Propositions

1. The presence of proteolytic microorganisms in dairy concentrate production lines contributes to the diversity and spore load of specific thermophiles in end products.
   (this thesis)

2. The pattern of gene expression relating to the biofilm development is dependent on the circumstance where biofilms are formed but is also largely dissimilar for different biofilm-forming organisms.
   (this thesis)

3. On the one hand, 'omics' data yield novel insights into the cellular inner workings of organisms; on the other hand, the abundance of data also presents many hurdles.

4. The difficulty of translating biological insights to practical use is often underestimated by scientists as well as the general public.

5. There is a need to educate the general public in food safety and nutrition related science.

6. Ideally a PhD study is a personal development trajectory, a process for generating scientific understanding, and a process that delivers practical insights.

7. Success is an accumulation of small steps taken and it is crucial to go in the right direction and not to give up.

8. Acceptance for the ownership of the problem is the first step towards solving a problem.

Propositions belonging to the thesis, entitled

Thermophilic sporeformers from dairy processing environments

Yu Zhao

Wageningen, 1 July 2020

Thermophilic sporeformers from dairy processing environments

Yu Zhao

Thesis committee

Promotor

Prof. Dr M.H. Zwietering

Professor of Food Microbiology

Wageningen University & Research

Other members

Prof. Dr M.A.J.S. van Boekel, Wageningen University & Research

Prof. Dr S. Brul, University of Amsterdam

Dr M.H.J. Wells-Bennik, NIZO food research, Ede

Dr M. van der Voort, Wageningen University & Research

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences)

Thermophilic sporeformers from dairy processing environments

Yu Zhao

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 1 July 2020
at 4 p.m. in the Aula.

Yu Zhao
Thermophilic sporeformers from dairy processing environments, 160 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)
With references, with summary in English

ISBN 978-94-6395-345-0
DOI https://doi.org/10.18174/517519

Table of contents

Chapter 1 General introduction

1.1 Dairy powder production

Dairy products are an important part of human diet. They contain many nutrients, including high-quality protein, and essential nutrients like minerals (e.g. calcium) and vitamins (e.g. vitamin D), which are considered necessary for homeostasis and therefore good health. Storage of raw milk is challenging since it is prone to microbial spoilage because of its neutral pH, high water content, and high nutrient content. In order to extend its shelf-life, different preservation techniques are applied (Law and Mabbitt, 1983). Powderization, for example, is a widely used method for the preservation of various dairy products, such as whole milk, non-fat milk, and dry buttermilk. Compared to butter, cheese or fluid cow milk products, milk powder consumption is, however, smaller. It is an important diet option for many people, especially for people in the countries where cooling facilities are not widely available. For example, China was the largest whole dry milk consumption area in the world in 2018, and the consumption of whole milk powder takes up 14% of total cow milk consumption (USDA/FAS, 2018).

In a nutshell, powderization is about transforming liquid milk into dry powder, which requires removal of (almost) all the water. Two main water removal processes used in milk powder industry are vacuum evaporation and spray drying, which can be further supplemented by other fluid removing technologies such as membrane processes or fluid bed drying (Pisecky, 2012). As an example of a dry powder process, the dairy powder production process in a New Zealand whole milk powder factory, as Scott et al. (2007) described, started with the raw milk's separation, pasteurization, and standardization. There the raw milk treatment runs were 6-8 hours in length. Raw milk was first preheated using a plate heat exchanger (PHE), and then separated into skim milk and cream. The skim milk and cream were then pasteurized separately. The skim milk and cream were then mixed to achieve a specified composition, in a process known as standardization. The standardized milk, which was stored at 4°C, was then directed to different dairy powder production processes to make products such as dry whole milk, non-fat dry milk, and dry dairy blends.

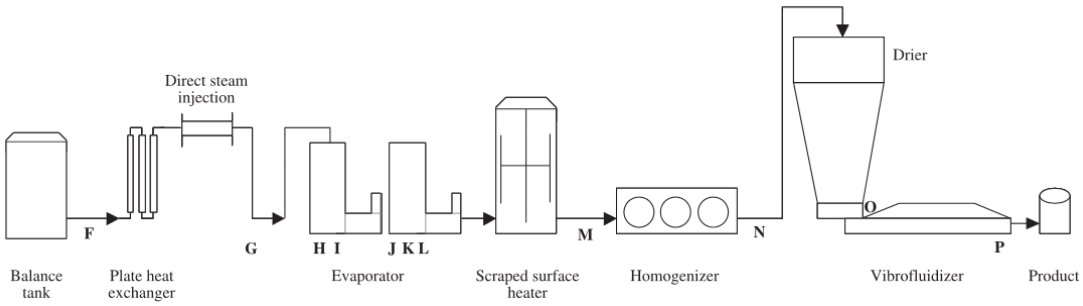


Figure 1.1 Schematic diagram of the evaporation and drying process (Scott et al., 2007).

In the case of dry whole milk at this New Zealand whole powder factory described in Scott et al. (2007), the milk powder manufacturing runs were approximately 18 hours in length. To begin with, the temperature of the milk is increased by pushing it through a plate heat exchanger (PHE) and a direct steam injection (DSI). Following the heating treatment by PHE and DSI, the milk goes into evaporators. Following evaporation, the concentrated milk is sent to a scraped surface pre-heater and then undergoes homogenization before being dried and packed (Figure 1.1) (Scott et al., 2007).

1.2 Spoilage organisms associated with dairy powder products

Dairy powder products (dairy-concentrate end products) are generally considered to be microbiologically stable, because of their low water content which prevents microbial cells from growing. However, bacterial spores may be present in the product. The ecology of spore formation is further discussed later in this chapter. After the powder is reconstituted with other products that have high water content, pre-existing spores can germinate and grow. This may result in enzyme and acid production, with the consequential development of an off-flavour, loss of structure, or coagulation in the end products (Chopra and Mathur, 1984; Chen et al., 2004).

Thermophilic bacilli are associated with contamination in dairy-concentrate processing environments (Burgess et al., 2010), and they can out-compete other bacteria in dairy-concentrate processing lines where high temperatures are applied,

and are a primary concern in such facilities (Watterson et al., 2014). In the dairy industry, thermophilic sporeformers are usually enumerated at 55 °C on aerobic plate count agar. Those that have been isolated from dairy products at this temperature can be divided into two groups: obligatory thermophiles and facultative thermophiles (also known as thermo-tolerant microorganisms). According to literature the obligate thermophiles grow at temperatures in the range of 30°C to 72°C; typical examples include *Anoxybacillus* spp. and *Geobacillus* spp. (Flint et al., 2001; Ronimus et al., 2003; Scott et al., 2007). Facultative thermophiles can grow at both mesophilic and thermophilic temperatures (approximately 15°C - 65°C). Examples of facultative thermophiles include *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus sporothermodurans*, and *Bacillus subtilis* (Crielly et al., 1994; Flint et al., 2001; Ronimus et al., 2003, Scheldeman et al., 2005). Obligatory thermophilic bacilli are less of a concern since they generally do not grow at temperatures below 37°C, while dairy-based concentrates are usually stored at temperatures below 37°C. However, exceptions have been reported, for example, obligatory thermophilic bacilli *Geobacillus stearothermophilus* are considered to be responsible for the flat sour spoilage of evaporated milk, which is milk with lowered water content and has not been subject to the final drying process which transforms it into milk powder (Kalogridou-Vassiliadou, 1992; Olson and Sorrells, 1992).

Facultative thermophilic bacilli are reported to be more involved in incidents of spoilage: strains of *B. licheniformis* are capable of producing a slimy extra-cellular substance that can affect the quality of pasteurized milk and cream; *B. subtilis* has been associated with ropiness in raw and pasteurized milk as well as the spoilage of UHT and canned milk products; *B. coagulans* have been connected to the spoilage of UHT and canned milk products due to their production of lactic acid (Burgess et al., 2010). Under favourable environmental conditions, evaporated milk may undergo flat sour spoilage when it contains viable spores capable of germinating and growing at both mesophilic and thermophilic temperatures, depending on the strain (Gordon et al., 1989). Table 1.1 presents the growth characteristics of several thermophilic bacilli. In this thesis, obligatory thermophilic sporeformers are studied, which are here defined to have a growth range from 30 to 75 °C. Mesophilic sporeformers are defined to grow from 5 to 35 °C.

Table 1.1 Growth characteristics of several thermophilic bacilli (Burgess et al., 2010).

	Anoxybacillus flavithermis	*Geobacillus stearothermo-philus*	*Geobacillus thermo-leovorans*	*Bacillus licheniformis*	*Bacillus subtilis*	*Bacillus coagulans*	*Bacillus pumilus*	*Bacillus sporo-thermodurans*
Maximum growth temperature (°C)	65-72	65-68	70	50-55	45-55	57-61	50-55	45-55
Minimum growth temperature (°C)	30-38	37	35-47	15	5.0-20.0	15-25	5.0-15.0	20
Anaerobic growth	Yes	No	No	Yes	No	Yes	No	No
pH range	6.0-9.0	6.0-8.0	5.2-8.0	5.5-8.5	5.5-8.5	4.0-10.5	5.5-8.5	unknown
Sporangium swollen	Yes	Yes	Yes	No	No	Variable	No	No
Spore position	Terminal	Terminal	Terminal	Central	Central	Sub-terminal	Central	Terminal
Voges-Proskauer	Positive	Negative	Negative	Positive	Positive	Variable	Positive	Negative
Growth in 7% NaCl	No	No	No	Yes	Yes	No	Yes	No
Nitrate reduced to nitrite	Yes	Variable	Yes	Yes	Yes	Variable	No	No
Casein hydrolysis	Yes	Variable	Variable	Yes	Yes	No	Yes	Weak
Gelatin hydrolysis	No	Yes	Variable	Yes	Yes	Variable	Yes	No

1.3 The origin of the contamination of dairy-concentrate processing environments by thermophilic bacilli

Thermophilic bacilli can be found in low numbers (< 1 log CFU/mL) in raw milk and may originate from the feed and milking equipment, where high numbers of heat resistant spores have been detected (Te Giffel et al., 2002). When animals consume feed contaminated by spore-forming bacteria, large quantities of spores can end up in their feces and in turn contaminate their udders and teats (Te Giffel et al., 2002). These low numbers of thermophilic bacilli in raw milk serve as the initial inoculum which could lead to the contamination by thermophilic bacilli in dairy-concentrate end products; however, studies have shown that the degree of contamination of thermophilic bacilli in dairy-concentrate end products is not related to the quality of raw milk (Scott et al., 2007). When the low number (<1 log CFU/ml) of thermophilic spores present in raw milk is introduced into the dairy-concentrate processing line, they cannot lead to contamination with cell counts larger than 10^6 CFU/g (the cell number which can lead to noticeable spoilage), provided that the cells are not given the chance to reside and proliferate during the processing. It has been observed that

a large increase of thermophilic bacilli spore counts occurs after the milk undergoes the PHE and evaporation processing steps (see Figure 1.1) (Scott et al., 2007). Moreover, the bacteria present in the fouling that remains in the DSI units and evaporators after cleaning in place (CIP) were predominantly in their spore form. This suggests that fouling can be a possible source of spores, which ultimately contaminate dairy-concentrate end products (Scott et al., 2007).

1.4 Factors contributing to the contamination of dairy powder products by thermophilic bacilli

There are several possible contributing factors to the capability of thermophilic bacilli to be a relevant contaminant in dairy powder products (dairy-concentrate end products). Although the initial levels of thermophilic sporeformers from the dairy farm can contribute to the contamination, the contributing factors relevant to the dairy-concentrates processing environment are discussed here.

The first factor is the ability of thermophilic bacilli to grow readily in dairy environments if temperature allows. Although milk is a nutrient-rich food, not all bacteria can grow in it, because milk contains several antimicrobial components, including lactoferrin, lactoperoxidase, lysozyme, and possibly N-acetyl-ß-D-glucosaminidase (Losnedahl et al., 1998), which are capable of reducing the growth of bacteria. However, thermophilic bacilli can grow readily in dairy environments if temperature allows, contributing to the growth or accumulation during dairy-concentrate processes and the contamination of the final power products. After rehydration, those sporeformers being present in the products can rapidly grow in the products under favourable conditions. During the growth, those thermophilic bacilli can produce acid, and degrade protein, this in turn will also lead to unfavourable product characteristics.

The second factor is the ability of thermophilic bacilli to form heat-resistant spores – i.e., bacteria in a dormant state – which contributes to the long survival in the dairy powder products and survival of heat treatments. Bacteria form spores mainly when they undergo environmental stresses (e.g. population density, lack of nutrients) (Tan and Ramamurthi, 2014). Since the spores of thermophilic bacilli are resistant to heat and chemicals (Setlow, 2006), when they are formed in a plant they are difficult to

eradicate, even with extreme heat processes, and can therefore end up in the dairy end products. The heat resistance of thermophilic bacilli spores varies widely, because it can be influenced by temperature, pH, and medium composition during the sporulation process (Burgess et al., 2010; Watterson et al., 2014). When the spores end up in the dairy powder products, they will usually remain dormant since well-kept dairy powder products will remain dry within the "best-before date" if stored as recommended. However, if the contaminated products are exposed to favourable conditions (so after mixing with other ingredients having more water), the thermophilic bacilli can be activated. For example, thermophilic sporeformers can be activated by heat, chemicals, or a decrease of pH (Kim and Foegeding, 1990; Rajan et al., 2006; de Vries, 2006; Ghosh et al., 2009), or activated by nutrients (e.g. L-alanine) or by high pressure, salts, or lysozyme (Setlow, 2003) and start germinating. This germination can then lead to spoilage of the products if the condition is favourable for the thermophilic sporeformers to grow. Thermophilic sporeformers are also notorious in this since they have a very short doubling time (at high temperatures), and they can spoil the product rapidly.

