy metal resistance characteristics; see Additional file 4). For example, we identified, in the genomes sequenced here, genes putatively encoding resistance to β -lactams (e.g. metallo- β -lactamases and TEM β -lactamase), aminoglycosides (e.g. aminoglycoside 3'phosphotransferase), chloramphenicol (e.g. chloramphenicol acetyltransferase), copper (e.g. CopC, and PcoD), vancomycin (e.g. VanZ), arsenic (e.g. Acr3), tellurite (e.g. TerD), and quaternary ammonium (e.g. EmrE) as well as genes encoding a number of efflux pumps (e.g. EmrB/QacA, Bcr/CflA). Antibiotic resistance genes have previously been reported in *Paenibacillus* from non-dairy sources [39,46], including reports of resistance genes similar to those found here in *P. vortex* and *P. larvae* genomes [26,47]. Similarly, presence of antimicrobial resistance genes appears to be common in *Bacillus* spp. [42,45,48].

Interestingly, the presence of genes putatively encoding resistance to a given antibiotic did not always correlate to the resistance phenotypes of the respective

Table 2 Phenotypic characterization of strains used in this study

Strain	Species	Relative growth, at day 21 (6°C) ¹	Relative proteolysis on Skim Milk Agar (24 h 32°C)	β -gal activity	Antimicrobial resistance pattern ³
FSL R5-860	<i>Bacillus</i> sp.	-	+	-	Lincomycin, Penicillin
FSL H7-687	<i>Bacillus weihenstephanensis</i>	++	+	-	Lincomycin, Penicillin
FSL R5-213	<i>Viridibacillus arenosi</i>	++	- ²	-	Lincomycin
FSL H8-237	<i>Paenibacillus odorifer</i>	++	+	+	Not tested ⁴
FSL R7-277	<i>Paenibacillus</i> sp.	+	+	+	Streptomycin
FSL R7-269	<i>Paenibacillus</i> sp.	+	+	+	Streptomycin
FSL R5-192	<i>Paenibacillus amylolyticus</i>	-	+	+	Lincomycin, Streptomycin
FSL H7-689	<i>Paenibacillus amylolyticus</i>	++	+	+	Pansusceptible
FSL R5-808	<i>Paenibacillus glucanolyticus</i>	-	-	+	Lincomycin, Penicillin
FSL H8-457	<i>Paenibacillus lautus</i>	-	-	+	Lincomycin, Penicillin

¹Relative growth in skim milk broth, plated on BHI at day 21 of shelf life (stored at 6°C): (-) <1 log cfu/mL, (+) 1-6 log cfu/mL, (++) >6 log cfu/mL.

²Isolate produced pinkish coloration of Skim Milk Agar.

³Isolates were tested for susceptibility against 16 antimicrobials included in the Gram positive NARMS panel (see Methods).

⁴Strain was considered too fastidious for susceptible testing.

isolates. For example, while genes encoding β -lactamases were detected in all 10 genomes, only 4 strains showed penicillin resistance (Table 2). Similarly, while the lincomycin resistance operon *lmrAB* was detected in all 10 genomes, only 6 strains showed phenotypic resistance to lincomycin. An alignment of *lmrAB* showed two clades, one that contains only resistant strains and one that contains resistant and susceptible strains (Additional file 5). These observations could be interpreted as being consistent with previous reports that proposed that, in *Paenibacillus* and *Bacillus* isolates, these “antimicrobial resistance genes”, are not necessarily involved in the resistance against anthropogenic antibiotics, but rather encode “defense mechanisms” that allow the bacteria carrying these genes to compete against antagonistic bacteria [26,28]. This hypothesis is also supported by studies that suggest that expression of resistance genes might be tightly regulated in environmental non-pathogenic bacteria and that these genes may not necessarily be expressed under laboratory conditions that are appropriate to evaluate antimicrobial resistance phenotypes in bacterial pathogens. For example, the *Bacillus subtilis* lincomycin resistance operon *lmrAB* has been shown to be dually regulated in response to flavonoids, rather than in response to lincomycin [49].

A number of bacteriocins and pathways for biosynthesis of non-ribosomal peptide antibiotics are encoded in the genomes of dairy associated *Paenibacillus* and *Bacillus*

Paenibacillus spp. have previously been reported to produce certain antimicrobials, like bacteriocins and non-ribosomal peptide antibiotics (NRPA) [18,50-52]. Here we investigated the presence of genes encoding bacteriocins and NRPA biosynthesis in the sequenced isolates, which could represent potential source for GRAS antimicrobials. Using HMM searches, we identified genes encoding putative class II bacteriocins in four of the genomes, including one *Bacillus* strain (FSL H7-687) and three *Paenibacillus* strains (FSL H8-237, FSL R5-192, and FSL H7-689) (Table 3). All of these bacteriocin genes were found in operons that encoded both the bacteriocins and the putative corresponding bacteriocin transporters [53,54]. In the genome of *B. weihenstephanensis* FSL H7-687 we specifically identified a bacteriocin operon that was annotated as encoding a bacteriocin cerein 7B precursor, an ABC-type bacteriocin exporter, and a bacteriocin secretion accessory protein. These proteins present an average of 87% of amino acid (aa) identity to the same operon in *B. cereus* (Table 3). In two of the *Paenibacillus* genomes (FSL H7-689 and FSL R5-192) we identified an operon annotated as encoding the lantibiotic mersacidin biosynthesis protein LanM and the corresponding transporter protein (Table 3). These proteins showed an average aa identity

(across both proteins) of 38% (FSL R5-192) and 40% (FSL H7-689) with the corresponding *B. cereus* proteins. In addition, we identified, in the genome of *Paenibacillus* FSL H8-237, another class II bacteriocin operon that was annotated as encoding a bacteriocin with a double-glycine leader peptide, a secretory protein, and an ABC-type bacteriocin transporter. The proteins encoded in this operon present 34% aa identity (across all 3 proteins) with a similar system in *Anoxybacillus flavithermus*. To search for unusual or un-annotated bacteriocins we conducted a search with BAGEL3 [55]. This analysis identified additional putative bacteriocins of class I (i.e., lanthipeptide, lassopeptide, and sactipeptide bacteriocins) in four genomes (FSL H7-687, FSL R7-277, FSL R7-269, and FSL R5-192), as well as two class II bacteriocins (i.e., UvB and un-classified class II bacteriocin) in two genomes (FSL R5-860 and FSL H8-237) (Table 3). While some of the bacteriocins identified show a high level of homology with previously identified bacteriocins, most of the bacteriocins identified in the *Paenibacillus* genomes show low level homology to previously identified bacteriocins, indicating that they may represent novel bacteriocins.

Genes encoding putative NRPA biosynthesis pathways were identified in the two *Bacillus* and five *Paenibacillus* strains (Table 3). In each of the two *Bacillus* strains, we identified two highly similar operons, one annotated as encoding proteins involved in gramicidin biosynthesis and one annotated as encoding proteins involved in biosynthesis of a putative, uncharacterized, antibiotic. The proteins encoded by these operons showed aa identities of >97% (across all proteins in a given operon) to the corresponding proteins in *B. cereus* (Table 3). Similarly, a number of different operons encoding proteins involved in the synthesis of NRPA were detected in five *Paenibacillus* genomes; *Paenibacillus* strains FSL R7-277, FSL R7-269, FSL R5-192, FSL R5-808, and FSL H7-689 were annotated as encoding two, three, three, one, and five NRPA biosynthesis operons, respectively. While NRPA have previously been identified in several *Paenibacillus* spp. [28,56,57], *Paenibacillus* strains without NRPA have also been described [29]. The operons identified here were annotated as encoding proteins required for the biosynthesis of NRPA similar to Mycosubtilin, Musaricidin, Gramicidin S, Cereulide, Iturin, Bacitracin, and Lichenysin (Table 3). Interestingly, many of the proteins encoding these putative NRPA biosynthesis pathways showed low levels of protein identity (e.g. 32%; see Table 3) to corresponding proteins in *Bacillus* genomes. This suggests that many of these operons may encode proteins that facilitate the biosynthesis of novel NRPA, which could potentially represent novel antimicrobials that inhibit dairy associated bacteria. While it is possible that these antimicrobials would be granted GRAS status

Table 3 Antimicrobial systems detected using HMM searches against the sequenced genomes