The third factor is the ability of thermophilic bacilli to form biofilms, which is also considered a very important factor contributing to their contamination of dairy powder products (Scott et al., 2007). It is well accepted that the biofilm life style is a feature common to most microorganisms in natural, medical, and engineered systems (Hobley et al., 2015). Biofilms consist of cells that are encompassed by complex biopolymer layers known as the extracellular matrix. These biopolymers can be proteins, DNA, and/or polysaccharides. There can exist single-species biofilms and mixed-species biofilms. In single-species biofilms, cells are able to differentiate into different variants, for example, more resistant variants can develop or cells can transform into spores, in order to promote the survival of the species under harsh conditions (Evans, 2015; Verplaetse et al., 2015). In mixed-species biofilms different species can reside in the biofilm, cooperating to support each other (Periasamy and Kolenbrander, 2009; Elias and Banin, 2012), or competing with each other for growth in the biofilm structure (Rendueles and Ghigo, 2012). Biofilm development consists of four stages: attachment, development, maturation, and dispersal. It has been reported that the attachment stage can be initiated both actively, through signalling molecules such as quorum sensing molecules (Davies et al., 1998; He et al., 2015), and passively, as a consequence of surface attraction between cell surface and substratum (Van Houdt and Michiels, 2005). After attaching to a substratum, while

maintaining their surface-attaching lifestyle, cells start to proliferate and differentiate. During the biofilm development, pathways important for regulation of biofilm formation have often mapped to global regulatory systems that mediate broad changes in cell physiology as a mean to adapt to specific environments (Monds et al., 2009). When cells reach the late growth phase, they will usually also disperse from the biofilm and return to planktonic environments, and the dispersed cells may then colonize other surfaces. The trigger for dispersal from the biofilm can also be either active or passive. Passive dispersal is caused by the sloughing of cells and erosion from the biofilm; active dispersal, on the other hand, is a highly regulated process. There are a range of environmental cues that trigger active dispersal from biofilm, including alterations in the availability of nutrients, such as carbon sources, oxygen depletion, low levels of nitric oxide, changes in temperature, and high or low levels of iron. In addition, there are several bacterially derived signals that can induce dispersal, including acyl-homoserine lactones, autoinducing peptides[1], diffusible fatty acids, and D-amino acids (McDougald et al., 2012). With regard to dairy powder production, biofilm formation during the process can contribute to the persistence of thermophilic bacilli in the processing line, because it increases the bacteria's ability to resist harsh conditions such as CIP cleaning.

1.5 Functional genomics as a tool to identify potential spoilers of dairy powder products

Currently, our knowledge about the mechanisms connected to the growth in certain niche environments and the biofilm formation of thermophilic bacilli, which can contaminate and spoil dairy powder products, is limited. Our knowledge can be built up through the application of functional genomics tools including genomics, transcriptomics, proteomics, and metabolomics; these are the new generation high-throughput tools which can be used to investigate the behaviour of bacteria at the DNA, RNA, protein, and metabolite levels, respectively. Moreover, these functional

[1] Extracellular peptides, ranging from 5 to 34 amino acids in length, that are generated by cleavage from precursor peptides and then further post-transcriptionally modified. These peptides are used by Gram-positive bacteria as cell communication signals. For example, the *S. aureus* quorum-sensing system is encoded by the accessory gene regulator (agr) locus and the communication molecule that it produces and senses is called an autoinducing peptide (Boles et al., 2008).

genomics techniques can generate high data volumes in the field in a short period. This contrasts with classical methods, in which hypotheses are made based on the prior knowledge and then limited numbers of selected aspects are assessed in experimental settings. Functional genomics tools thus allow us to investigate with a holistic, unbiased approach, in which a massive number of cellular molecules are studied in chosen experimental settings and time frames (Brul et al., 2006). In addition, biostatistics and the comparative analysis of the collected data, can result in new biological insights. In our research, besides using transcriptomic analysis to investigate biofilm-forming mechanisms (Chapter 4 of this thesis), we used a comparative genomic approach, that is, we connected comparative genomics analyses with phenotypic experiments, to explore genomic characteristics linked to phenotypic indicators of milk spoilage at high temperature (Chapter 5 of the thesis). Processes that are specifically active in biofilm-phase cells were found, and potential biomarkers that could predict the contamination potential of a thermophilic bacillus in dairy powder products were identified. Apart from creating some useful biological insights and providing support for further knowledge-based hypotheses, functional genomic analysis also generates a great deal of information which requires careful interpretation and validation. This aspect of functional genomics approach is further discussed in Chapter 6 of this thesis.

1.6 Thesis outline

Thermophilic sporeformers, mainly thermophilic bacilli, are a primary concern for plants producing dairy-based concentrates. Over the last few years, some knowledge has been gained regarding the prevalence of these bacteria, and several techniques have been developed for cleaning contaminated environments. Nevertheless, the prevalence and growth characteristics of thermophilic sporeformers in the dairy-concentrate producing plants are not known. This sets major limits to the development of an efficient method to control the contamination of thermophilic bacilli in the end products.

This study was aimed to study the genomic, physiological and environmental aspects contributing to the contamination of dairy powder products by thermophilic bacilli in a dairy-concentrate processing plant, with a strong focus on the surface-attached bacterial community. Another objective of the study was to investigate factors

contributing to the outgrowth of the persistent and common thermophilic spoilers. In addition, the genomic characteristics as indicators for the potential of thermophilic bacilli to contaminate dairy powder products were explored.

Chapter 2 describes thermophilic sporeformers present at different locations in a dairy-concentrate processing environment. This chapter also emphasizes the importance of abiotic and microbiotic factors for niche colonization in dairy plants, where the presence of thermophilic bacilli can affect the quality of end products. Moreover, in Chapter 2, a growth dependence of one major thermophilic contaminant, *G. thermoglucosidans*, on other dairy isolates was found.

Chapter 3 further communicates the study on the mechanism behind the previously described growth dependence of *G. thermoglucosidans* in skim milk. Different possibilities were investigated using both comparative genomic methods and phenotypic assays. In the end, the release of glucose and galactose by the other dairy isolate was found to be the element which *G. thermoglucosidans* was dependent on for rapid growth in skim milk.

Chapter 4 documents the differences of gene expression profiles of *G. thermoglucosidans* between biofilm and planktonic phases during biofilm development from a transcriptomic study. The main objective of this study was to find out biofilm formation specific gene expressions. The results indicate that there are discernible differences of expression profiles between biofilm-phase cells and planktonic-phase cells of *G. thermoglucosidans*. Categories of genes significantly up- or down-regulated in biofilm development processes were further studied and also described in this chapter.

Chapter 5 focuses on the study of spoilage capabilities of 22 thermophilic sporeformers in dairy-concentrate processing environments. Their genomes and their abilities in the spoilage related activities were evaluated. Together with the additional experimental evidences and genomic analysis of selected hypothetical biomarkers, the data described in this chapter can facilitate the identification of targets for the detection and control of contamination of thermophilic sporeformers in dairy-concentrate processing environments.

Finally, **Chapter 6** integrates the topics discussed in this thesis, setting them in perspective.

1.7 References

Boles, B. R., and A. R. Horswill. 2008. Agr-mediated dispersal of *Staphylococcus aureus* biofilms. PLoS Pathogens. 4: e1000052.

Brul, S., F. Schuren, R. Montijn, B. J. Keijser, H. Van Der Spek and S. J. Oomes. 2006. The impact of functional genomics on microbiological food quality and safety. International Journal of Food Microbiology. 112:195-199.

Burgess, S. A., D. Lindsay and S. H. Flint. 2010. Thermophilic bacilli and their importance in dairy processing. International Journal of Food Microbiology. 144:215-225.

Bylund G., B. Malmgren, A. Holanowski, M. Hellman, G. Mattsson and B. Svensson. 2015. Dairy processing handbook. Tetra pak processing system AB. Sweden.

Chen, L., T. Coolbear and R. M. Daniel. 2004. Characteristics of proteinases and lipases produced by seven *Bacillus* sp. isolated from milk powder production lines. International Dairy Journal. 14:495-504.

Chopra, A. K. and D. K. Mathur. 1984. Isolation, screening and characterization of thermophilic *Bacillus* species isolated from dairy products. Journal of Applied Bacteriology. 57:263-271.

Crielly, E. M., N. A. Logan and A. Anderton. 1994. Studies on the *Bacillus* flora of milk and milk products. Journal of Applied Microbiology. 77:256-263.

Davies, D. G., M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton and E. P. Greenberg. 1998. The involvement of cell-to-cell signals in the development of a bacterial biofilm. Science. 280:295-298.

De Vries, Y. P. 2006. *Bacillus cereus* spore formation, structure, and germination. PhD Thesis, Wageningen University.

Elias, S. and E. Banin. 2012. Multi-species biofilms: Living with friendly neighbours. FEMS Microbiology Review. 36:990-1004.

Evans, T. J. 2015. Small colony variants of *Pseudomonas aeruginosa* in chronic bacterial infection of the lung in cystic fibrosis. Future Microbiology. 10:231-239.

Flint, S. H., L. J. Ward, and K. M. Walker. 2001. Functional grouping of thermophilic *Bacillus* strains using amplification profiles of the 16S–23S internal spacer region. Systematic and Applied Microbiology. 244:539-548.

Ghosh, S., P. Zhang, Y. Q. Li and P. Setlow. 2009. Super-dormant spores of *Bacillus* species have elevated wet-heat resistance and temperature requirements for heat activation. Journal of Bacteriology. 191:5584-5591.

Gordon, R. E., W. C. Haynes and C. H-N. Pang. 1973. The genus *Bacillus*. US Department of Agriculture Handbook. Washington. D.C.

He, Z., J. Liang, Z. Tang, R. Ma, H. Peng and Z. Huang. 2015. Role of the luxS gene in initial biofilm formation by *Streptococcus mutans*. Journal of Molecular Microbiology and Biotechnology. 25:60-68.

Hobley, L., C. Harkins, C. E. MacPhee and N. R. Stanley-Wall. 2015. Giving structure to the biofilm matrix: An overview of individual strategies and emerging common themes. FEMS Microbiology Review. 39:649-669.

Kalogridou-Vassiliadou, D. 1992. Biochemical activities of *Bacillus* species isolated from flat sour evaporated milk. Journal of Dairy Science. 75:2681-2686.

Kim, J. and P. M. Foegeding. 1990. Effect of heat treatment, $CaCl_2$ treatment and ethanol treatment on activation of *Bacillus* spores. Journal of Applied Bacteriology. 69:414-420.

Law, B. A. and L. A. Mabbitt. 1983. New methods for controlling the spoilage of milk and milk products. The Society for Applied Bacteriology Symposium Ser. 11:131-150.

Losnedahl, K. J., H. Wang, M. Aslam, S. Zou and W. L. Hurley. 1998. Antimicrobial factors in milk. Illinois DairyNet Papers, University of Illinois.

McDougald, D., S. A. Rice, N. Barraud, P. D. Steinberg and S. Kjelleberg. 2012. Should we stay, or should we go: Mechanisms and ecological consequences for biofilm dispersal. Nature Reviews Microbiology. 10:39-50.

Monds, R. D., G. A. O'Toole. 2009. The developmental model of microbial biofilms: Ten years of a paradigm up for review. Trends Microbiology. 17:73-87.

Olson, K. and K. Sorrells. 1992. Thermophilic flat sour sporeformers. Compendium of methods for the microbiological examination of foods, 3rd Edition. American Public Health Association, Washington, DC.

Periasamy, S. and P. E. Kolenbrander. 2009. *Aggregatibacter actinomycetemcomitans* builds mutualistic biofilm communities with *Fusobacterium nucleatum* and *Veillonella* species in saliva. Infection and Immunity. 77:3542-3551.

Písecký, J. 2012. Handbook of milk powder manufacture. GEA Process Engineering A/S. Copenhagen, Denmark.

Rajan, S., S. Pandrangi, V. M. Balasubramaniam and A. E. Yousef. 2006. Inactivation of *Bacillus stearothermophilus* spores in egg patties by pressure-assisted thermal processing. LWT-Food Science and Technology. 39:844-851.

Rendueles, O. and J. M. Ghigo. 2012. Multi-species biofilms: How to avoid unfriendly neighbours. FEMS Microbiology Review. 36:972-989.

Ronimus, R. S., L. E. Parker and H. W. Morgan. 1997. The utilization of RAPD-PCR for identifying thermophilic and mesophilic *Bacillus* species. FEMS Microbiology Letters. 147:75-79.

Scheldeman, P., A. Pil, L. Herman, P. De Vos and M. Heyndrickx. 2005. Incidence and diversity of potentially highly heat-resistant spores isolated at dairy farms. Applied and Environmental Microbiology. 71:1480-1494.

Scott, S. A., J. D. Brooks, J. Rakonjac, K. M. Walker and S. H. Flint. 2007. The formation of thermophilic spores during the manufacture of whole milk powder. International Journal of Dairy Technology. 60:109-117.

Setlow, P. 2003. Spore germination. Current Opinion in Microbiology. 6:550-556.

Setlow, P. 2006. Spores of *Bacillus subtilis*: Their resistance to and killing by radiation, heat and chemicals. Journal of Applied Microbiology. 101:514-525.

Tan, I. S. and K. S. Ramamurthi. 2014. Spore formation in *Bacillus subtilis*. Environmental Microbiology Reports. 6:212-225.

Te Giffel, M. C., A. Wagendorp, A. Herrewegh and F. Driehuis. 2002. Bacterial spores in silage and raw milk. Antonie Van Leeuwenhoek. 81:625-630.

USDA/FAS. 2018. Dairy: world markets and trade. <http://www.fas.usda.gov/psdonline/circulars/dairy.pdf>.

Van Houdt, R. and C. W. Michiels. 2005. Role of bacterial cell surface structures in *Escherichia coli* biofilm formation. Research in Microbiology. 156:626-633.

Verplaetse, E., L. Slamti, M. Gohar and D. Lereclus. 2015. Cell differentiation in a *Bacillus thuringiensis* population during planktonic growth, biofilm formation, and host infection. MBio. 6:e00138-15.

Watterson, M. J., D. J. Kent, K. J. Boor, M. Wiedmann and N. H. Martin. 2014. Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest. Journal of Dairy Science. 97:2487-2497.

Chapter 2 Abiotic and microbiotic factors controlling biofilm formation of thermophilic spore formers

Yu Zhao,
Martien P. M. Caspers,
Karin I. Metselaar,
Paulo de Boer,
Guus Roeselers,
Roy Moezelaar,
Masja Nierop Groot,
Roy C. Montijn,
Tjakko Abee,
Remco Kort

This chapter was published in Applied and Environmental Microbiology (2013) 79 (18), 5652–5660

2.1 Summary

One of the major concerns in the production of dairy concentrates is the risk of contamination by heat-resistant spores from thermophilic bacteria. In order to acquire more insight in the composition of microbial communities occurring in the dairy concentrate industry, a bar-coded 16S-amplicon sequencing analysis was carried out on milk, final products and fouling samples taken from dairy concentrate production lines. The analysis of these samples revealed the presence of DNA from a broad range of bacterial taxa, including a majority of mesophiles and a minority of (thermophilic) spore forming bacteria. Enrichments of fouling samples at 55°C showed the accumulation of predominantly *Brevibacillus* and *Bacillus*, whereas enrichments at 65°C led to the accumulation of *Anoxybacillus* and *Geobacillus* species. Bacterial population analysis of biofilms grown using fouling samples as an inoculum indicated that both *Anoxybacillus* and *Geobacillus* preferentially form biofilms on surfaces at air-liquid interfaces rather than on submerged surfaces. Three of the most potent biofilm forming strains isolated from the dairy factory industrial samples, including *Geobacillus thermoglucosidans*, *Geobacillus stearothermophilus*, and *Anoxybacillus flavithermus* have been characterized in detail with respect to their growth conditions and spore resistance. Strikingly, *Geobacillus thermoglucosidans*, which forms the most thermostable spores of these three species, is not able to grow in dairy intermediates as a pure culture but appears dependent for growth on other spoilage organisms present, probably as a result of their proteolytic activity. These results underscore the importance of abiotic and microbiotic factors on niche-colonization in dairy factories, where the presence of thermophilic spore formers can affect the quality of end products.

2.2 Introduction

Contamination by spore-forming bacteria is an important concern in the production of dairy concentrates. Besides mesophilic bacteria, thermophiles are problematic in food-producing industrial facilities operating from 40°C to 65°C, as these temperatures support growth and biofilm formation of thermophilic spore formers (Burgess et al., 2010b). The growth of these thermophiles in biofilms in factories can result in numbers of up to 10^6 CFU/g of bacteria and spores released in the final products, including whey and milk concentrates (Scott et al. 2007). These spores could germinate when the conditions are favourable, finally resulting in high numbers of bacteria and off-flavour in end products (Scheldeman et al., 2005, Scott et al., 2007). In order to prevent the presence and outgrowth of the accumulated spores, costly precautions such as frequent cleaning, short production runs and intensive microbial product control are required.

Most thermophilic spore formers, which have been identified so far in dairy processing lines and products, belong to the genera of *Bacillus, Geobacillus* and *Anoxybacillus* (Flint et al.,1997b, Scott et al., 2007, Yuan et al., 2012). *Geobacillus* spp. and *A. flavithermus* are the most frequently reported species in thermophilic dairy biofilms (Burgess et al., 2010b). The presence of spores from these thermophilic bacilli in the final products most likely results from the detachment of spores from biofilms on stainless steel surfaces found within a milk powder plant (Scott et al., 2007). However, it is not evident that these organisms are the only organisms important for biofilm formation in dairy processing environments. Insight in the species diversity and the contribution of both thermophilic and mesophilic species in microbial populations at the different sites in dairy concentrate production lines is currently lacking. In this study we applied a bar-coded 16S-amplicon sequencing approach (Nocker et al., 2010) to get insight in the microbial composition of fouling samples in dairy concentrate processing plants and evaluated the effect of enrichments at high temperatures, at air-liquid interface or on different surfaces. We isolated three thermophilic species on the basis of their ability to grow at high temperatures and efficiency to form biofilms under lab conditions. We provide evidence that suggests that growth in milk-based media of *G. thermoglucosidans* is dependent on proteolytic activity of other species present in dairy concentrate processing environments.

2.3 Results

2.3.1 Enrichment of *Geobacillus* and *Anoxybacillus* at high temperatures.

The contribution of thermophilic spore formers to the contamination of the dairy processing lines and end products was evaluated by an analysis of the microbial composition of dairy fouling samples by bar-coded 16S-rRNA amplicon sequencing up to the genus level. At the phylum level, the dairy fouling samples were dominated by Firmicutes and Proteobacteria (55% and 42% respectively). The majority of the 16S rRNA sequences in each sample represented a wide variety of mesophilic genera (Figure 2.1), covering many genera of the classical milk microbial flora (Delbes et al., 2007, Ercolini et al., 2009, Lafarge et al., 2004, Scheldeman et al., 2005, De Jonghe et al., 2008); only a minor fraction of 16S-rRNA sequences were associated to the thermophilic genera *Anoxybacillus* and *Geobacillus***.** It should be noted that the standard enumeration method for thermophilic species at 55°C also provides conditions for some mesophilic species to grow. Therefore, the composition analysis on the 14 dairy samples was also carried out after enrichment at 65°C (Figure 2.1). Overnight (O/N) incubation of the 14 dairy samples at 55°C resulted in the enrichment of spore forming genera *Bacillus* (four samples) or *Brevibacillus* (six samples) and in some cases in the enrichment of thermophilic spore forming genera *Geobacillus* (two samples) and *Anoxybacillus* (one sample). An increase of the enrichment temperature to 65°C resulted in a higher predominance of thermophilic genera, including *Geobacillus* (seven samples) and *Anoxybacillus* (three samples). In eight samples little or no growth occurred (- or -/+), showing a similar composition to that present in the samples prior to enrichment at 65°C.

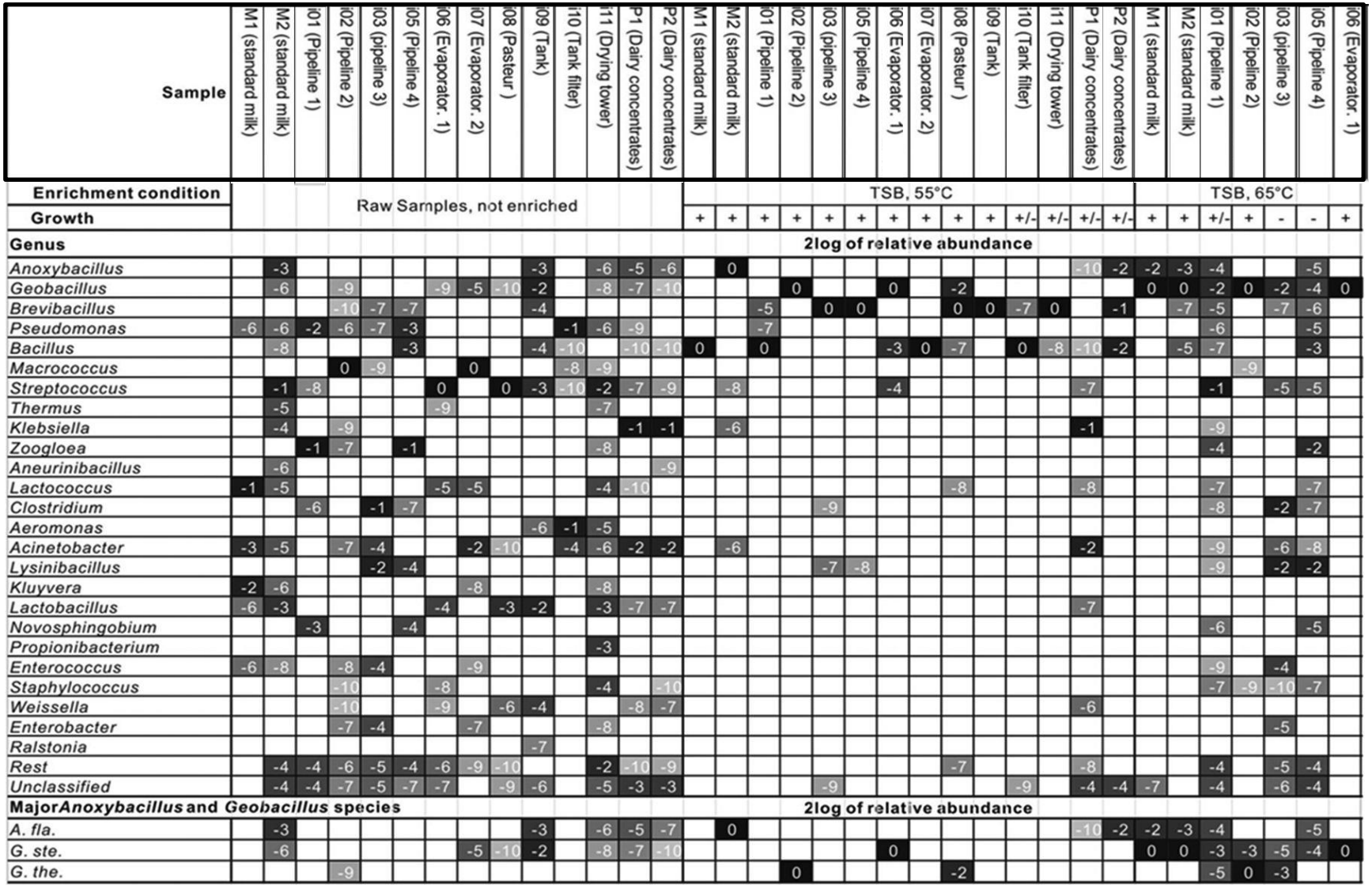


Figure 2.1 Microbial inventory of raw and enriched dairy samples. Genomic DNA isolated from industrial samples (raw, enriched overnight at 55°C or 65°C in TSB medium) was analysed by mass-sequencing 16S-genotyping (500-2000 sequences per sample). Grey levels of cells represent relative abundance of microbial composition at the genus level or species level; from black to white indicates high and low abundance levels. Number represents 2log (relative abundance). From top to the bottom of the figure, each row describes the abundance of one genus or species in each of 14 locations selected from a dairy processing plant. Abbreviations: A. fla: Anoxybacillus flavithermus; G. ste: *Geobacillus stearothermophilus*; G. the.: *Geobacillus thermoglucosidans*

2.3.2 Preference of thermophiles for air-liquid-interface or submerged biofilms.

The next experiment was aimed at the identification of thermophilic genera in different types of biofilms formed at high temperatures. Static biofilm systems were inoculated with three of our previously isolated dairy samples (two standard milk samples, and one whey evaporator sample). The incubations were carried out at 55°C and 65°C in the submerged steel biofilm system and the standing steel biofilm model system, which includes an air-liquid interface, as described in the materials and methods and displayed in Figure S2.1. The total viable counts of the different fractions in the standing steel biofilm model (medium, standing steel coupon and plastic well) were determined (Table S2.2). The counts in the planktonic fraction at 55°C and 65°C were approximately 1000-fold higher than the initial counts of the dairy samples at 55°C, indicating that enrichment of thermophiles occurred at 55°C and 65°C in milk.

Subsequently, the different fractions in the biofilm model system were analysed for their microbiological composition by bar-coded 16S-amplicon sequencing. The thermophilic genera *Anoxybacillus* and *Geobacillus* dominated in most of the samples (Figure 2.2). Relatively high numbers of *Anoxybacillus* were found after enrichment at both 55°C and 65°C, whereas 16S rRNA gene sequences affiliated to *Geobacillus* dominated the population when samples were incubated at 65°C. Besides, the mesophilic spore forming genus *Anaerinibacillus* was enriched at 55°C. We observed that the contribution of the thermophilic genera *Geobacillus* and *Anoxybacillus* in biofilms was higher in the air-liquid interface biofilms (standing steel) compared to the submerged biofilms, where the genus *Pseudomonas* dominated at 55°C and 65°C. Although the latter genus is not a known thermophilic biofilm former, it should be noted that a thermophilic *Pseudomonas* species has been described growing at 55°C (Manaia and Moore, 2002). Thermophilic populations which adhere to steel and plastic surfaces were found to be nearly identical in our model system (Figure 2.2). The presence of the species *A. flavithermus*, *G. stearothermophilus*, and *G. thermoglucosidans* is shown in the three bottom rows of Figure 2.2. While *A. flavithermus*, *G. stearothermophilus* were frequently enriched in the standing steel biofilm system, *G. thermoglucosidans* was not found in any of the samples enriched in milk medium. Apparently, this species does not readily accumulate in milk medium, possibly resulting from a growth-dependence, as described below.

Sample	M1 (standard milk)													M2 (standard milk)							Evaporator (i7)			
																						+ Sterile Milk		
Enrichment condition	Raw	55°C			65°C			55°C			65°C			Raw	55°C			65°C			Raw	65°C		
Culture fraction		M	subm.S		M	subm.S		M	st.S	P	M	st.S	P		M	st.S	P	M	st.S	P		M	st.S	P
Genus	**2log of relative abundance**																							
Anoxybacillus		0		-7	-2	-4	-2	-1	0	0	0	0	0	-3	-1	0	0	-2	-1	-1		-3		
Geobacillus									-8		-2	-3	-3	-6				-1	-1	-1	-5	0	0	0
Brevibacillus				-8	-6																	-8		
Pseudomonas	-6		0	0	-1	0	-1							-6										
Bacillus				-8	-3	-4								-8								-8		
Macrococcus																					0	-5	-6	
Streptococcus											-10			-1				-6		-7		-9		
Thermus											-9			-5	-8			-3	-4	-4				
Klebsiella														-4	-8			-7						
Zoogloea																								
Aneurinibacillus								-2	-4	-7				-6	-1	-4	-7					-5		
Lactococcus	-1					-7	-4							-5							-5			
Clostridium																								
Aeromonas																								
Acinetobacter	-3						-5	-10						-5	-8						-2	-7		
Lysinibacillus																								
Kluyvera	-2													-6							-8			
Lactobacillus	-6									-7				-3					-7					
Novosphingobium																								
Propionibacterium																				-7		-9	-9	
Enterococcus	-6													-8							-9			
Staphylococcus																						-9		
Weissella																								
Enterobacter																					-7			
Ralstonia						-4	-4																	
Rest							-5	-8	-10					-4	-7	-7					-9	-8	-8	-8
Unclassified						-7	-5		-10		-7	-8	-7	-4	-8			-6	-6			-9	-7	
Major *Anoxybacillus* and *Geobacillus* species	**2log of relative abundance**																							
A. fla.				-8	-2	-4	-2	-1	-1	-1	0	0	0	-3	-1	0	0	-2	-2	-1		-3		
G. ste.									-8		-2	-3	-3	-6				-1	-1	-1	-5	0	0	0
G. the.																								