Strain (FSL) ¹	Species	Antimicrobial system ³	Locus tag	pBLAST identity in % ⁶ [species = GenBank acc. number]
Bacteriocins ²				
R5-860	<i>B. sp.</i>	Bacteriocin UvB ⁴	C175_07686 to C175_07731	100% [<i>B. thuringiensis</i> = WP_003302295]
H7-687	<i>B. weihenstephanensis</i>	Bacteriocin cerein 7B	C174_01754 to C174_01764	87.0% [<i>B. cereus</i> = CAJ32354.1]
		Lanthipeptide_class_1 ⁴	C174_05773 to C174_05843	100% [<i>B. cereus</i> = WP_002128181.1]
		Lasso_peptide ⁴	C174_07587 to C174_07617	100% [<i>B. cereus</i> = WP_002128586.1]
R5-213	<i>V. arenosi</i>	-	-	-
H8-237	<i>P. odorifer</i>	Putative bacteriocins with double-glycine leader peptide	C17_06947 to C171_06937	34.3% [<i>A. flavithermus</i> = YP_002316950.1]
		Class II ⁴	C171_31256	82% [<i>Paenibacillus</i> sp. = WP_018755215.1]
R7-277	<i>P. sp.</i>	Lasso_peptide ⁴	C173_26127 to C173_26182	78% [<i>Paenibacillus</i> sp. = WP_019913509.1]
		Sactipeptides ⁴	C173_26762 to C173_26742	93% [<i>Paenibacillus</i> sp. = WP_017690652.1]
R7-269	<i>P. sp.</i>	Lasso_peptide ⁴	C162_11816 to C162_11871	78% [<i>Paenibacillus</i> sp. = WP_019913509.1]
		Sactipeptides ⁴	C162_32049 to C162_32089	48% [<i>Dorea</i> sp. = WP_005337554.1]
R5-192	<i>P. amylolyticus</i>	Lantibiotic mersacidin	C161_23334 to C161_23344	38.0% [<i>B. cereus</i> = ZP_17590371]
		Lasso_peptide ⁴	C161_09323 to C161_09363	56% [<i>Paenibacillus</i> sp. = YP_003009997.1]
		Sactipeptides ⁴	C161_03994 to C161_03969	49% [<i>Paenibacillus</i> sp. = WP_016818495.1]
H7-689	<i>P. amylolyticus</i>	Lantibiotic mersacidin	C170_27843 to C170_27833	40.0% [<i>B. cereus</i> = ZP_17590371]
R5-808	<i>P. glucanolyticus</i>	-	-	-
H8-457	<i>P. lautus</i>	-	-	-
Non ribosomal peptide antibiotics				
R5-860	<i>B. sp.</i>	uncharacterized antibiotic ⁵	C175_22452 to C175_22455	99.8% [<i>B. cereus</i> = NP_832072.1]
		Gramicidin	C175_21698 to C175_21708	98.0% [<i>B. cereus</i> = ZP_17623965.1]
H7-687	<i>B. weihenstephanensis</i>	Uncharacterized antibiotic ⁵	C174_09897 to C174_09917	99.0% [<i>B. weihenstephanensis</i> = YP_001645029]
		Gramicidin	C174_03013 to C174_03023	98.0% [<i>B. cereus</i> = ZP_17640177.1]
R5-213	<i>V. arenosi</i>	-	-	-
H8-237	<i>P. odorifer</i>	-	-	-
R7-277	<i>P. sp.</i>	Mycosubtilin	C173_20341 to C173_20411	33.0% [<i>B. laterosporus</i> = ZP_16304742.1]
		Cereulide	C173_11825 to C173_11870	32.3% [<i>N. punctiforme</i> = YP_001866562.1]
R7-269	<i>P. sp.</i>	Uncharacterized antibiotic ⁵	C162_12938 to C163_12948	37.3% [<i>N. punctiforme</i> = YP_001866562.1]
		Gramicidin	C162_20656 to C162_20696	40.8% [<i>A. cellulolyticus</i> = ZP_09464526]
		Fusaricidin	C162_28814 to C162_28839	62.5% [<i>B. thuringiensis</i> = ZP_04087647.1]
R5-192	<i>P. amylolyticus</i>	Gramicidin/polymyxin	C161_02905 to C161_02925	87.2% [<i>P. polymyxa</i> = YP_003872502.1]
		Mycosubtilin	C161_22119 to C161_22129	53.5% [<i>B. atrophaeus</i> = YP_003973351.1]
		Iturin	C161_04861 to C161_04866	34.0% [<i>P. polymyxa</i> = YP_005960904.1]

Table 3 Antimicrobial systems detected using HMM searches against the sequenced genomes (Continued)

H7-689	<i>P. amylolyticus</i>	Gramicidin	C170_23095 to C170_23105	36.3% [<i>P. curdlanolyticus</i> = ZP_07387013.1]
		Gramicidin/polymyxin	C170_04453 to C170_04473	90.2% [<i>P. polymyxa</i> = YP_003872502.1]
		Mycosubtilin	C170_20065 to C170_20070	53.5% [<i>B. atrophaeus</i> = YP_003973351.1]
		Iturin	C170_00119 to C170_00129	38.0% [<i>P. polymyxa</i> = YP_005960904.1]
		Gramicidin S	C170_21789	91.0% [<i>P. polymyxa</i> = AEZ51516]
R5-808	<i>P. glucanolyticus</i>	Gramicidin	C169_08418 to C169_08438	77.2% [<i>P. vortex</i> = ZP_07900219.1]
H8-457	<i>P. lautus</i>	-	-	-

¹The full strain designation includes the prefix FSL, e.g., FSL R5-860.

²Bacteriocin genes identified are predicted to encode class I and II bacteriocins.

³The designation assigned to a given putative antimicrobial system represents the designation associated with the previously annotated system that showed the highest similarity to the query sequence, or the designation predicted by Bagel3.

⁴Identified by BAGEL3.

⁵The designation "uncharacterized antibiotic" was given to ORFs that contain domains involved in peptide antibiotic synthesis.

⁶Average protein BLAST identity for all proteins encoded in a given operon.

(as there is evidence for consumption of these spoilage organisms by a considerable number of consumers), additional scientific safety studies may be needed, particularly if efficacy against relevant target organisms is shown.

Both *Paenibacillus* and *Bacillus* genomes encode proteolytic systems with a putative role in casein breakdown

Phenotypically, both *Bacillus* isolates sequenced here as well as five of the seven *Paenibacillus* isolates showed proteolysis (on skim milk agar at 32°C). BLAST searches of all 10 genomes sequenced here were thus performed to screen for the presence of genes encoding proteolytic systems and proteins previously reported as facilitating casein breakdown in lactic acid bacteria [58]. Using protein sequences for appropriate cell-wall proteinases (Prt), peptide transporters (e.g. oligo- di-tri-peptide transporters), and 13 different peptidases (e.g. PepP, PepA, PepE) [59] as query sequences, we identified cell-wall proteinases as well as oligo-peptide (OppABCD) and di-peptide (DppABCD) transporters in all 10 genomes (Additional file 6). Genes encoding Tri-peptide transporters (DtpT) were found in four genomes (FSL H8-237, FSL H7-687, FSL R5-860, and FSL R5-213), including one strain that did not show proteolytic activity in the phenotypic assay (strain FSL R5-213, see Table 2 and Additional file 6). We also identified genes encoding peptidases in all 10 genomes; the number of peptidase encoding genes ranged from 15 (in FSL R7-269) to 22 (in FSL H7-687 and FSL R5-860) peptidase genes per genome; genes encoding PepD, PepO and PepE/PepC, which had previously been shown to contribute to growth in milk in lactic acid bacteria [60], were not found in any of the 10 genomes (Additional file 6). Overall, even though some of the 10 strains characterized here were negative for proteolysis on skim milk agar, all strains appear to encode proteins for all major steps required for casein breakdown in lactic acid bacteria (i.e., cell-wall proteinase activity, peptide transport and intracellular peptidases). Future experimental work will thus be required to dissect the function of the different putative casein hydrolysis-related proteins encoded in the genomes sequenced here.

***Paenibacillus* genomes carry several genes encoding β -galactosidase activity**

Consistent with previous phenotypic studies [5], we found that only the seven *Paenibacillus* isolates expressed β -galactosidase activity; the two *Bacillus* and the *Viridibacillus* isolates characterized here were β -galactosidase negative (Table 2). An HMM search for β -galactosidase domains also identified genes with these types of domains only in the seven *Paenibacillus* genomes,

but not in the *Bacillus* and *Viridibacillus* genomes (Additional file 7). Interestingly, we identified between 9 and 19 genes encoding putative β -galactosidases in each *Paenibacillus* genome (Additional file 7). A maximum likelihood phylogeny of the aa sequences for all β -galactosidases (Figure 2) not only revealed considerable diversity among these proteins, but also allowed us to classify these proteins into 33 β -galactosidase types (β -gal 1 to 33); putative β -galactosidases were assigned the same type if they clustered into the same clade (see Figure 2). While five β -gal types (2, 4, 5, and 11) were found in each of the seven genomes (Figure 2 & Additional file 8); 13 β -gal types (i.e., 1, 6, 7, 10, 12, 13, 15, 17, 18, 19, 21, 29, and 31) were absent in both strains classified into *Paenibacillus* clade III (FSL H8-457, FSL R5-808) and one β -gal types (β -gal17) was only found in clade II strains (Additional file 8). Both *Paenibacillus* clade I strains contained the identical repertoire of β -gal types.