Figure 2.2 Microbial inventory of thermophilic bacteria in a static biofilm model. Standard milk (M1 and M2) or sterile milk inoculated with industrial fouling from a whey evaporator (i7) was cultured O/N at 55 or 65°C in plastic culture wells containing steel coupons (static biofilm model) in order to study thermophile enrichment in the medium (planktonic) and on steel and plastic surfaces (biofilm). Genomic DNA isolated from raw milk and fouling (raw), milk, steel or submerged steel and plastic well-wall fractions were analysed by mass-sequencing 16S-genotyping (500-2000 sequences per sample). Grey values, numbers and abbreviations are the same as in Fig 1. M: media; subm.s: submerged-stainless-steel-surface attached biofilm; St. s: standing-stainless-steel-surface attached biofilm; P: plastic-surface attached biofilm

2.3.3 Isolation and characterization of novel thermophilic biofilm and spore formers.

In this study, approximately 200 strains were isolated by selection of colonies from TSA plates incubated at 55°C after inoculation with fouling samples from the dairy concentrate production line. Twenty strains with morphologically different colonies were characterized with respect to their 16S-rRNA genotype, growth rates and biofilm-forming performance (data not shown). All culturable isolates from the standard milk (M1 and M2) were typed as *A. flavithermus*, except for two isolates of *Bacillus licheniformis*; the isolate from the dairy concentrate end product was typed as *G. stearothermophilus*, and those from the fouling samples show a higher variety, including *A. flavithermus*, *G. stearothermophilus*, and *G. thermoglucosidans*. The occurrence of the thermophilic isolates *A. flavithermus*, *G. stearothermophilus*, and *G. thermoglucosidans* in raw and enriched samples was confirmed by an exact match to the 16S-rRNA sequences of these species (see the three bottom rows in Figure 2.1). Most isolates showed significant biofilm formation at 60°C and 70°C, as derived from the OD-values from crystal violet-staining of surface-attached biomass after growth. On the basis of their ability to efficiently form biofilms in a laboratory model system, the isolates *A. flavithermus* TNO-09.006, *G. stearothermophilus* TNO-09.008 and *G. thermoglucosidans* TNO-09.020 were selected and their species identity was confirmed by DNA-DNA hybridizations with genomic DNA isolated from the three corresponding type strains. The percentage of relatedness to the type strain matched the > 70% criterion for the assignment of all three bacterial species (Table 2.1). The full genomes sequences of the three strains were determined (Zhao et al., 2012, Caspers et al., 2013) and the strains were characterized regarding their temperature growth range and optimum, sporulation efficiency, and spore heat resistance (Tables 2.1 and 2.2). In addition, their ability to sporulate was confirmed by microscopic examination showing the formation of phase bright endospores at the poles (Figure S2.3). Heat-resistant spores were enumerated in culture-medium and stainless-steel biofilm fractions during a cultivation experiment of 30 hours, indicating an increase in the number of spores over time in both fractions up to 10^5 CFU's per ml (Figure S2.4). Interestingly, the growth at high temperatures was observed over a temperature window of 19°C for all three thermophilic species, including 43-62°C, 48-67°C, and 50-69°C for *A. flavithermus*, *G. stearothermophilus* and *G. thermoglucosidans*, respectively. The preference for *Geobacillus* to grow at relatively high temperatures is reflected in the enrichment experiments, showing accumulation at 65°C rather than

55°C degrees (Figures 2.1, 2.2). The *Geobacillus* produce heat-resistant spores, with decimal reduction values ranging from 18-20 min at 110°C, whereas the D-value of *Anoxybacillus* is only 2 min. at this temperature (Table 2.2). The ability to efficiently form biofilms and generate highly heat resistant spores with high efficiency under lab conditions renders *G. thermoglucosidans* an interesting model organism. Biofilm forming behaviour of the TNO-09.020 isolate on a stainless-steel coupon in the static biofilm model system with Tryptone-based medium was analysed microscopically. Examination of biofilms stained with Auramine indicated the presence of the multicellular structures that predominantly formed at the air-liquid interphase (Figure 2.3A). The bacterial spores formed within these biofilms appeared more or less randomly distributed (Figure 2.3 B, C).

Table 2.1 Typing and growth characteristics of selected model strains. Species assignment of model strains was confirmed by DNA-DNA hybridizations with reference strains from the LMG culture collection. The T_{min} (°C) and T_{max} (°C) are defined as the maximum and minimum temperatures at which still growth could be detected under the conditions used (see Materials and Methods). The T_{opt} (°C) is the temperature at the highest growth rate, which is expressed in the doubling time t_D (min).

Strain ID	*DNA-DNA hybridization (% homology)*	*Growth temperature range*			*t_D (min)*
		Tmin (°C)	Tmax (°C)	Topt (°C)	
TNO-09.006	*Anoxybacillus flavithermus* LMG 18397T (75 ± 8 %)	43	62	57	52
TNO-09.008	*Geobacillus stearothermophilus* LMG 6939T (86 ± 9 %)	48	67	61	35
TNO-09.020	*Geobacillus thermoglucosidans* LMG 7137T (88 ± 13 %)	50	69	60	32

Table 2.2 Sporulation efficiency of thermophilic spore formers and heat resistance of their spores. The sporulation efficiency was expressed as the number spores (CFU after heat inactivation) divided by the total number of bacterial cells and spores (CFU before heat inactivation). The D-values are expressed in minutes of treatment at indicated temperature for a 10-fold CFU reduction; the z-values are expressed in °C temperature increase required for a 10-fold reduction of the D-value; the calculations are described in detail in Material and Methods and Figure S2.2.

Strain ID	*Sporulation (on NA++ agar plates)*	*Heat resistance of spores*	
	Average sporulation efficiency (%)	D_{110} (min)	z-value (°C)
Anoxybacillus flavithermus TNO-09.006	77 ± 40 %	2	13
Geobacillus stearothermophilus TNO-09.008	38 ± 31 %	18	11
Geobacillus thermoglucosidans TNO-09.020	91 ± 3%	20	8

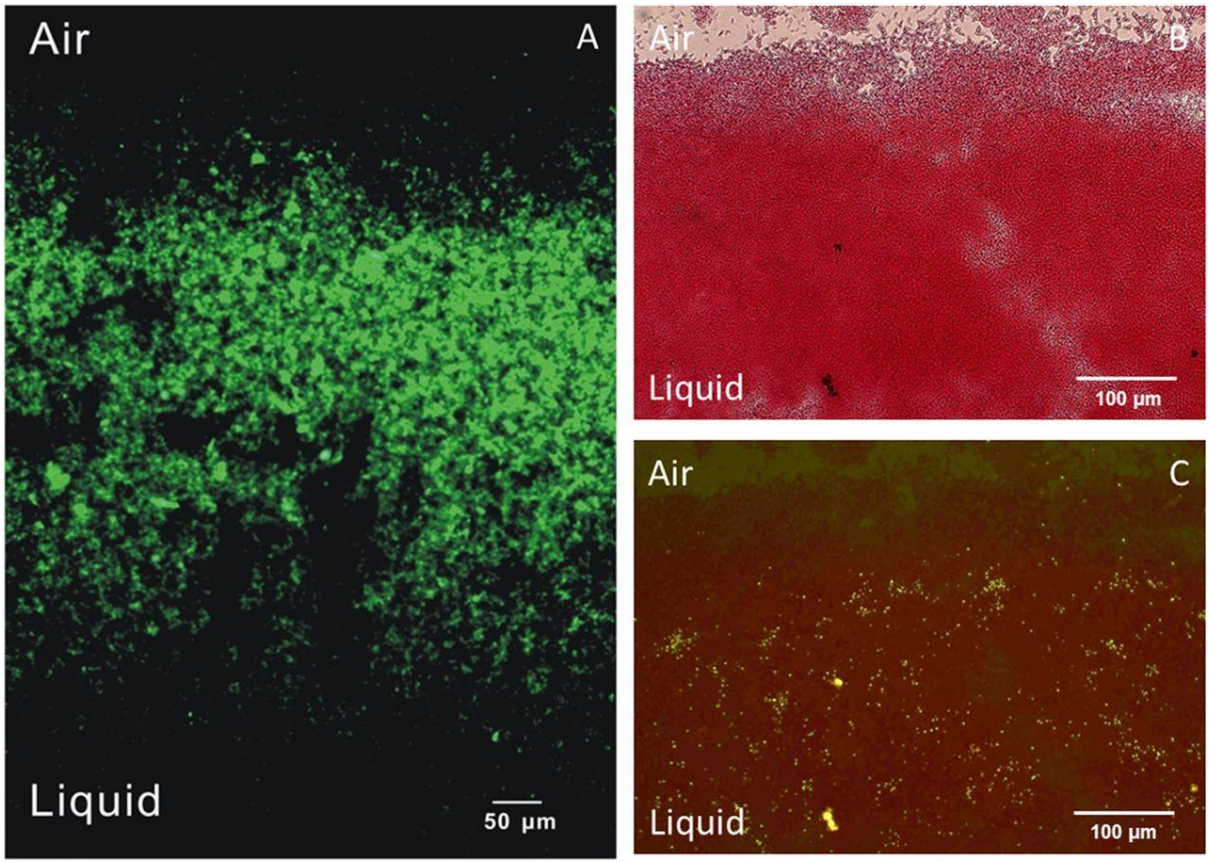


Figure 2.3 *Geobacillus thermoglucosidans* TNO-09.020 biofilms at the air-liquid interface. A) Fluorescence microscopy image of Auramine-stained biofilm on a standing stainless-steel coupon after 10 hours of batch cultivation at 65°C. **B)** Bright field and **C)** fluorescence microscopy image of Auramine and Safranine-stained biofilm on a standing glass coupon after 16 hours of batch cultivation at 65°C.

2.3.4 Growth-dependence of *Geobacillus thermoglucosidans*.

The selected species *A. flavithermus, G. steathermophilus,* and *G. thermoglucosidans* were further characterized for their ability to grow on different nutrient plates. Interestingly, the *G. thermoglucosidans* strains TNO-09.020 and TNO-09.023 were not capable of growing on milk-plates. However, they were capable of growing on plates containing casein, the major protein component of milk, if the casein was proteolytically digested (data not shown). Therefore, we hypothesized that *G. thermoglucosidans* is dependent on the proteolytic activity of other bacteria for growth in milk. To test this, we analysed growth of *G. thermoglucosidans* TNO-09.020 and *A. flavithermus* TNO-09.006 in a cell culture insert setup that enables cultivation of the two strains separated by a permeable membrane. This membrane allows the diffusion of enzymes and small organic molecules between the two compartments. *A. flavithermus* TNO-09.006 readily started growth after 3 hours and continued growing until approximately 12 hours in the presence and absence of TNO-09.020, after which the CFU number started to decrease (Figure 2.4). As expected, *G. thermoglucosidans* TNO-09.020 inoculated in milk did not show any growth, with CFUs remaining below 3 log units per ml. However, when *G. thermoglucosidans* TNO-09.020 was inoculated in the presence of *A. flavithermus* TNO-09.006, growth started after a long lag time of ≥ 12 hours, reaching a CFU value of approx. 5 log units after 24 hours (Figure 2.4). Clearly, *G. thermoglucosidans* TNO-09.020 is dependent on *A. flavithermus* TNO-09.006 for growth in the milk medium. The second *G. thermoglucosidans* strain isolated in this study, TNO-09.023 was also tested in this cell culture insert setup and showed similar behaviour (data not shown).

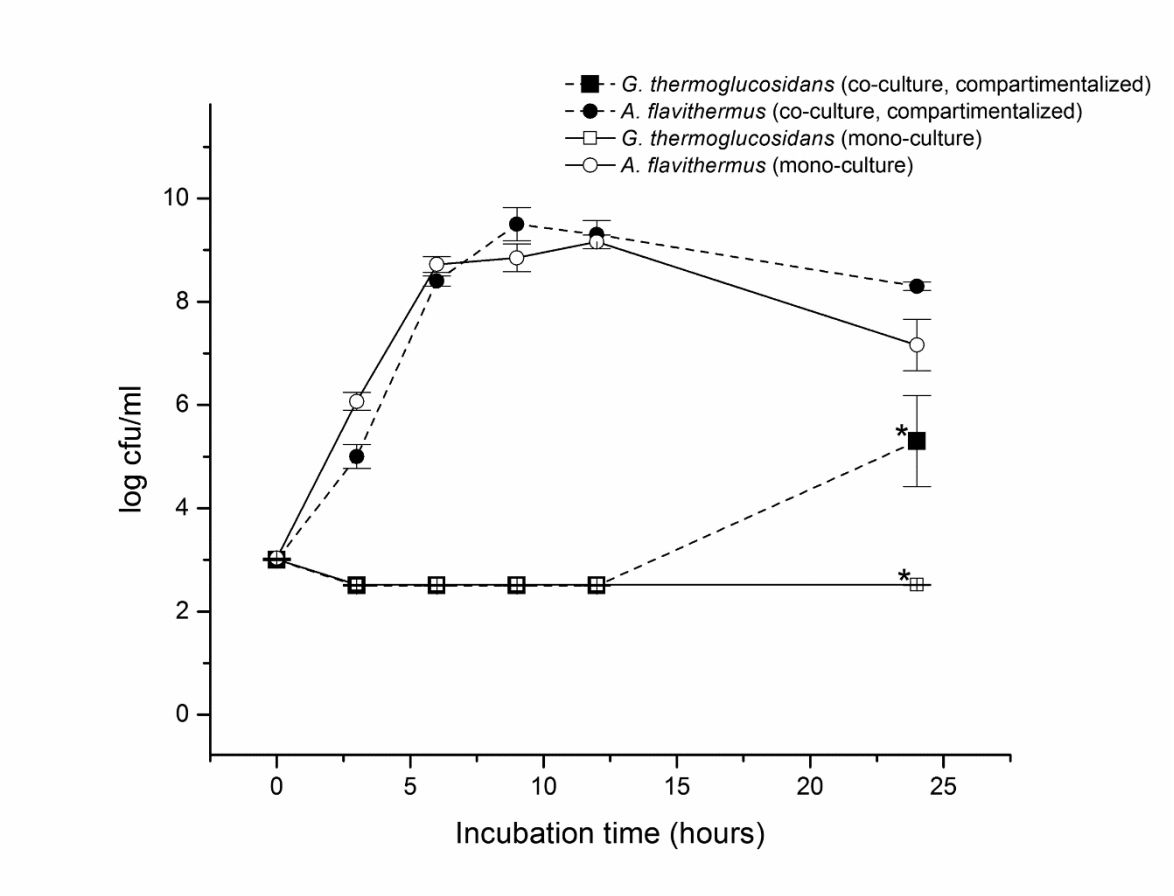


Figure 2.4 Compartmentalized growth of *Geobacillus thermoglucosidans* TNO-09.020 and *Anoxybacillus flavithermus* TNO-09.006. Graphical representation of bacterial counts from two strains in a compartmentalized growth experiment in UHT skim milk with the BD FalconTM Cell Culture insert system, allowing growth of strains in two compartments separated by a permeable membrane that permits diffusion of media components (pore size 0.4 μm). (■) Cell counts of TNO-09.020 (with TNO-09.006 in the other compartment); (●) Cell counts of TNO-09.006 (with TNO-09.020 in the other compartment); (□) Cell counts of TNO-09.020 in the absence of TNO-09.006; (○) Cell counts of TNO-09.006 in the absence of TNO-09.20. The bacterial cultures were enumerated at 6 different time points, each point represents the mean and standard deviation, of triplicate measurements. The * indicates a significant difference for growth (log CFU) of *G. thermoglucosidans* TNO-09.020 in the presence or absence of *A. flavithermus* TNO-09.006 in the other compartment (T-test, $P < 0.02$).

Next, a co-culture experiment in milk was conducted with *G. thermoglucosidans* TNO-09.020 and *A. flavithermus* TNO-09.006, and biofilm development was monitored by determining the total number of viable cells in the biofilm attached to stainless steel coupons, and the number of colony forming units of *G. thermoglucosidans* TNO-09.020 was selectively determined as they appear as white colonies on TSA-X-Gal plates at 55°C, in contrast to colonies of the *A. flavithermus* TNO-09.006 strain that appear blue on TSA X-gal plates, as a result of its galactosidase activity (Figure 2.5). In agreement with the results of the

compartmentalized growth experiment, the strain TNO-09.020 is only able to form biofilms when TNO-09.006 is present, and the number of colony forming units of TNO-09.020 in the biofilm reached a level of approximately 10^5 CFU/biofilm after 24 hours, and approximately 10^7 CFU/biofilm fraction after 48 hours (Figure 2.5). The TNO-09.006 strain grows well in milk in the absence of TNO-09.020, reaching approximately 10^7 CFU/ml in the milk medium and 10^5 CFU/biofilm after 8 hours. However, no CFU's of this strain could be detected after 48 hours in either biofilm or milk medium when TNO-09.020 was present (Figure 2.5).

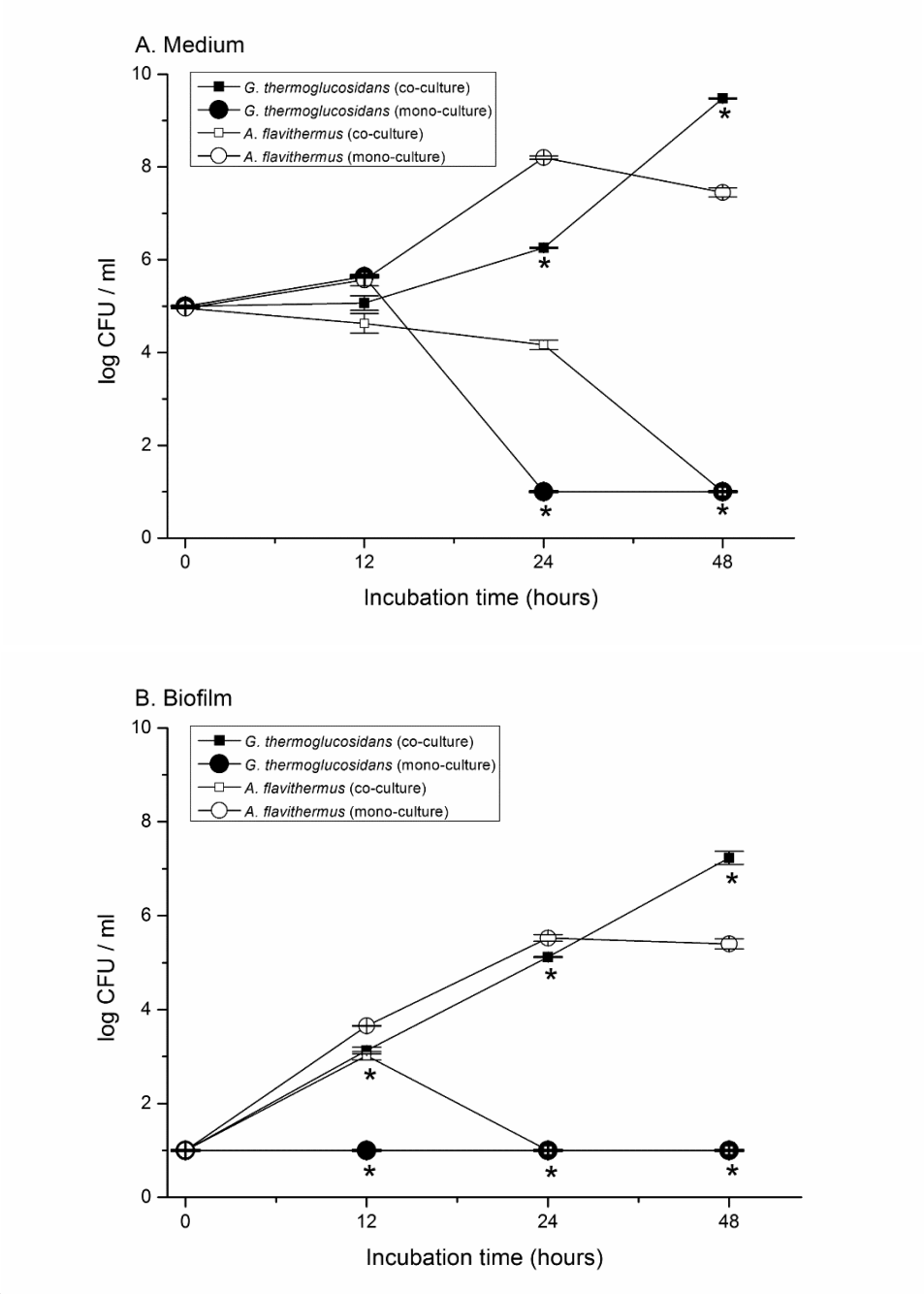


Figure 2.5 Colony forming units of planktonic cells and biofilms in co-culture of thermophiles. Figures A and B show the results of a total of three batch cultivation experiments in milk of the standing steel biofilm model system, including one co-culture and two monocultures. A) Bacterial cell counts in the 3-ml milk medium fraction of a co-culture of planktonic cells of *Geobacillus thermoglucosidans* TNO-09.020 (■), and *Anoxybacillus flavithermus* TNO-09.006 (□). For reference the results of monocultures of TNO-09.020 (●), and TNO-09.006 (○) in milk were plotted. B) Bacterial cell counts of the biofilm attached to stainless steel in a co-culture of TNO-09.020 (■), and TNO-09.006 cells (□). For reference the results were plotted of biofilms obtained by monocultures of TNO-09.020 cells (●), and TNO-09.006 cells (○). The bacteria were enumerated at 4 different time points, each bar represents the mean and the error bar the standard deviation, from two experiments of triplicate measurements. The * indicates a significant difference for growth (log CFU) of *G. thermoglucosidans* TNO-09.020 in the presence or absence of *A. flavithermus* TNO-09.006 during co-culture (T-test, $P < 0.01$).

2.4 Discussion

In this study we developed and applied a number of novel approaches to study growth and biofilm forming capacity of spore formers associated with the dairy industry. This work included a cultivation-independent approach to study contaminants in milk, factory fouling samples, end products and enrichments thereof. We have grown biofilms with milk samples as an inoculum and screened factory isolates for their ability to form biofilms under laboratory conditions in multi-well plates. We have characterized three of these thermophilic biofilm forming isolates and their spores in detail. Three major findings resulted from this work: (i) dairy processing environments harbour species-rich microbial communities, (ii) the thermophilic spore formers studied, preferentially form biofilms at air-liquid interfaces, and (iii) the thermophilic spore former *Geobacillus thermoglucosidans* depends on other thermophilic species present for growth and biofilm formation in milk-based media.

The results revealed a wide diversity of genera in the processing lines. Fouling samples taken from the processing line where high temperatures were applied, were not dominated by thermophilic spore formers, but significant numbers mesophilic bacteria were identified. The mass sequencing applied here detects DNA molecules, encoding 16S-rRNA molecules, thus not necessarily viable bacteria, so this may lead to overestimation of the viable microbiota present (see also reference (Ronaghi 2001)). Dairy-associated microbiota shows remarkable diversity, as previously reported and reflected in the assignment of dairy farm isolates to seven spore-forming genera, i.e., *Aneurinibacillus, Bacillus, Brevibacillus, Geobacillus, Paenibacillus, Ureibacillus,* and *Virgibacillus* (Scheldeman et al., 2005). In the current study, we confirmed the presence of the spore forming genera *Aneurinibacillus*, *Bacillus, Brevibacillus, Geobacillus* and *Anoxybacillus*.

Enrichment at 55°C of the fouling materials resulted in selection of thermophilic genera and spore formers. After enrichment at 65°C, the thermophilic genera (spore formers or non-spore formers) are dominant in most samples. The fact that these thermophiles are not always detected by this method in the fouling samples, is because their numbers were below the detection limit of the method used. The enrichment of fouling samples in biofilm model systems shows that the predominant spoilage genera associated with biofilm formation are *Geobacillus* and *Anoxybacillus*, of which species have been isolated from milk powders and dairy concentrate

processing factories (Flint et al., 1997b). According to the bar-coded 16S-amplicon sequencing data from this study, a number of other thermophilic genera are present in the fouling samples, even after enrichments, including *Thermus*, *Brevibacillus* and *Aneuribacillus*, but which do not appear among the cultured isolates; possibly these species are easily outcompeted by *Anoxybacillus* and *Geobacillus* on TSB plates at 55°C.

Concerning the abiotic conditions which control biofilm formation of the thermophiles studied here, we identified in this study no evident correlation between composition of surface-attached microbiota and the nature of the surface, including steel and plastic. However, a clear difference in preferred environment for biofilm formation of microbial genera was identified, as *Anoxybacillus* and *Geobacillus* preferentially reside at air-liquid interface, whereas *Pseudomonas* accumulated at the surface of submerged steel. We hypothesize that the oxygen concentration may play a crucial role in selective accumulation of bacteria and spores on the stainless-steel surface at the air-liquid interface. This suggests that biofilms of thermophilic spore formers and associated spores may particularly develop at elevated temperature and in industrial piping systems that are only partly filled and as a result are exposed to oxygen during operation.

Finally, we present data in this study suggesting that the *G. thermoglucosidans* strain, which produces the most thermostable spores, is dependent on proteolytic strains for outgrowth in the dairy environment. Several observations support this: the *G. thermoglucosidans* was not enriched from the industrial milk or fouling samples in milk medium, probably due to its long lag phase before outgrowth (bottom row, Figure 2.2). In addition, the *G. thermoglucosidans* strain TNO-09.020 and the strain TNO-09.023 cannot readily grow or form biofilms in undigested casein or milk medium. However, they grow well and form biofilms in pre-digested casein or milk medium or alternatively, when the proteolytic strain *A. flavithermus* TNO-09.006 was also present in undigested casein or milk medium. Although there is no evidence for a mutual relationship between TNO-09.020 and TNO-09.006, our observation may bear some resemblance to that of the yogurt consortium, where the proteolytic activity of *L. bulgaricus* results in the supply of amino acids for *S. thermophiles* (Sieuwerts et al., 2008a). The ecology and interrelationship between the selected isolates will be elucidated using gene-trait matching approaches based on whole genome sequence information (Zhao et al., 2012, Caspers et al., 2013). Such information would be relevant because our results suggest that the presence of

proteolytic microorganisms in the dairy concentrate production line may contribute to the diversity and spore load of specific thermophiles in end products.

2.5 Experimental procedures

Sampling, culturing and enrichment.

A number of fouling sites were selected along dairy concentrate production lines for the bar-coded 16S-amplicon sequencing analysis of the microbial flora (Table S2.1). Samples of standard milk, fouling material isolated from the processing line and final products were collected. Standard milk was flash frozen by dripping in liquid nitrogen. The frozen milk pellets were stored at -80°C. Fouling samples were scraped from pipelines, dispersed 1:1 (w/v) in sterile antifreeze Microbank medium (Pro-Lab Diagnostics, Canada) and stored at -80°C. Final products were dissolved in sterile water (2 -10% w/v) and stored at -80°C.

Viable counts were carried out for all samples analysed with bar-coded 16S-amplicon sequencing. Growth analysis of strains and colony forming unit (CFU) determination were performed on Tryptone Soy Broth (TSB) or Tryptone Soy Agar (TSA) (Tritium Microbiologie, The Netherlands). All CFU determinations in this study were performed by plating 80 µl on TSA plates followed by overnight (O/N) incubation at 30°C (for non-thermophilic CFU determination) or at 55°C (for thermophilic CFU determination). Dilution series were made in PPS (0.1% peptone, 0.9% NaCl). Thermophilic aerobic spore counts at 55°C were similarly determined after pre-treatment of the samples at 100°C for 30 minutes to eliminate vegetative cells and to activate thermo-resistant spores (Scott et al., 2007). The CFU determinations of samples after enrichment at 55 or 65°C were obtained by plating on TSA and incubating at the respective enrichment temperatures.

Thermophilic enrichment was carried out by O/N culturing of 50 µl of a sample in 2 ml TSB at 55 or 65°C followed by inspection for growth by increase of optical density (no growth (-), little growth (+/-) or outgrowth (+)). Initially, enrichment was performed at 55°C in TSB, since it is a classical method to determine dairy thermophilic bacterial loads (Scott et al., 2007). However, to prevent extensive overgrowth of mesophilic species at 55°C, enrichment at 65°C was included as well, in order to facilitate selection of the thermophilic species.

Biofilm model systems.

In order to study biofilm formation by thermophilic spore forming dairy isolates on a laboratory scale, a standing steel biofilm model system was developed. This biofilm system included a sterile, vertically standing, 14 x 14 mm stainless steel coupon (P. 316 grade) in a well of a sterile 24-well-plate (Corning, The Netherlands). The plate was incubated in a tight plastic bag containing a wetted paper towel to limit evaporation of the culture media (Figure S2.1). In addition, a submerged steel biofilm system was developed, consisting of a steel coupon lying horizontally on the well bottom of a 24-well-plate.

For enrichment in the static biofilm models, 2 ml industrial milk samples (standardized milk with a standardized composition), and 50 µl of a fouling sample in 2-ml heat sterilized milk (120°C, 20min), were cultured O/N at 55 or 65°C (non-shaken) in the separately wells. After O/N incubation the various fractions (including culture medium, polystyrene well wall and coupon surfaces) were harvested and directly subjected to CFU determinations or stored at -80°C until DNA isolation (see below). The medium fractions were directly harvested from the culture wells. The metal coupons and empty wells were gently rinsed with sterile PPS (3 x 3 ml) and separately swabbed (coupons were first transferred to clean sterile wells) with sterile cotton swabs, each in 2 x 150 µl sterile PPS.

Air-liquid interface biofilms of industrial isolates were studied as well by the use of vertical, sterile 15 x 15 mm glass coupons (cut from standard microscopy object glasses) in 12-well-plates. After O/N cultivation, the glass coupons were gently washed with demineralized water and fixed by drying for 10 min at 60°C. Culture wells were washed with sterile water (3x 3ml/well) and fixed by incubation for 10 minutes at 60°C. Water washed and air-dried coupons or culture wells were used for Crystal Violet (CV) staining (5 min 1% w/v CV, 3 x water washing). CV-stained coupons were analysed by light microscopy. CV-stained culture wells were distained for 5 minutes at room temperature with 33% acetic acid (1.1 x volume originally cultured in well) and the OD between 580 – 600 nm was measured with a plate well-reader (TECAN, Switzerland) to determine the amount of CV-stainable biofilm.