Similarly to our findings, presence of multiple genes encoding β -galactosidases has been previously described for haloarchaeal bacteria [61]. Interestingly, β -galactosidases have also been identified in microorganisms typically found in lactose-free environments [62] and β -galactosidases in these bacteria have been shown to hydrolyze pectin-like plant polysaccharides [62]. For example, a β -d-galactosidase from *Paenibacillus thiaminolyticus* was found to have fucosidase activity [63]. Further experimental work will be needed to define the role of the multiple β -galactosidases reported here in dairy related *Paenibacillus* spp. isolates. For example, the different β -galactosidases may be expressed under different conditions (e.g. different temperatures) or may facilitate use of diverse oligo- and polysaccharides in addition to lactose. In addition, our data suggest that *Paenibacillus* may represent a potential source of GRAS β -galactosidases for industrial applications.

Comparison of the two *Paenibacillus* strains in clade I revealed cold-growth related genomic features only in the strain able to grow at 6°C in skim milk broth

Strains FSL H7-689 and FSL R5-192 are closely related strains of *Paenibacillus*, which both have *rpoB* allelic type 23 (Table 1) [5]. Accordingly, these two strains clustered in the same clade in the phylogenetic analysis (Figure 1) and presented a high genome-wide average nucleotide identity (97.39%; see Additional file 3). However, these strains differed in their abilities to grow at 6°C in skim milk broth (monitored over 21 days); while FSL H7-689 was able to grow under these conditions, FSL R5-192 was not (Table 2). We thus performed a comparative analysis of the genomes for these two strains to identify potential genomic features that may contribute to these distinct phenotypes. This analysis identified 216 and 479 functionally annotated genes

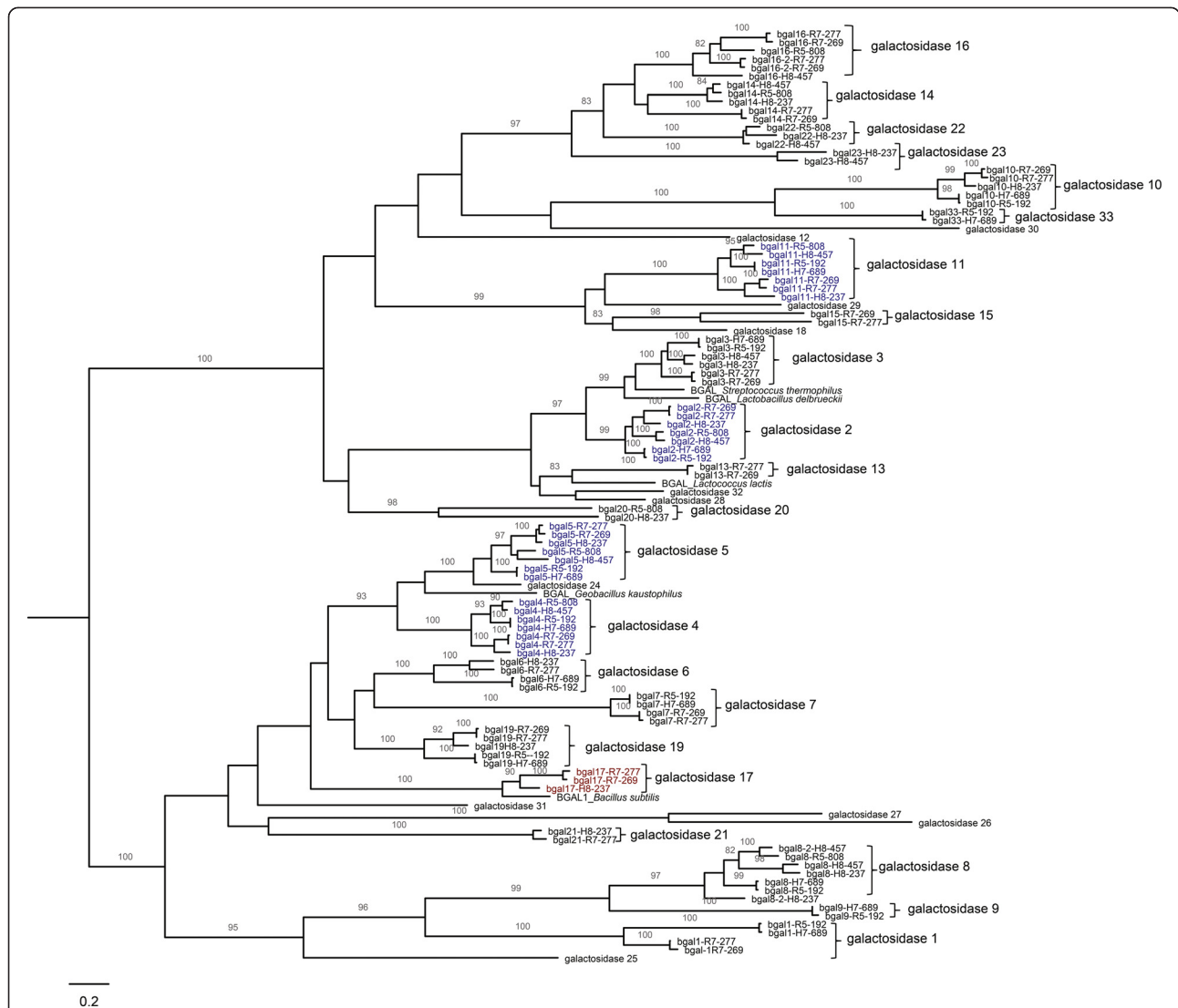


Figure 2 Maximum likelihood phylogeny of the β -galactosidases identified in the *Paenibacillus* genomes. β -galactosidases present in all seven *Paenibacillus* genomes are shown in blue; the β -galactosidase only present in clade II strains is shown in red. Numbers on the branches represent bootstrap values based on 100 replicates. Bootstrap values <80 are not included. Five previously characterized β -galactosidases were included in the analysis, including LacZ of *Lactococcus lactis* (uniprot: Q48727), *Lactobacillus delbrueckii* (uniprot: POC1Y0), and *Streptococcus thermophilus* (uniprot: P23989), YesZ of *Bacillus subtilis* (uniprot: O31529), and BgaB of *Geobacillus kaustophilus* (uniprot: P19668).

unique to strains FSL H7-689 and FSL R5-192, respectively (Additional file 9). Genes only found in FSL H7-689 encode proteins involved in a number of functions including, for example, synthesis of non-ribosomal peptide antibiotics (NRPA), transmembrane transport (e.g. Fe³⁺ transport, drug exporters, sugar transport systems), antibiotic resistance (e.g. an aminoglycoside 3'-phosphotransferase), as well as mobile elements (e.g. transposases and phage related proteins) and a toxin-antitoxin system (see Additional file 9). While genes only found in FSL R5-192 encode many proteins involved in similar functions (e.g. NRPA synthesis, transmembrane transport, antibiotic resistance), more mobile elements were

found only in this strain. For example, we identified two putative prophages that appear complete, and multiple transposons in FSL R5-192 (Additional file 9), while no complete prophages and only two transposons were identified in FSL H7-689. Importantly, prophages in FSL R5-192 do not seem to be interrupting genes related with cold growth. Different stressful conditions (e.g. acidity, osmolarity, high temperature [64,65]), could induce the prophage lytic cycle, and consequently lyse the cells; experimental evidence will be necessary through to determine whether cold-growth could induce the lytic cycle in the prophage identified here, thus preventing cold growth.

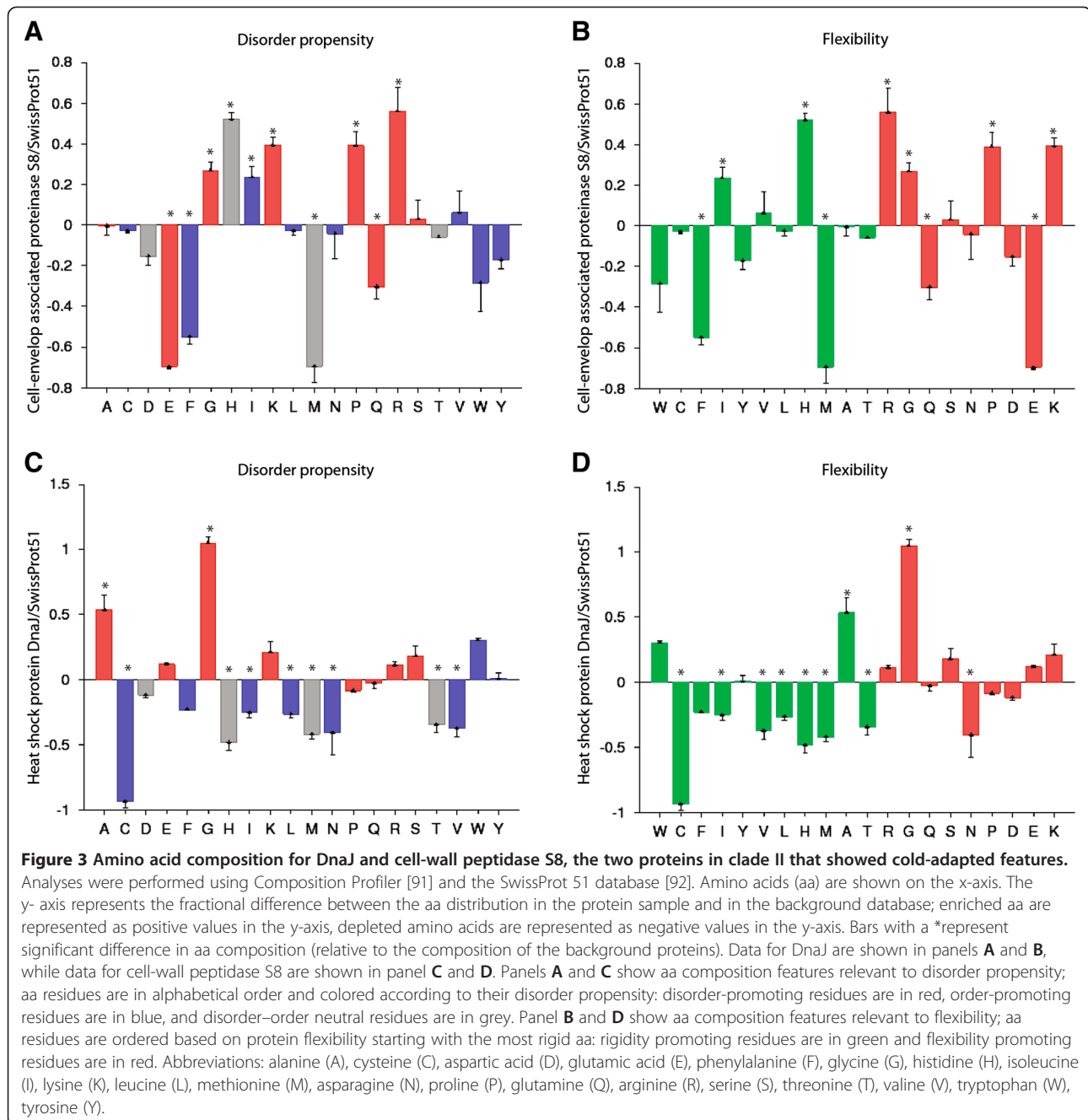
Interestingly, we identified, in the genome of FSL H7-689, one cluster of genes annotated as encoding a DNA methylase, a restriction endonuclease, and a DEAD box helicase (Additional file 9); these genes were absent in FSL R5-192 with upstream and downstream genes located in a single contig, indicating that genes are not missing due to a sequence gap. DEAD box helicases are proteins that participate in a number of functions, including cold adaptation [66-68]. In *Bacillus subtilis*, deletion of DEAD box helicases was found to cause a cold-sensitive phenotype [66] and in *Listeria monocytogenes* deletion mutants of DEAD box helicases were unable to grow at 3°C [69]. Another DEAD box helicase (CshA) was found in both strains of clade I. While the presence of two DEAD box helicases only in the strain able to grow at 6°C in skim milk broth could be one of the genomic features that contributes to the cold growth phenotype of this strain, further experimental work is necessary to elucidate the role of this protein in cold adaptation of FSL H7-689. In addition, novel unrelated DEAD box helicases cannot be excluded as contributing to the ability of some *Paenibacillus* to grow at low temperatures.

Comparison of *Paenibacillus* clade II and clade III genomes revealed a number of genes encoding cold adapted features and lactose utilization in clade II strains

Clade II strains present different phenotypic characteristics than clade III strains; specifically, only clade II strains were able to grow at 6°C in skim milk broth and showed proteolysis on skim milk agar (Table 2). Initial comparative genomic analyses identified 914 and 1,070 orthologs present in only clade II or clade III strains, respectively. To identify genomic features and genes that may be linked to the unique phenotypic characteristics of clade II strains (e.g. cold growth), we performed an enrichment analysis for RAST subcategories among clade II and III genomes. This analysis identified nine RAST subcategories that were overrepresented in the clade II genomes (Additional file 10 and Additional file 11); these subcategories were also significantly overrepresented in a comparisons of “cold adapted *Paenibacillus*” (i.e., clade II and clade I strain FSL H7-689) and “non-cold adapted *Paenibacillus*” (i.e., clade III and clade I strain FSL R5-192) (Additional file 10). While most of these subcategories are involved in energy acquisition or in bacterial growth; some of these subcategories (e.g. “protein and nucleoprotein secretion system, type IV”; “di- and oligo-saccharides”) encode proteins that could facilitate growth of clade II strains in milk. The RAST subcategory “di- and oligo-saccharides” includes genes encoding essential enzymes for lactose and galactose uptake and utilization (Additional file 10). These findings suggest that clade II strains contain unique genomic

features that may facilitate their growth in milk and dairy production associated environments.

We also identified a number of specific genes that were only identified in clade II strains and that encode for properties that could be involved in the ability of these strains to grow in milk at cold temperatures, including genes encoding β -galactosidases (included in the RAST subcategory “di- and oligo-saccharides”) as well as genes encoding (i) peptide transport systems and peptidases and (ii) cold-adaptation related proteins. While we already reported the presence of several proteinases and peptide transport systems in all *Paenibacillus* genomes (in a previous section), some of these proteins appear uniquely encoded in the genomes of clade II strains. For example, two operons for peptide transport (oligopeptide and dipeptide transport systems) as well as several peptidases (cell-wall associated proteinase S8 and oligopeptidase F) were identified specifically in clade II (Additional file 12). Interestingly, these enzymes are associated with the three steps of bacterial casein breakdown, which includes (i) hydrolysis of casein into oligo- and dipeptides by cell envelope proteinases, (ii) peptide-transport-system-facilitated import of these peptides into the cell, and (iii) peptide degradation by intracellular peptidases [58,59]. We further analyzed the amino acid composition of the clade II specific proteins involved in casein hydrolysis to test the hypothesis that these proteins may contain specific structural features associated with cold-adaptation. In cold-adapted enzymes the protein structure is more flexible, which allows for catalysis at cold temperatures [70]; features associated with increased structural flexibility include disorder regions and amino acid bias (e.g. increase of glycine and decrease of acidic residues) [70-73]. Our analyses found that the cell-envelope associated peptidase S8 encoded by clade II strains (i) shows significant ($p = 0.017$) enrichment of disorder-promoting aa residues and (ii) shows enrichment of certain aa associated with cold adaptation (i.e., glycine and lysine) (Figure 3A, B). The peptidases of the S8 family have been found to function at extreme temperatures [74,75] and their stability is enhanced by calcium [74], important features for the growth in milk at refrigeration temperatures. In addition, genes encoding three β -galactosidases (β -gal types 10, 17, and 19) and several glycosyl hydrolases with possible β -galactosidase activity were identified in the genomes of clade II strains, but not in clade III strains (Additional file 12). We, however, did not identify cold-adapted features in any of these proteins. Interestingly, we found though that genes classified into glycosyl hydrolase families that have been shown to specifically have β -galactosidase activity (i.e., families 1, 2, 35 and 42) were over-represented in clade II ($p = 0.044$; overall Fisher’s exact test), suggesting that clade II may have



unique carbohydrate metabolic capabilities (and specifically capabilities to hydrolyze lactose) that may facilitate growth in milk.

We also identified, in clade II strains, genes encoding several cold growth associated proteins that have been previously identified in other bacteria (Additional file 12), including (i) the low temperature requirement protein A (LtrA), which has been found to be important for growth at 4°C in *L. monocytogenes* [76]; (ii) chaperone DnaJ, which has been found to be up-regulated in *E. coli* and *Lactococcus* at cold temperatures [77,78], (iii)

glutamate transport ATP-binding protein, which has been found to be up-regulated in *B. subtilis* at cold temperatures, and (iv) D-alanyl-D-alanine carboxypeptidase, which could decrease the cell wall resistance at cold temperature [70]. Among these proteins, DnaJ was identified to have cold-adapted features (flexibility and disorder promoting aa were both significantly enriched [$p < 0.001$]; see Figure 3C, D). Interestingly, two DnaJ-encoding genes were identified in clade II strains, one that is present in both the 3 clade II and the 2 clade III strains, and one that is present only in the 3 clade II

strains and that encodes a DnaJ with cold-adapted features. In addition, in clade II strains we identified four aspartate aminotransferases, enzymes reported to be essential for growth of *Lactococcus lactis* in milk [79], as well as a number of proteins with no clear association with cold-adaptation (e.g. tellurite resistance operon, nitrogen fixation operon, and lactate dehydrogenase). Overall, we identified a number of proteins that could facilitate the growth of strains in clade II in milk at 6°C. Mapping of clade III genome sequencing reads against the clade II genome contigs that contained the clade II specific genes mentioned above (e.g. peptidase S8, β -galactosidases) confirmed that these genes were absent from the clade III genomes (see Additional file 12). To further understand the features associated with the cold tolerant phenotype of these strains; additional experimental data, including expression studies, are needed. The genes identified here do represent potential targets for these types of studies.