Fluorescence microscopy

Coupons were incubated for 2 min with 0.1 % Auramine (Merck, The Netherlands) for visualizing the attached cells (Bartholomew et al., 1965). Spores were stained in the water-washed and air-dried coupons by the Auramine-Safranine method (Bartholomew et al., 1965). Briefly, stainless steel or glass coupons were incubated for 2 min with 0.1 % Auramine (Merck, The Netherlands), water-washed, incubated for 1 minute with 0.25 % Safranin (BD Biosciences), water-washed, and air-dried for 10 min at 60°C; Bright field (glass coupons) and fluorescence microscopy (Zeiss, Axio Observer Z1, filter set "Endow GFP" Ex BP 470/40, BS FT 495, EM BP 525/55) was performed directly on the stained, dried and covered-glass coupons.

DNA isolation

Genomic DNA (gDNA) was isolated from the (enriched) fouling samples and fractions from the static biofilm model. The bacterial samples (50 – 200 µl) were added to a 1.5 ml screw-cap Eppendorf tube with 0.3 g zirconium-silica beads (0.1 mm bead size), 800 µl phenol (pH 8.0) and 400 µl Agowa buffer without detergent. Next, the samples were homogenized with a BeadBeater Bio101 (Biospec Products, USA) for 2x 45 seconds with a 30 seconds interval of cooling on ice and spinned down for 10 minutes at 10,000 g. The upper, aqueous phase was taken and extracted with the AGOWA mag Mini DNA Isolation Kit (AGOWA, Germany), eluted in 45 µl AGOWA BL-buffer. Quality and quantity of gDNA was determined on agarose gel and by Nanodrop ND-1000 (NanoDrop Technologies, USA).

Bar-coded 16S-amplicon sequencing.

Mass sequencing was performed as described earlier (Nocker et al., 2010). Briefly, barcoded 16S rRNA fragments were amplified with forward 785F (5'-gcctccctcgcgccatcagggattagatacccbrgtagtc-3') and reverse primer 1175R (5'-gccttgccagcccgctcagnnnn-acgtcrtccccdccttcctc-3'). Pyrosequencing of equimolar mixes of 24 amplicon pools was performed by Keygene N.V. (The Netherlands) using the Roche Genome Sequencer-20 (GS-20) and FLX 454 pyrosequencing technology yielding on average 1145 reads per amplicon pool (standard deviation 456; minimum 277; maximum 2583). The FASTA format sequences and corresponding quality scores were extracted from the .sff data files generated by the GS-FLX system using the GS Amplicon software package (Roche, Branford, CT). Sequence data was processed

using modules implemented in the Mothur v. 1.25.0 software platform (Schloss, Westcott et al., 2009). Sequences were binned by sample of origin by the unique barcodes sequences in each amplicon pool. For further downstream analyses, barcodes and primer sequences were trimmed and low-quality reads were excluded from the analyses. The data set was simplified by using the "unique.seqs" command to generate a non-redundant (unique) set of sequences. Unique sequences were aligned using the "align.seqs" command and an adaptation of the Bacterial SILVA SEED database as a template (available at: http://www.mothur.org/wiki/Alignment_database). In order to ensure that we were analysing comparable regions of the 16S rRNA gene across all reads, sequences that started before the 2.5-percentile or ended after the 97.5-percentile in the alignment were filtered. Sequences were denoised using the "pre.cluster" command. This command applies a pseudo-single linkage algorithm with the aim of removing sequences that are likely due to pyrosequencing errors (Huse et al., 2010). Potentially chimeric sequences were detected and removed using the "chimera.slayer" command (Haas et al., 2011). High quality aligned sequences were classified using the RDP-II naïve Bayesian Classifier implemented into the Mothur platform. Aligned sequences were clustered into OTUs (defined by 97% similarity) using the average linkage clustering method. Typing to the level of *Anoxybacillus* and *Geobacillus* species was performed using the most abundant unique sequence of these OTUs in the Seqmatch tool of RDP. Relative abundance of genera and species were calculated as fractions of the total reads per sample.

Typing of industrial isolates.

A set of around 100 bacterial isolates (single colonies) were obtained from raw and enriched samples. These isolates were cultured to determine growth and biofilm formation at temperatures of 30, 60, 65 and 70 °C in TSB medium. Of all 100 isolates tested, 20 isolates were able to grow (OD>0.08) and form biofilms (OD>0.11) at 60°C and 70°C. DNA of 20 industrial isolates was isolated as described above. For typing of industrial isolates, the 16S rRNA gene region 8-1408 was PCR-amplified from gDNA using forward (F) and reverse (R) primers 8F (5'-agagtttgatchtggytcag-3') and 1408R (5'-tgacgggcggtgtgtacaa-3'). PCR amplicons were purified and bidirectionally sequenced by GATC-biotech AG, Germany, using primers 8F, 27F (5'-agagtttgatcmtggctcag-3'), 1408R, and 1392R (5'-acgggcggtgtgtgtrc-3'). The sequences were typed at the species level with the RDP SeqMatch tool (http://rdp.cme.msu.edu/) (Cole et al., 2009), and by selection of the best hit reported

from the RDP database (type strains, non-type strains, unculturable strains and isolates with a size of > 1200 bp and of good quality). Growth curves of selected model strains from thermophilic spore forming species *A. flavithermus* TNO.09-006*, G. stearothermophilus* TNO.09-008*, G. thermoglucosidans* TNO.09-020 were determined in TSB medium at various temperatures by the use of a temperature gradient in a PCR machine (100 µl of culture per well; range: 38 - 74°C; model DNA Engine Tetrad, PTC-225), and multiple OD measurements during cultivation (each 20 min; 50 µl/well of 384-well plate, Tecan F500 plate reader at 600 nm). Exponential growth rates µ and doubling time t_D were calculated with the equations $\mu = \ln[d(OD)]/dt$ and $t_D = \ln 2/\mu$, respectively. The T_{min} and T_{max} are defined as the minimum and maximum temperature at which growth could be detected under the conditions used. The T_{opt} is defined as the temperature with the highest growth rate, as expressed in doubling time (t_D).

DNA-DNA hybridizations.

Genomic DNA was extracted from pure cultures according to a modification of the procedure by Gevers et al. (2001). Hybridizations were performed in the presence of 50% formamide at 39°C according to a method adapted from Ezaki et al. (1990). The DNA-DNA hybridization percentages reported are the means of at least 6 hybridisations.

Assessment of casein degrading activity.

The selected model strains from thermophilic spore forming species *A. flavithermus* TNO.09-006*, G. stearothermophilus* TNO.09-008*, G. thermoglucosidans* TNO.09-020 were examined for their capability to utilize milk protein. Media included casitone plates (25g/L casitone, 5 g/L NaCl, 2.5g/L K_2HPO_4, 1.5% (w/v) agar), Tryptone plates (10 g/L Tryptone, 5 g/L NaCl, 2.5 g/L K_2HPO_4, 1.5% (w/v) agar), Ca-caseinate plates (1.25% (w/v) containing Ca-caseinate (Friesland-Campina, NL) and 0.8% agarose, and pancreatin-digested Ca-caseinate plates. The latter plates were prepared by digestion of Ca-caseinate (1.25% (w/v)) with 10 mg/ml pancreatin (Sigma P3292) for 3 hours at 37° C, followed by heat inactivation at 100° C for 10 min. Plates were inoculated by transfer of bacterial cells taken from a TSA plate.

Heat resistance of spores.

A suspension prepared from an overnight grown plate culture was spread on NA++ plates (Nutrient agar with supplementation of analytical grade 1.13 mM $CaCl_2$ and 0.99 mM $MnSO_4$) and incubated for 2 days at 55°C. Bacterial lawns containing spores were harvested, and washed with sterile demi-water as described (Kort et al., 2005). This water washing procedure was repeated three times in order to obtain pure spore suspensions. The spore suspensions were stored at -20°C. The heat inactivation kinetics of the spores were determined as follows: Micropipettes of 100 µl were filled with spore suspensions and both ends of the micropipettes were sealed by heating. Micropipettes were incubated within a time window at serial temperatures above 100°C in an oil bath filled with glycerol. The spore suspension was diluted 100 times in PPS, series of dilutions were made, and plated on to TSA plates. The D-values of the spore batches were derived from plots with log CFU versus incubation time by fitting a log-linear model with tail to the data and the z-values were calculated by plotting the logD value against the temperature and performing a linear regression (Figure S2.2).

Compartmentalized growth experiments.

The determination of growth dependencies in ultra-heat treated (UHT) skim milk was performed using the BD Falcon™ Cell Culture insert system. This system allows growth of strains in two compartments separated by a permeable membrane that permits diffusion of media components (pore size 0.4 µm). Both the well and the cell culture insert were filled with 3 ml UHT skim milk and inoculated with approximately 4 x 10^3 CFU (*A. flavithermus* TNO-09.006) and 3 x 10^3 CFU (*G. thermoglucosidans* TNO-09.020), respectively. As controls, wells were filled with 3 ml UHT skim milk and inoculated with either *A. flavithermus* TNO-09.006 or *G. thermoglucosidans* TNO-09.020 with the same amounts of CFUs. All measurements were performed in triplicate. Following inoculation, the 6-well plates were wrapped in a plastic bag and sealed in order to prevent evaporation and incubated at 65°C at 50 RPM. Sampling was performed at 3, 6, 9, 12, and 24 hours. CFU counts of each fraction were determined by serial dilutions poured in TSA.

Formation of biofilms on stainless steel coupons in co-cultures.

Stainless steel coupons were placed in the wells of a polystyrene 12-well plate (Falcon, Becton Dickinson, France). The wells were half filled with 3ml of UHT skim milk, which was inoculated with 1% (v/v) overnight culture of either a mixture or single strain of *A. flavithermus* strain TNO-09.006, and *G. thermoglucosidans* TNO-09.020. The plates were wrapped with plastic bags and wet tissues and incubated for 12h, 24h and 48 h at 65°C. The total number of bacterial cells present in the milk or attached to the surface of the stainless-steel coupon, was determined by CFU counting. The coupons were washed in sterile UHT skim milk 3 times. Then they were placed in 50-ml tubes filled with 3 ml UHT skim milk and 0.5 g of glass beads (100 µm diameter). Tubes were mixed by vortex for 1 min to detach the cells from the stainless-steel coupon. Serial dilutions were made and plated on TSA-X-Gal plates for counting after 24h of incubation at 55°C. Biofilm formation was assessed in triplicate in two independent experiments.

2.6 Acknowledgments

We thank Carlo Brouwer, RuiRui Liu and Rob Leer for expert assistance. This work was supported by the Top Institute Food and Nutrition (TIFN).

2.7 References

Bartholomew J. W., M. D. Lechtman and H. Finkelstein. 1965. Differential spore and lipid staining at room temperature by use of fluorescent dye. Journal of Bacteriology. 90:1146 –1147.

Burgess S. A, D. Lindsay and S. H. Flint. 2010. Thermophilic bacilli and their importance in dairy processing. International Journal of Food Microbiology. 144:215–225.

Burgess S. A., J. D. Brooks, J. Rakonjac and K. M. Walker and S. H. Flint. 2009. The formation of spores in biofilms of *Anoxybacillus flavithermus*. Journal of Applied Microbiology. 107:1012–1018.

Caspers M. P. M., J. Boekhorst, T. Abee, R. Siezen and R. Kort. 2013. Complete genome sequence of *Anoxybacillus flavithermus* TNO-09.006, a thermophilic sporeformer associated with a dairy-processing environment. Genome Announcement. 1:e00010-13. doi:10.1128/genomeA.00010-13.

Cole J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity and J. M. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research. 37:D41–D145.

De Jonghe V., A. Coorevits, J. Vandroemme, J. Heyrman, L. Herman, P. De Vos and M. Heyndrickx. 2008. Intraspecific genotypic diversity of *Bacillus* species from raw milk. International Journal of Dairy Technology. 18:496 –505.

Delbès C, L. Ali-Mandjee and M. C. Montel. 2007. Monitoring bacterial communities in raw milk and cheese by culture-dependent and -independent 16S rRNA gene-based analyses. Journal of Applied & Environmental Microbiology. 73:1882–1891.

Ercolini D, F. Russo, I. Ferrocino and F. Villani. 2009. Molecular identification of mesophilic and psychrotrophic bacteria from raw cow's milk. Food Microbiology. 26:228–231.

Ezaki T., S. Adnan and M. Miyake. 1990. Quantitative microdilution plate hybridization to determine genetic relatedness among bacterial strains. Nihon Saikingaku Zasshi 45:851–857. (In Japanese.)

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

Gevers D., G. Huys and J. Swings. 2001. Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. FEMS Microbiology Letters. 205:31–36.

Haas B. J., D. Gevers, A. M. Earl, M. Feldgarden, D. V. Ward, G. Giannoukos, D. Ciulla, D. Tabbaa, S. K. Highlander, E. Sodergren, B. Methé, T. Z. DeSantis, The Human Microbiome Consortium, J. F. Petrosino, R. Knight and B. W, Birren. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research. 21:494 – 504.

Huse S. M., D. M. Welch, H. G. Morrison HG and M. L. Sogin. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology. 12:1889 –1898.

Kort R., A. C. O'Brien, I. H. M. Van Stokkum, S. J. C. M. Oomes, W. Crielaard, K. J. Hellingwerf and S. Brul. 2005. Assessment of heat resistance of bacterial spores from food product isolates by fluorescence monitoring of dipicolinic acid release. Applied and Environmental Microbiology. 71:3556 –3564.

Lafarge V., J. C. Ogier, V. Girard, V. Maladen, J. Y. Leveau, A. Gruss, A. Delacroix-Buchet. 2004. Raw cow milk bacterial population shifts attributable to refrigeration. Applied and Environmental Microbiology. 70:5644 –5650.

Manaia C.M. and E. R. Moore. 2002. *Pseudomonas thermotolerans* sp. nov., a thermotolerant species of the genus *Pseudomonas* sensu stricto. International Journal of Systematic and Evolutionary Microbiology. 52:2203–2209.

Nocker A., T. Richter-Heitmann, R. Montijn, F. Schuren and R. Kort. 2010. Discrimination between live and dead cells in bacterial communities from environmental water samples analysed by 454 pyrosequencing. International Microbiology. 13:59–65.

Ronaghi M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Research. 11:3–11.