Conclusions

Genomic analysis of selected sporeforming Bacillales isolated predominantly from milk allowed for identification of a number of genes encoding potentially relevant GRAS proteins such as potentially novel bacteriocins and NRPA. Further comparative genomic analyses, focusing on *Paenibacillus* clades and strains that differed in their ability to grow in milk at refrigeration temperatures, identified a number of genomic features that likely allow some strains to grow at refrigeration temperatures and to break down milk associated proteins and carbohydrates at these temperatures. This study thus demonstrates how a genomic approach can be used to develop an improved understanding of microbial food spoilage, which will facilitate subsequent identification of genomic targets (e.g. β -galactosidases or peptidases only present in clade II) that can be used to control food spoilage or rapidly screen foods or raw materials for presence of bacteria with specific spoilage related genomic features.

Methods

Isolate selection

Isolates were selected for genome sequencing from a large collection of dairy-associated sporeforming Bacillales (>1,200) [5]. The ten isolates used here for genome sequencing (Table 1) were selected to (i) represent different genera of dairy associated sporeforming bacteria; (ii) represent various phenotypic and metabolic characteristics (Table 2) and (iii) represent the predominant *rpoB* allelic types associated with the dairy production system, using previously reported data [5]. We specifically selected the isolates to present strains that differ in their ability to (i) growth in skim milk broth at 6°C (monitored over 21 days), (ii) show proteolytic activity on skim

milk agar, and (iii) show β -galactosidase hydrolysis on BHI X-gal. This approach was used to allow for comparative analyses to explore genomic features associated with the ability, of sporeforming Bacillales, to grow in milk at refrigeration temperatures.

Illumina sequencing, assembly, and annotation

Genomic DNA from the selected isolates was prepared using phenol/chloroform DNA extraction [80]. Genomic DNA was quantified on the NanoDrop (NanoDrop products, Wilmington, DE) and standardized to 200 ng/ml. Isolate identity was confirmed via *rpoB* PCR prior to sequencing [5]. Genome sequencing was performed on the Illumina HiSeq 2000 at the Cornell University Life Sciences Core Laboratories Center (Ithaca, NY). Paired-end 2x100 bp reads were obtained and assembled *de novo* using Velvet and the VelvetOptimiser [81]. Assembled contigs were filtered to only include contigs >199 bp. Genome annotations were performed using a combination of RAST [82] and the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) [83]. GenBank accession numbers are available in Table 1.

Whole genome comparison and phylogenetic analysis

Previously sequenced closed genomes representing *Bacillus* spp. (n = 9) and *Paenibacillus* spp. (n = 6) were retrieved from NCBI (see Additional file 1) and used for comparative analyses. Genome alignments were performed with MAUVE [84], mapping alignments of homologous genes to a reference genome was conducted with GeneWiz 0.94 [85]. Phylogenetic analysis was performed using maximum likelihood methods implemented in RaxML version 7.0.4 [86] with 1,000 rapid bootstrap replicates, using amino acid sequences for 31 marker genes [36]. For this analysis, we selected 52 genomes representing sequenced Firmicutes (Additional file 1); genomes representing *Clostridium* spp. were selected to root the phylogenetic tree. Average nucleotide identity was calculated with JSpecies v1.2.1 [38].

Phenotypic characterization

Isolates were stored at -80°C (15% glycerol, 85% BHI culture) and were grown on Brain-Heart Infusion (BHI) agar plates, which were incubated at 32°C for 24 h. β -galactosidase activity of isolates was assessed as previously described [5]. Briefly, isolates were streaked onto two BHI agar plates, one with and one without 100 μ l of a 40 μ g/ml solution of bromo-chloro-indolyl-galactopyranoside (X-Gal).

To assess cold growth, a single colony for each given isolate was first inoculated into BHI broth and grown for 24 h at 32°C. Cultures were diluted and inoculated into skim milk broth (BD, Franklin Lakes, NJ) to reach an initial inoculum of approx. 10^2 to 10^3 (plating showed that bacterial

numbers ranged from $1.3\text{--}7.4 \times 10^2$ CFU/ml). The inoculated skim milk broth was held at 6°C for 21 days and bacterial numbers were enumerated every 7 days by plating in appropriate dilutions on plate count agar (PCA; EMD Millipore, Swedesboro, NJ) using an Autoplate 4000 (Spiral Biotech, Norwood, MA). Bacterial colonies were counted after 24–48 h (depending on colony growth morphology) using Q-count.

Antimicrobial resistance determination

The National Antimicrobial Resistant Monitoring System (NARMS) protocol for Gram positive bacteria was used to test for antimicrobial resistance. Minimal inhibitory concentrations (MIC) were determined using the Sensititre system (TREK Diagnostic Systems, Cleveland, OH), for the following 16 antimicrobials: chloramphenicol, ciprofloxacin, daptomycin, erythromycin, gentamicin, kanamycin, lincomycin, linezolid, nitrofurantoin, penicillin, streptomycin, synergid, tetracycline, tigecycline, tylosin, and vancomycin. The Clinical and Laboratory Standards Institute (CLSI) cut-offs, for bacterial isolates from animal origin, were used to interpret MIC values.

Identification of proteins related to key phenotypes

Hidden Markov Models (HMM) for (i) relevant proteins (i.e., β -galactosidases, bacteriocins, glycosyl hydrolases) and (ii) proteins previously shown to be involved in relevant phenotypes (i.e., synthesis of non-ribosomal peptide antibiotics) were obtained from Pfam 26.0 protein families' database [87] (see Additional file 13 for Pfam models used for the searches). Searches for protein sequence similarities were conducted with HMMER3 [88]. We also conducted an additional search with the bacteriocin mining tool implemented in BAGEL3 [55]. Additional analyses were performed to identify, in the sequenced genomes, genes encoding proteolytic systems; briefly, proteins previously described to be involved in casein breakdown in lactic acid bacteria [59] were obtained from UniProt (<http://www.uniprot.org/>) and used as query sequences for BLAST searches against the 10 genomes sequenced. In addition, text searches of the genome annotations were performed to identify genes annotated as putative antimicrobial resistance genes.

Ortholog comparisons

To identify genes that are unique to specific strains or clades (e.g. clades that only contain isolates with ability to grow at 6°C in skim milk broth), BLAST all against all searches were conducted with OrthoMCL v. 1.4 using the default settings [89]. For orthologs that were identified as unique to a given strain or clade, pBLAST searches against the GenBank database were performed to confirm the initial gene annotations. To confirm that

genes classified as only present in clade II were absent from the clade III genomes, we also mapped the genome sequencing reads for the clade III genomes against clade II genome contigs that contained clade II specific genes. This read mapping was performed with SMALT version 0.7.5 [90]; using a word length of 17 and a sampling step size of 1.

Analysis of amino acid composition

Amino acid sequences of selected proteins were analyzed for structural features related with cold-adapted enzymes [70]. Specifically, amino acids bias, disorder promoting regions and flexibility were analyzed using Composition Profiler [91] and the SwissProt 51 database [92] as a reference; 10,000 bootstrap iterations were performed.

Enrichment analysis for RAST subcategories

Genes in all seven *Paenibacillus* were classified, by RAST, in 102 gene subcategories [82]. We assessed whether genes in each subcategory and genes identified by the glycosyl hydrolase HMMs were over-represented in a group of interest (e.g. clade II genomes) using the one-sided Fisher's exact test implemented in R (version 2.13.0).

Availability of supporting data

The sequence data supporting the results of this article are available on GenBank under the following accession numbers [GenBank: ASPZ000000000, ASPY000000000, ASQ A000000000, ASPV000000000, ASPX000000000, ASPS000000000, ASPR000000000, ASPU000000000, ASPT000000000, ASP W000000000].

Additional files

Additional file 1: List of closed genomes used for phylogenetic analysis. PDF file containing a table with the list of isolates representing *Bacillus*, *Paenibacillus*, and selected Firmicutes.

Additional file 2: Circular representation of BLAST comparisons for the *Paenibacillus* and *Bacillus* genomes. A) BLAST comparison for all seven *Paenibacillus* using *P. lautus* Y4.12MC10 as reference genome. The inner circles in red and blue represent CDS in leading and lagging strands of *P. lautus* Y4.12MC10. In red from inside to outside are the genomes of FSL H8-457, FSL R5-808, FSL H8-237, in blue are the genomes of FSL R7-277 and FSL R7-269, and in green are the genomes of FSL H7-689 and FSL R5-192. B) BLAST comparison for the *Bacillus* and *Viridibacillus* genomes. The inner circles in red and blue represent CDS in leading and lagging strands of *B. weihenstephanensis* KBAB4. In red from inside to outside are the genomes of FSL H7-687, FSL R5-860, and FSL R5-213. Legends for BLAST identity percentages are found underneath for the respective figures.

Additional file 3: BLAST average nucleotide identity (ANIb), determined with the BLAST algorithm implemented in Jspecies, for *Paenibacillus* spp. PDF file containing a table with the nucleotide identity values.

Additional file 4: List of putative antimicrobial and heavy metal resistance genes identified. PDF file containing a table with the antimicrobial resistance genes identified in a given genome.

Additional file 5: Neighbor joining tree of the lincomycin resistant operon *ImrAB*. PDF file containing the tree. In red are the strains that showed resistance. In parenthesis are the minimal inhibitory concentrations (MIC) determined for lincomycin.

Additional file 6: Number of putative casein breakdown associated proteins identified in the 10 genomes. PDF file containing the number of proteins identified in the genomes sequenced here.

Additional file 7: Table containing the number of proteins representing putative β -galactosidases identified in the 10 genomes. PDF file containing a table that details the numbers and types of β -galactosidases identified in the 10 genomes sequenced here.

Additional file 8: Heatmap of the distribution of β -galactosidases in the seven *Paenibacillus* genomes. Blue indicates absence, green indicates presence, and dark green indicates two copies of a given β -galactosidase.

Additional file 9: Comparison of unique genes identified in FSL H7-689 (grows at 6°C) and FSL R5-192 (does not grow at 6°C). PDF file containing a table that lists the proteins that were unique to either FSL H7-689 or FSL R5-192; hypothetical proteins are not included in the list.

Additional file 10: Subcategories over-represented in clade II and in cold-adapted strains. PDF file containing a table that details the subcategories that are over-represented in clade II as compared to clade III, as well as in cold-adapted versus non-cold adapted strains.

Additional file 11: Number of genes in RAST subcategories shown to be significantly over-represented in clade II genomes as compared to clade III genomes. For each RAST subcategory, gene numbers were color coded as representing (i) high number of genes in a specific subcategory (light gray) and (2) low number of genes in a given subcategory (dark gray); the cut-off between these categories were arbitrary. The two non cold-adapted clade III strains are shown above the three cold-adapted clade II strains.

Additional file 12: Clade II-specific genes that are putatively associated with the ability to grow in milk at refrigeration temperature. PDF file containing clade II-specific genes.

Additional file 13: PFAM Models used in identification of genes encodings bacteriocins, non-ribosomal peptide antibiotics, β -galactosidases, and glycosyl hydrolases. PDF file containing the accession numbers for the models used.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AA, MLR, NM, and RI: designed experiments; AA, NM and RI: conducted experiments; HdB and AA: participated in the sequence assembly; HdB, RHO, and AIMS: participated in genome analysis and interpretation; AIMS, MW, and KJB: wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the New York State Dairy Promotion Advisory Board (through the New York State Department of Agriculture), representing New York State dairy farmers committed to producing high-quality milk. The funding source had no role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the paper for publication.

Received: 30 July 2013 Accepted: 8 January 2014
Published: 14 January 2014

References

1. United States Census Bureau. Department of Commerce: **Global population profile**. 2002. <http://www.census.gov/population/international/files/wp02/wp-02003.pdf>.
2. **Dairy farming today**. <http://www.dairyfarmingtoday.org/Learn-More/FactsandFigures/Pages/StateStatistics.aspx>.
3. U.S. Environmental protection agency: **Dairy in the US**. <http://www.epa.gov/agriculture/ag101/dairy.html>.
4. International Dairy Foods Association: *Dairy facts*. Washington DC: International Dairy Foods Association; 2010:66–76.
5. Ivy RA, Ranieri ML, Martin NH, den Bakker HC, Xavier BM, Wiedmann M, Boor KJ: **Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing**. *Appl Environ Microbiol* 2012, **78**:1853–1864.
6. Cleto S, Matos S, Kluskens L, Vieira MJ: **Characterization of contaminants from a sanitized milk processing plant**. *PLoS One* 2012, **7**:e40189.
7. Ranieri ML, Boor KJ: **Short communication: bacterial ecology of high-temperature, short-time pasteurized milk processed in the United States**. *J Dairy Sci* 2009, **92**:4833–4840.
8. Ranieri ML, Huck JR, Sonnen M, Barbano DM, Boor KJ: **High temperature, short time pasteurization temperatures inversely affect bacterial numbers during refrigerated storage of pasteurized fluid milk**. *J Dairy Sci* 2009, **92**:4823–4832.
9. Huck JR, Sonnen M, Boor KJ: **Tracking heat-resistant, cold-thriving fluid milk spoilage bacteria from farm to packaged product**. *J Dairy Sci* 2008, **91**:1218–1228.
10. Huck JR, Woodcock NH, Ralyea RD, Boor KJ: **Molecular subtyping and characterization of psychrotolerant endospore-forming bacteria in two New York State fluid milk processing systems**. *J Food Prot* 2007, **70**:2354–2364.
11. Ranieri ML, Ivy RA, Mitchell WR, Call E, Masiello SN, Wiedmann M, Boor KJ: **Real-time PCR detection of *Paenibacillus* spp. in raw milk to predict shelf life performance of pasteurized fluid milk products**. *Appl Environ Microbiol* 2012, **78**:5855–5863.
12. Huck JR, Hammond BH, Murphy SC, Woodcock NH, Boor KJ: **Tracking spore-forming bacterial contaminants in fluid milk-processing systems**. *J Dairy Sci* 2007, **90**:4872–4883.
13. Klee SR, Brzuszkiewicz EB, Nattermann H, Brüggemann H, Dupke S, Wollherr A, Franz T, Pauli G, Appel B, Liebl W, Couacy-Hymann E, Boesch C, Meyer FD, Leendertz FH, Ellerbrok H, Gottschalk G, Grunow R, Liesegang H: **The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids**. *PLoS One* 2010, **5**:e10986.
14. Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, Willner K, Nolan N, Lentz S, Thomason MK, Sozhamannan S, Mateczun AJ, Du L, Read TD: **Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis***. *Genome Res* 2012, **22**:1512–1524.
15. Lapidus A, Goltzman E, Auger S, Galleron N, Ségurens B, Dossat C, Land ML, Broussolle V, Brillard J, Guinebretiere MH, Sanchis V, Nguen-The C, Lereclus D, Richardson P, Wincker P, Weissenbach J, Ehrlich SD, Sorokin A: **Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity**. *Chem Biol Interac* 2008, **171**:236–249.
16. Adlakhia N, Ritturaj Kushwaha H, Rajagopal R, Yazdani SS: **Draft genome sequence of the *paenibacillus* sp. Strain ICGEB2008 (MTCC 5639) isolated from the Gut of *helicoverpa armigera***. *Genome Announc* 2013, **1**.
17. Beneduzi A, Campos S, Ambrosini A, De Souza R, Granada C, Costa P, Arruda L, Moreira F, Vargas LK, Weiss V, Tieppo E, Faoro H, De Souza EM, Pedrosa FO, Passaglia LMP: **Genome sequence of the diazotrophic gram-positive rhizobacterium *Paenibacillus riograndensis* SBR5T**. *J Bacteriol* 2011, **193**:6391–6392.
18. Ding R, Li Y, Qian C, Wu X: **Draft genome sequence of *Paenibacillus elgii*B69, a strain with broad antimicrobial activity**. *J Bacteriol* 2011, **193**:4537.
19. Djukic M, Becker D, Poehlein A, Voget S, Daniel R: **Genome sequence of *Paenibacillus alvei* DSM 29, a secondary invader during European fowlbrood outbreaks**. *J Bacteriol* 2012, **194**:6365.
20. Huang E, Guo Y, Yousef AE: **Draft genome sequence of *Paenibacillus* sp. strain OSY-SE, a bacterium producing the novel broad-spectrum lipopeptide antibiotic paenibacterin**. *J Bacteriol* 2012, **194**:6306.