Scheldeman P., A. Pil, L. Herman, P. De Vos and M. Heyndrickx. 2005. Incidence and diversity of potentially highly heat-resistant spores isolated at dairy farms. Applied and Environmental Microbiology. 71:1480 –1494.

Schloss P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, F. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, C. F. Weber. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 75:7537–7541.

Scott S. A., J. D. Brooks, J. Rakonjac, K. M. R. Walker and S. H. Flint. 2007. The formation of thermophilic spores during the manufacture of whole milk powder. International Journal of Dairy Technology. 60:109 –117.

Sieuwerts S., F. A. de Bok, J. Hugenholtz, J. E. van Hylckama Vlieg. 2008. Unraveling microbial interactions in food fermentations: from classical to genomics approaches. Applied and Environmental Microbiology. 74:4997– 5007.

Yuan D., G. Liu, D. Ren, L. Zhao, C. Kan, Y. Yang, W. Ma, Y. Li and L. Zhang. 2012. A survey on occurrence of thermophilic bacilli in commercial milk powders in China. Food Control. 25:752–757.

Zhao Y., M. P. M. Caspers, T. Abee, R. J. Siezen and R. Kort. 2012. Complete genome sequence of *Geobacillus thermoglucosidans* TNO-09.020, a thermophilic sporeformer associated with a dairy-processing environment. Journal of Bacteriology. 194: 4118-4118.

2.8 Supplemental figures and tables

2.8.1 Supplemental figures

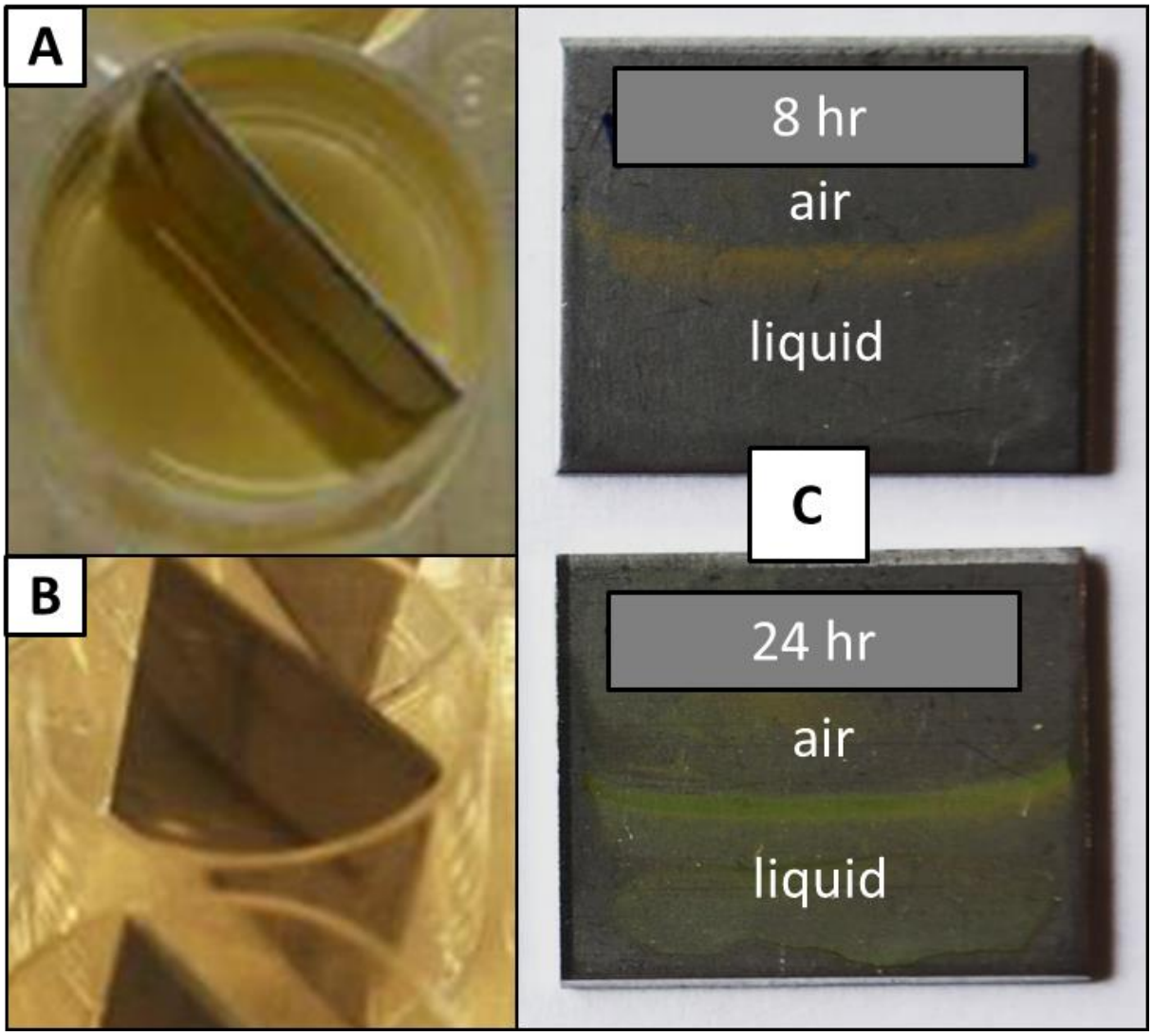
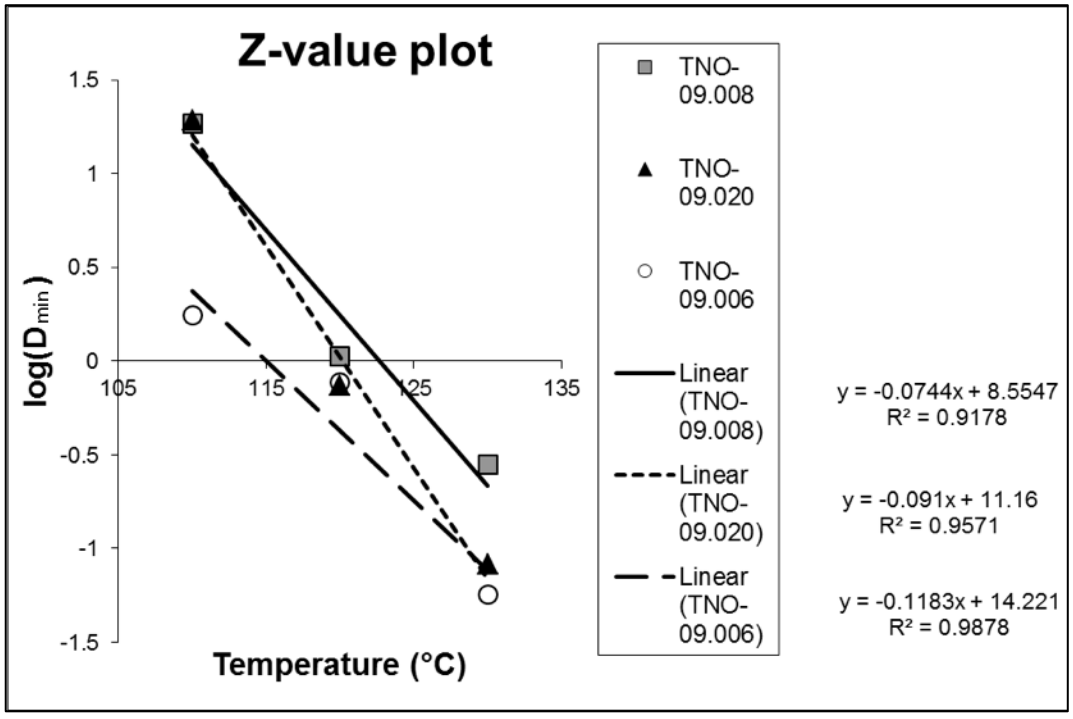


Figure S2.1 Standing steel biofilm model system. *Geobacillus thermoglucosidans* TNO-09.020 was cultured 0-24h at 65°C. **(A)** Culture well with tryptone medium and stainless steel (SS) coupon. **(B)** Same well after 24h culturing: horizontal biofilm visible on washed SS coupon. **(C)** Auramine-stained coupons harvested, washed and stained after 8 or 24 hours of culturing. The submerged and air-exposed parts of the coupons have been indicated (liquid air)

A)

Inactiv. Temp.	Strain	*Kmax* (min^{-1})	*Nres*	*D* (min)	*R square*
110 °C	**TNO-09.006**	1.25	3.11	2	0.9643
	TNO-09.008	0.13	1	18	0.9872
	TNO-09.020	0.12	1	19	0.8921
120 °C	**TNO-09.006**	3.01	2.92	0.8	0.9924
	TNO-09.008	2.19	4.03	1.1	0.9546
	TNO-09.020	3.09	5.01	0.8	0.9786
130 °C	**TNO-09.006**	40.34	2.08	0.06	0.9788
	TNO-09.008	8.24	1.49	0.28	0.9892
	TNO-09.020	27.88	2.22	0.08	0.9775

B)



C)

Strain	*slope*	*Z*	*R square*
TNO-09.006	-0.0744	13	0.9178
TNO-09.008	-0.091	11	0.9571
TNO-09.020	-0.1183	8	0.9878

Figure S2.2 Heat inactivation kinetics of spores from thermophilic strains. The heat inactivation kinetics of spores isolated in this study indicated by (A) D-values and other inactivation equation parameters, (B) linear logD/temperature equations, and (C) their slopes and Z-values. The D-values of the spore batches were derived from plots with log CFU versus incubation time by fitting the following log-linear model with tail to the data by using Excel add-in GInaFiT:

$logN_t = (logN_0 - logN_{res}) * exp(-k_{max} * t) + logN_{res}$

In which log N_t is the spore count at a certain time point, $logN_0$ is the spore count at t=0, $logN_{res}$ is the number of spores in the tail and kmax is the inactivation constant. The D-value was derived by $D=ln(10)/k_{max}$. In case of no tailing ($logN_{res} = 0$), the model was reduced to a log-linear inactivation model. The z-values were calculated by plotting the logD value against the temperature and performing a linear regression. The z-value was calculated by z=-1/slope.

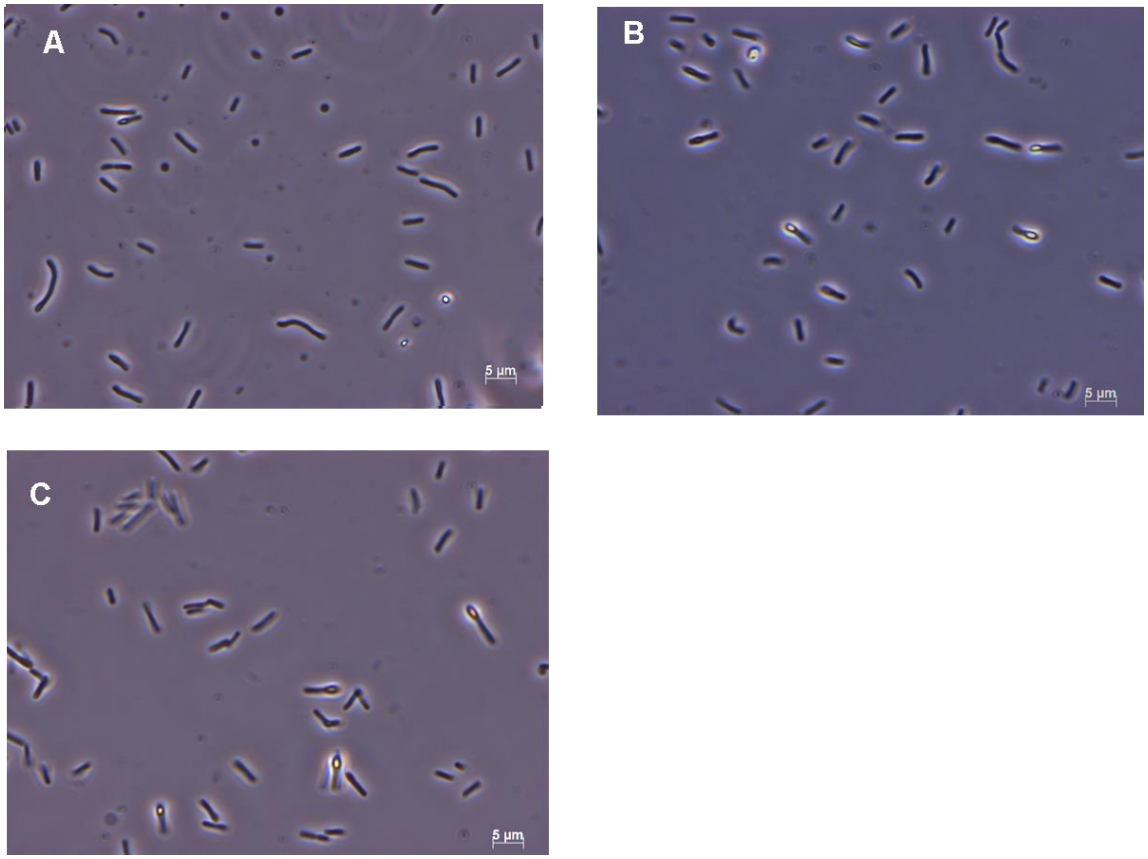
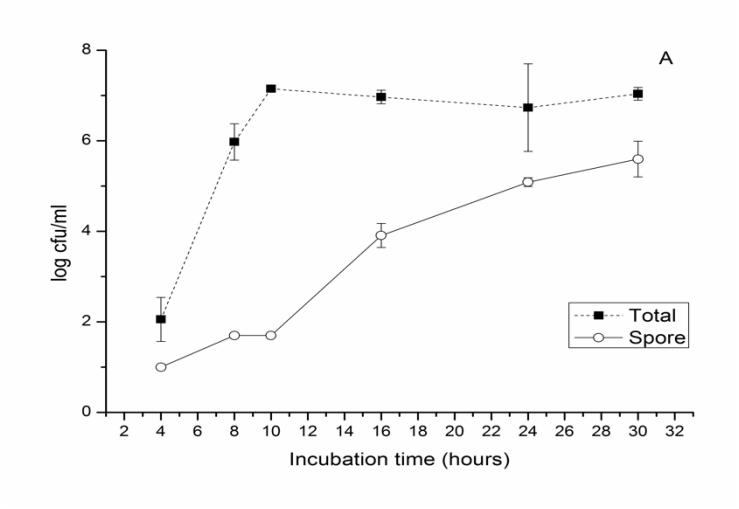
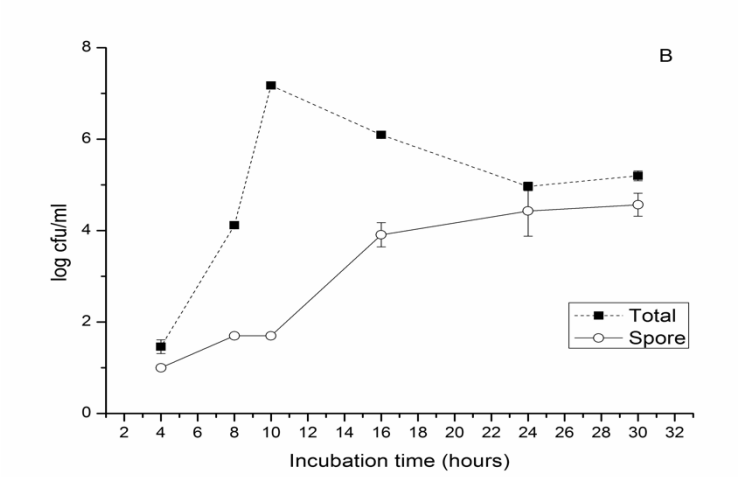


Figure S2.3 Phase-contrast micrographs of thermophilic sporeformers. Planktonic cells and spores of the 3 thermophilic spore forming isolates. **A**) *Anoxybacillus flavith*ermus TNO-09.006; **B)** *Geobacillus stearothermophilus* TNO-09.008; **C)** *Geobacillus thermoglucosidans* TNO-09.020.