21. Huang E, Yousef AE: **Draft genome sequence of *Paenibacillus polymyxa* OSY-DF, which coproduces a lantibiotic, paenibacillin, and polymyxin E1.** *J Bacteriol* 2012, **194**:4739–4740.
22. Jeong H, Choi SK, Park SY, Kim SH, Park SH: **Draft genome sequence of *Paenibacillus peoriae* strain KCTC 3763 T.** *J Bacteriol* 2012, **194**:1237–1238.
23. Jeong H, Park SY, Chung WH, Kim SH, Kim N, Park SH, Kim JF: **Draft genome sequence of the *Paenibacillus polymyxa* type strain (ATCC 842 T), a plant growth-promoting bacterium.** *J Bacteriol* 2011, **193**:5026–5027.
24. Li NZ, Xia T, Xu YL, Qiu RR, Xiang H, He D, Peng YY: **Genome sequence of *Paenibacillus* sp. strain Aloe-11, an endophytic bacterium with broad antimicrobial activity and intestinal colonization ability.** *J Bacteriol* 2012, **194**:2117–2118.
25. Ma M, Wang Z, Li L, Jiang X, Guan D, Cao F, Chen H, Wang X, Shen D, Du B, Li J: **Complete genome sequence of *Paenibacillus mucilaginosus* 3016, a bacterium functional as microbial fertilizer.** *J Bacteriol* 2012, **194**:2777–2778.
26. Chan QW, Cornman RS, Birol I, Liao NY, Chan SK, Docking TR, Jackman SD, Taylor GA, Jones SJ, De Graaf DC, Evans JD, Foster LJ: **Updated genome assembly and annotation of *Paenibacillus larvae*, the agent of American foulbrood disease of honey bees.** *BMC Genomics* 2011, **12**:450.
27. Chow V, Nong G, St John FJ, Rice JD, Dickstein E, Chertkov O, Bruce D, Dettler C, Brettin T, Han J, Woyke T, Pitluck S, Nolan M, Pati A, Martin J, Copeland A, Land ML, Goodwin L, Jones JB, Ingram LO, Shanmugam KT, Preston JF: **Complete genome sequence of *Paenibacillus* sp. strain JDR-2.** *Stand Genomic Sci* 2012, **6**:1–10.
28. Sirota-Madi A, Olender T, Helman Y, Brainis I, Finkelshtein A, Roth D, Hagai E, Leshkowitz D, Brodsky L, Galatenko V, Nikolaev V, Gutnick DL, Lancet D, Ben-Jacob E: **Genome sequence of the pattern-forming social bacterium *Paenibacillus dendritiformis* C454 chiral morphotype.** *J Bacteriol* 2012, **194**:2127–2128.
29. Mead DA, Lucas S, Copeland A, Lapidus A, Cheng J-F, Bruce DC, Goodwin LA, Pitluck S, Chertkov O, Zhang X, Dettler JC, Han CS, Tapia R, Land M, Hauser LJ, Chang YJ, Kyrpidis NC, Ivanova NN, Ovchinnikova G, Woyke T, Brumm C, Hochstein R, Schoenfeld T, Brumm P: **Complete genome sequence of *Paenibacillus* strain Y4.12MC10, A novel *paenibacillus lautus* strain isolated from Obsidian Hot Spring in Yellowstone National Park.** *Stand Genomic Sci* 2012, **6**:381–400.
30. Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE: **Non-contiguous finished genome sequence and description of *Paenibacillus senegalensis* sp. nov.** *Stand Genomic Sci* 2012, **7**:70–81.
31. Qin X, Evans JD, Aronstein KA, Murray KD, Weinstock GM: **Genome sequences of the honey bee pathogens *Paenibacillus larvae* and *Ascosphaera apis*.** *Insect Mol Biol* 2006, **15**:715–718.
32. McSpadden Gardener BB: **Ecology of *Bacillus* and *Paenibacillus* spp. In agricultural systems.** *Phytopathology* 2004, **94**:1252–1258.
33. Piuri M, Sanchez-Rivas C, Ruzal SM: **A novel antimicrobial activity of a *Paenibacillus polymyxa* strain isolated from regional fermented sausages.** *Lett Appl Microbiol* 1998, **27**:9–13.
34. Gioia J, Yerrapragada S, Qin X, Jiang H, Igboeli OC, Muzny D, Dugan-Rocha S, Ding Y, Hawes A, Liu W, Perez L, Kovar C, Dinh H, Lee S, Nazareth L, Blyth P, Holder M, Buhay C, Tirumalai MR, Liu Y, Dasgupta I, Bokhetache L, Fujita M, Karouia F, Eswara Moorthy P, Siefert J, Uzman A, Buzumbo P, Verma A, Zwiya H, et al: **Paradoxical DNA repair and peroxide resistance gene conservation in *Bacillus pumilus* SAFR-032.** *PLoS One* 2007, **2**:e928.
35. Kim JF, Jeong H, Park SY, Kim SB, Park YK, Choi SK, Ryu CM, Hur CG, Ghim SY, Oh TK, Kim JJ, Park CS, Park SH: **Genome sequence of the polymyxin-producing plant-probiotic rhizobacterium *Paenibacillus polymyxa* E681.** *J Bacteriol* 2010, **192**:6103–6104.
36. Wu M, Scott AJ: **Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2.** *Bioinformatics* 2012, **28**:1033–1034.
37. Ash C, Priest FG, Collins MD: **Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*.** *Antonie Van Leeuwenhoek* 1993–1994, **64**:253–260.
38. Richter M, Rosselló-Móra R: **Shifting the genomic gold standard for the prokaryotic species definition.** *Proc Natl Acad Sci U S A* 2009, **106**:19126–19131.
39. Alippi AM, López AC, Reynaldi FJ, Grasso DH, Aguilar OM: **Evidence for plasmid-mediated tetracycline resistance in *Paenibacillus larvae*, the causal agent of American Foulbrood (AFB) disease in honeybees.** *Vet Microbiol* 2007, **125**:290–303.
40. Guardabassi L, Perichon B, Van Heijenoort J, Blanot D, Courvalin P: **Glycopeptide resistance vanA operons in *Paenibacillus* strains isolated from soil.** *Antimicrob Agents Chemother* 2005, **49**:4227–4233.
41. Pednekar PB, Jain R, Thakur NL, Mahajan GB: **Isolation of multi-drug resistant *Paenibacillus* sp. From fertile soil: an imminent menace of spreading resistance.** *J Life Sci* 2010, **4**:15.
42. Rather MA, Aulakh RS, Gill JPS, Mir AQ, Hassan MN: **Detection and sequencing of plasmid encoded tetracycline resistance determinants (tetA and tetB) from food-borne *Bacillus cereus* isolates.** *Asian Pac J Trop Med* 2012, **5**:709–712.
43. Fenselau C, Havey C, Teerakulkittipong N, Swatkoski S, Laine O, Edwards N: **Identification of beta-lactamase in antibiotic-resistant *Bacillus cereus* spores.** *Appl Environ Microbiol* 2008, **74**:904–906.
44. Luna VA, King DS, Gulleddge J, Cannons AC, Amuso PT, Cattani J: **Susceptibility of *Bacillus anthracis*, *Bacillus cereus*, *Bacillus mycoides*, *Bacillus pseudomycoides* and *Bacillus thuringiensis* to 24 antimicrobials using Sensititre automated microbroth dilution and Etest agar gradient diffusion methods.** *J Antimicrob Chemother* 2007, **60**:555–567.
45. Bisicchia P, Bui NK, Aldridge C, Vollmer W, Devine KM: **Acquisition of VanB-type vancomycin resistance by *Bacillus subtilis*: the impact on gene expression, cell wall composition and morphology.** *Mol Microbiol* 2011, **81**:157–178.
46. Murray KD, Aronstein KA, De León JH: **Analysis of pMA67, a predicted rolling-circle replicating, mobilizable, tetracycline-resistance plasmid from the honey bee pathogen, *Paenibacillus larvae*.** *Plasmid* 2007, **58**:89–100.
47. Sirota-Madi A, Olender T, Helman Y, Ingham C, Brainis I, Roth D, Hagi E, Brodsky L, Leshkowitz D, Galatenko V, Nikolaev V, Mugasimangalam RC, Bransburg-Zabary S, Gutnick DL, Lancet D, Ben-Jacob E: **Genome sequence of the pattern forming *Paenibacillus vortex* bacterium reveals potential for thriving in complex environments.** *BMC Genomics* 2010, **11**:710.
48. Schlesinger SR, Kim SG, Lee JS, Kim SK: **Purification development and characterization of the zinc-dependent metallo- β -lactamase from *Bacillus anthracis*.** *Biotechnol Lett* 2011, **33**:1417–1422.
49. Hirooka K, Kunikane S, Matsuoka H, Yoshida KI, Kumamoto K, Tojo S, Fujita Y: **Dual regulation of the *Bacillus subtilis* regulon comprising the *ImrAB* and *yxgH* operons and *yxaf* gene by two transcriptional repressors, *LmrA* and *YxaF*, in response to flavonoids.** *J Bacteriol* 2007, **189**:5170–5182.
50. Ding R, Wu XC, Qian CD, Teng Y, Li O, Zhan ZJ, Zhao YH: **Isolation and identification of lipopeptide antibiotics from *Paenibacillus elgii* B69 with inhibitory activity against methicillin-resistant *Staphylococcus aureus*.** *J Microbiol* 2011, **49**:942–949.
51. Lal S, Tabacchioni S: **Ecology and biotechnological potential of *Paenibacillus polymyxa*: a minireview.** *Indian J Microbiol* 2009, **49**:2–10.
52. Wen Y, Wu X, Teng Y, Qian C, Zhan Z, Zhao Y, Li O: **Identification and analysis of the gene cluster involved in biosynthesis of paenibactin, a catecholate siderophore produced by *Paenibacillus elgii* B69.** *Environ Microbiol* 2011, **13**:2726–2737.
53. Tagg JR, Dajani AS, Wannamaker LW: **Bacteriocins of gram-positive bacteria.** *Bacteriol Rev* 1976, **40**:722–756.
54. Kjos M, Borrero J, Opsata M, Birri DJ, Holo H, Cintas LM, Snipen L, Hernández PE, Nes IF, Diep DB: **Target recognition, resistance, immunity and genome mining of class II bacteriocins from gram-positive bacteria.** *Microbiol* 2011, **157**:3256–3267.
55. Van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP: **BAGEL3: Automated identification of genes encoding bacteriocins and (non-) bacteriocidal posttranslationally modified peptides.** *Nucleic Acids Res* 2013, **41**:448–453.
56. Kajimura Y, Kaneda M: **Fusaricidin A, a new depsipeptide antibiotic produced by *Bacillus polymyxa* KT-8. Taxonomy, fermentation, isolation, structure elucidation and biological activity.** *J Antibiot* 1996, **49**:129–135.
57. Kajimura Y, Kaneda M: **Fusaricidins B, C and D, new depsipeptide antibiotics produced by *Bacillus polymyxa* KT-8: isolation, structure elucidation and biological activity.** *J Antibiot* 1997, **50**:220–228.
58. Savijoki K, Ingmer H, Varmanen P: **Proteolytic systems of lactic acid bacteria.** *Appl Microbiol Biotechnol* 2006, **71**:394–406.
59. Liu M, Bayjanov JR, Renckens B, Nauta A, Siezen RJ: **The proteolytic system of lactic acid bacteria revisited: a genomic comparison.** *BMC Genomics* 2010, **11**:36.
60. Christensen JE, Dudley EG, Pederson JA, Steele JL: **Peptidases and amino acid catabolism in lactic acid bacteria.** *Antonie Van Leeuwenhoek* 1999, **76**:217–246.

61. Anderson I, Scheuner C, Göker M, Mavromatis K, Hooper SD, Porat I, Klensk H-P, Ivanova N, Kyrpides N: **Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes.** *PLoS One* 2011, **6**:e20237.
62. Shipkowski, Stephanie: *Alternative functions for the Beta-Galactosidases of microorganisms found in environments without lactose*, PhD Thesis. University Park, PA 16802, United States: The Pennsylvania State University. Biochemistry, Microbiology, and Molecular Biology.
63. Benesová E, Lipovová P, Dvoráková H, Králová B: **beta-D-Galactosidase from *Paenibacillus thiaminolyticus* catalyzing transfucoosylation reactions.** *Glycobiology* 2010, **20**:442–451.
64. Lunde M, Aastveit AH, Blatny JM, Nes IF: **Effects of diverse environmental conditions on {phi}LC3 prophage stability in *Lactococcus lactis*.** *Appl Environ Microbiol* 2005, **71**:721–727.
65. Ivy RA, Wiedmann M, Boor KJ: ***Listeria monocytogenes* grown at 7°C shows reduced acid survival and an altered transcriptional response to acid shock compared to *L. monocytogenes* grown at 37°C.** *Appl Environ Microbiol* 2012, **78**:3824–3836.
66. Lehnik-Habrink M, Rempeters L, Kovács ÁT, Wrede C, Baierlein C, Krebber H, Kuipers OP, Stülke J: **DEAD-Box RNA helicases in *Bacillus subtilis* have multiple functions and act independently from each other.** *J Bacteriol* 2013, **195**:534–544.
67. Owttrim GW: **RNA helicases: diverse roles in prokaryotic response to abiotic stress.** *RNA Biol* 2013, **10**:96–110.
68. Kuhn E: **Toward understanding life under subzero conditions: the significance of exploring psychrophilic “cold-shock” proteins.** *Astrobiology* 2012, **12**:1078–1086.
69. Markkula A, Mattila M, Lindström M, Korkeala H: **Genes encoding putative DEAD-box RNA helicases in *Listeria monocytogenes* EGD-e are needed for growth and motility at 3°C.** *Environ Microbiol* 2012, **14**:2223–2232.
70. Bakermans C, Bergholz PW, Rodrigues DF, Vishnivetskaya TA: **Genomic and expression analyses of cold-adapted microorganisms.** In *Polar microbiology: life in a deep freeze*. Edited by Miller RV, Whyte LG. Washington, D.C: ASM Press; 2012:126–155.
71. Georgette D, Blaise V, Collins T, D'Amico S, Gratia E, Hoyoux A, Marx J-C, Sonan G, Feller G, Gerday C: **Some like it cold: biocatalysis at low temperatures.** *FEMS Microbiol Rev* 2004, **28**:25–42.
72. Metpally RP, Reddy BVB: **Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins.** *BMC Genomics* 2009, **10**:11.
73. Kahlke T, Thorvaldsen S: **Molecular characterization of cold adaptation of membrane proteins in the *Vibrionaceae* core-genome.** *PLoS One* 2012, **7**:e51761.
74. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DJ, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH: **CDD: conserved domains and protein three-dimensional structure.** *Nucleic Acids Res* 2013, **41**:D348–D352.
75. Dong D, Ihara T, Motoshima H, Watanabe K: **Crystallization and preliminary X-ray crystallographic studies of a psychrophilic subtilisin-like protease Apa1 from Antarctic *Pseudoalteromonas* sp. strain AS-11.** *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2005, **61**:308–311.
76. Zheng W, Kathariou S: **Differentiation of epidemic-associated strains of *Listeria monocytogenes* by restriction fragment length polymorphism in a gene region essential for growth at low temperatures (4 degrees C).** *Appl Environ Microbiol* 1995, **61**:4310–4314.
77. Garnier M, Matamoros S, Chevret D, Pilet MF, Leroi F, Tresse O: **Adaptation to cold and proteomic responses of the psychrotrophic biopreservative *Lactococcus piscium* strain CNCM I-4031.** *Appl Environ Microbiol* 2010, **76**:8011–8018.
78. Yamanaka K: **Cold shock response in *Escherichia coli*.** *J Mol Microbiol Biotechnol* 1999, **1**:193–202.
79. Dudley E, Steele J: ***Lactococcus lactis* LM0230 contains a single aminotransferase involved in aspartate biosynthesis, which is essential for growth in milk.** *Microbiol* 2001, **147**:215–224.
80. Sambrook J, Russell DW: *Molecular cloning*, Vol. 1–3. 3rd edition. New York: Cold Spring Harbour Laboratory Press; 2001.
81. Zerbino DR: **Using the Velvet de novo assembler for short-read sequencing technologies.** *Curr Protoc Bioinform* 2010. doi:10.1002/0471250953.bi1105s31.
82. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O: **The RAST Server: rapid annotations using subsystems technology.** *BMC Genomics* 2008, **9**:75.
83. **NCBI prokaryotic genome annotation pipeline.** <http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>.
84. Darling ACE, Mau B, Blattner FR, Perna NT: **Mauve: multiple alignment of conserved genomic sequence with rearrangements.** *Genome Res* 2004, **14**:1394–1403.
85. Hallin PF, Staerfeldt HH, Rotenberg E, Binnewies TT, Benham CJ, Ussery DW: **GeneWiz browser: an interactive tool for visualizing sequenced chromosomes.** *Stand Genomic Sci* 2009, **1**:204–215.
86. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**:2688–2690.
87. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40**:D290–D301.
88. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res* 2011, **39**:W29–W37.
89. Li L, Stoeckert CJ, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13**:2178–2189.
90. Pongstingl HNZ: **SMALT - a new mapper for DNA sequencing reads.** *In Intell Syst Mol Biol 2010 Meet* 2010, **U14**:1–1.
91. Vacic V, Uversky VN, Dunker AK, Lonardi S: **Composition profiler: a tool for discovery and visualization of amino acid composition differences.** *BMC Bioinformatics* 2007, **8**:211.
92. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**:45–48.

doi:10.1186/1471-2164-15-26

Cite this article as: Moreno Switt *et al.*: Genomic comparison of sporeforming bacilli isolated from milk. *BMC Genomics* 2014 **15**:26.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