A)



B)

Figure S2.4 Colony forming units of *Geobacillus thermoglucosidans* TNO-09.020 in the standing steal biofilm system. Graphical representation of bacterial counts development (CFU/ml) in the static standing steal biofilm system with 1% Tryptone at 60 °C. The 2 panels represent **(A)** the planktonic cells from the medium and **(B)** biofilm cells from the steel coupon. The total CFUs were determined, vegetative cells and spores (■), and for the same samples treated for 30 min at 100 °C, allowing spore counts (○). Means and standard deviations from 3 replicate static culture systems are shown.

2.8.2 Supplemental tables

Table S2.1 Colony forming units of dairy factory samples used in this study.

Sample description	*Id*	*Temperature in factory*	*Concen-tration or volume*	*Log (CFU/ml)*		
				Total 30°C	Total 55°C	Spore 55°C
M1 (Standard milk)	M1	4°C	undiluted	7.0	3.0	2.0
M2 (Standard milk)	M2	4°C	undiluted	4.5	3.3	1.1
Pipeline 1	i01	75°C	0.07 g/ml	8.7	5.0	4.0
Pipeline 2	i02	75°C	1.5 ml	8.7	4.0	4.0
Pipeline 3	i03	75°C	0.33 g/ml	6.0	5.0	4.0
Pipeline 4	i05	75°C	1.5 ml	2.0	4.0	3.0
Evaporator 1	i06	65°C	1.0 g/ml	8.0	4.0	2.0
Evaporator 2	i07	65°C	1.0 g/ml	6.1	3.4	3.0
Pasteur	i08	68°C	0.90 g/ml	6.0	3.0	3.0
Tank	i09	45°C	1.0 g/ml	8.0	4.0	3.0
Tank filter	i10	37° - 45°C	1.0 g/ml	7.0	3.0	2.0
Drying tower	i11	60°C	0.15 g/ml	7.0	3.0	2.0
Dairy concentrates	i14	20° - 60°C	0.10 g/ml	2.1	2.2	<1.8
Dairy concentrates	i15	20° - 60°C	0.10 g/ml	<2.1	1.8	1.8

Table S2.2 Colony forming units of enrichments at 55ºC and 65ºC. Standard milk (M1 and M2) or sterile milk inoculated with industrial fouling from a whey evaporator (i07) was cultured O/N at 55 or 65°C in plastic culture wells containing standing steel coupons (static biofilm model). The table shows the CFUs in the different fractions in the biofilm system: fraction in the medium (2 ml), on the stainless-steel surface, and attached on the plastic well.

Medium	Temperature	Fraction	Log (CFU) / Fraction
M2 (standard milk)	55°C	medium	5.7
		steel	5.4
		well	5.5
	65°C	medium	3.3
		steel	1.6
		well	2.1
M3 (Standard milk)	55°C	medium	5.4
		steel	5.5
		well	5.6
	65°C	medium	6.3
		steel	4.8
		well	5.8
Evaporator (i7)	55°C	medium	7.0
		steel	5.5
		well	5.4
	65°C	medium	5.7
		steel	4.2
		well	4.7

Chapter 3 Growth of dairy isolates of *Geobacillus thermoglucosidans* in skim milk depends on lactose degradation products supplied by *Anoxybacillus flavithermus* as secondary species

Y. Zhao,
M. Kumar,
M. P. M. Caspers,
M. N. Nierop Groot,
J.M.B.M van der Vossen,
T. Abee

This chapter was published in the Journal of Dairy Science (2018) 101:1013-1019

3.1 Summary

Thermophilic bacilli such as *Anoxybacillus* and *Geobacillus* are important contaminants in dairy powder products. Remarkably, one of the common contaminants, *Geobacillus thermoglucosidans,* displayed poor growth in skim milk, whereas significant growth was observed in the presence of an *Anoxybacillus flavithermus* dairy isolate. In the present study, the underlying reason for this growth dependence of *G. thermoglucosidans* is investigated. Whole genome sequences of 4 *A. flavithermus* strains and 4 *G. thermoglucosidans* strains were acquired with special attention given to carbohydrate utilization clusters and proteolytic enzymes. Zooming in on traits relevant for dairy environments, comparative genomic analysis revealed that all *G. thermoglucosidans* strains (i) lack genes necessary for lactose transport and metabolism, (ii) show poor growth in skim milk and (iii) produce white colonies on X-gal plates indicating the lack of β-galactosidase activity. Tested *A. flavithermus* isolates scored positive in these tests, in line with the presence of a putative lactose utilization gene cluster. All tested isolates from both species showed proteolytic activity on MPCA plates. Subsequent experiments in liquid skim milk with added glucose or galactose supported growth of *G. thermoglucosidans* isolates, in line with the presence of respective monosaccharide utilization gene clusters in the genomes. HPLC analysis of *A. flavithermus* TNO-09.006 culture filtrate indicated that the previously described growth dependence of *G. thermoglucosidans* in skim milk, is based on the supply of glucose and galactose by *A. flavithermus* TNO-09.006.

3.2 Introduction

Heat resistant spores from thermophilic bacilli form a major concern in dairy powder processing facilities posing a risk for product contamination. This group of bacteria is able to grow in sections of skim milk manufacturing plants, such as heat exchangers and evaporation sections where elevated temperatures, typically between 40 °C to 65 °C, are applied (Seale et al., 2015). Although these bacilli are not pathogenic, their presence as spores in end products may lead to quality issues of respective reconstituted products when conditions are favourable for germination and outgrowth (Setlow and Johnson, 2013; Watterson et al., 2014; Wells-Bennik et al., 2016). Obviously, such quality issues could also lead to severe economic losses.

Spore-forming bacteria survive through their innate ability to resist adverse conditions in dairy manufacturing processes including heat, mechanical disruption, and a wide variety of chemicals (Burgess et al., 2010). Moreover, both spores and vegetative cells can attach to stainless steel and fouled surfaces in dairy processing lines. Once attached to the surface, the spores may germinate, grow out, and form biofilms. Spores from thermophilic bacilli detected in the end products conceivably originate from biofilms formed in dairy processing lines (Seale et al., 2015). Thermophilic bacteria generally have a high growth rate with generation times typically in the range of 15-20 minutes under optimal conditions. Thus, high cell counts can be reached in a short period of time. Additionally, the resistance of the endospores to heat and chemicals makes it difficult to fully eliminate the thermophilic bacteria from dairy processing environments (Burgess et al., 2010; Wells-Bennik et al., 2016). Understanding of the proliferation and survival of spore forming bacteria within the dairy processing environments and dairy products is therefore a prerequisite to develop more effective methods to control and reduce contamination.

Thermophilic bacteria predominantly isolated from the dairy processing industry are of the species *Anoxybacillus flavithermus* and the genus *Geobacillus* spp., formerly classified as *Bacillus* (Burgess et al., 2010; Zhao et al., 2013; Sadiq et al., 2016). *A. flavithermus* typically dominates the preheating section of the skim milk powder process, whereas a mix of *A. flavithermus* and *Geobacillus* spp. was found in the evaporation and drying stages of the process. *Geobacillus* spp. was predominantly isolated from fouling sites (Zhao et al., 2013).

To proliferate in skim milk, bacteria must utilize the carbon and nitrogen sources in skim milk. Raw milk is a nutrient-rich source supporting bacterial growth. It contains approximately 3.4% protein, 3.7% fat, 4.6% lactose, and 0.7% ash by weight (Jensen, 1995). However, the *G. thermoglucosidans* strains TNO-09.020 and TNO-09.023, previously isolated from a fouling site in the dairy-processing pipeline, showed poor growth on skim milk plates and in liquid skim milk. This growth deficiency, and its biofilm forming capacity in skim milk, was restored when *A. flavithermus*, a proteolytic thermophilic sporefomer from the microbiota of the same dairy-concentrate processing plant, was added as secondary species (Zhao et al., 2013). To understand the mechanism of growth dependence of *G. thermoglucosidans* on the presence of *A. flavithermus* as secondary species in skim milk, we performed a comparative genomic analysis on the isolates of both species. Next to the 5 dairy isolates (*Anoxybacillus flavithermus* TNO-09.006, TNO-09.014, TNO-09.016; *Geobacillus thermoglucosidans* TNO-09.020, TNO-09.023), 1 *A. flavithermus* and 2 *G. thermoglucosidans* hot spring isolates (*Anoxybacillus flavithermus* WK1, *Geobacillus thermoglucosidans* C56-YS93, *Geobacillus thermoglucosidans* Y4-1MC1) available in public genome databases were included in the study (Table 3.2). Particularly, the presence and absence of genes involved in metabolic pathways including nitrogen (proteolytic system) and carbohydrate metabolism, required for growth in skim milk, were taken into consideration. Genes only present in 1 of the 2 species were enumerated and their annotated function was inspected for possible roles in growth in skim milk. Moreover, growth experiments in skim milk without and with added supplements that conceivably could support the outgrowth of *G. thermoglucosidans* were conducted. Finally, the cell free culture-filtrate of *A. flavithermus* was analyzed to identify compounds produced by this organism, which could support the growth of *G. thermoglucosidans* in skim milk.

3.3 Results and discussion

Comparative genomics revealed that genomes of *A. flavithermus* strains (except for *A. flavithermus* WK1 isolated from a hot spring) encompass genes encoding a lactose ABC transporter cassette and β-galactosidase activity that are involved in the uptake and the utilization of lactose (De Vos and Vaughan, 1994), whereas these genes are lacking in the *G. thermoglucosidans* genomes. Because *G. thermoglucosidans* lacks these genes it is conceivably unable to utilize lactose present in skim milk for its

activity and growth. Notably, genome analysis revealed all strains to contain an extensive repertoire of proteolytic enzymes including genes encoding putative proteases and peptidases (Y. Zhao, unpublished data). This would suggest that acquisition of nitrogen sources by *G. thermoglucosidans* is not a limiting factor for growth in skim milk, as previously suggested (Zhao et al., 2013). Based on this genome analysis it was hypothesized that *G. thermoglucosidans* strains cannot utilize lactose present in skim milk, and cannot grow readily in skim milk where lactose is the sole energy source for the growth initiation; while *A. flavithermus* strains, apart from the hot spring strain WK1, are able to do so because they are equipped with the genes involved in the uptake and the utilization of lactose.

We subsequently validated this hypothesis by combining comparative genome analysis with results obtained from growth experiments on X-gal plates, on skim milk plates, in liquid skim milk, and on 1% MPCA plates, respectively. All colonies of *G. thermoglucosidans* strains on X-gal plates appeared white indicating the lack of β-galactosidase activity, while the *A. flavithermus* strains, except WK1, produced blue colonies indicating hydrolysis of X-gal resulting from β-galactosidase activity (Table 3.1). This phenotypic result matches with the prediction from the comparative genome analysis. Additionally, we determined the growth capacity of the *G. thermoglucosidans* and *A. flavithermus* strains on skim milk plates and in liquid skim milk. Results showed that all *A. flavithermus* strains, except *A. flavithermus* WK1, could form visible colonies on skim milk plates, whereas tested *G. thermoglucosidans* strains (like *A. flavithermus* WK1) did not show visible growth on skim milk plates. Growth assays in liquid milk with the *A. flavithermus* and *G. thermoglucosidans* strains showed the same trend compared to growth on skim milk plates (Table 3.2). All *A. flavithermus* strains, except *A. flavithermus* WK1, were able to grow and reached cell densities of 7.5 – 7.7 log CFU/mL within 6 hours in skim milk, whereas the *G. thermoglucosidans* strains and *A. flavithermus* WK1 remained around the inoculation cell density in 6 hours. In conclusion, the presence of genes encoding β-galactosidase activity and a lactose ABC transporter cassette in these thermophilic bacilli correlates with their capacity to grow rapidly in skim milk, which points to an essential and differentiating role of the identified putative lactose utilization cluster in these strains. Notably, all tested isolates from both species showed proteolytic activity on MPCA plates (Table 3.1). This indicates that the inability of *G. thermoglucosidans* to grow in skim milk is not due to the lack of proteolytic digestion of milk proteins. Thus, we hypothesize that the inability to digest lactose

and subsequent access to the monosaccharides (glucose and galactose) as the lactose hydrolysing products, is a limiting factor for rapid growth of *G. thermoglucosidans* strains in milk.

Table 3.1 Growth of *Anoxybacillus flavithermus* and *Geobacillus* thermoglucosidans strains on skim milk plates, in liquid skim milk, on X-gal plates and 1% MPCA plates

Strains	*Growth on skim milk agar plate*[1]	*Initial cell counts (log CFU/mL)*	*Cell counts at 6h (log CFU/mL)*	*On MPCA (halo)*[2]	*On X-gal plate*[3]
Anoxybacillus flavithermus TNO-09.006	+	3.6 (±0.1)	7.5(±0.4)	+	Blue
Anoxybacillus flavithermus TNO-09.014	+	4.1(±0.4)	7.7(±0.3)	+	Blue
Anoxybacillus flavithermus TNO-09.016	+	3.7(±0.5)	7.7(±0.2)	+	Blue
Anoxybacillus flavithermus WK1	-	3.7(±0.2)	4.0(±1.0)	+	White
Geobacillus thermoglucosidans TNO-09.020	-	3.8(±0.2)	4.4(±0.5)	++	White
Geobacillus thermoglucosidans TNO-09.023	-	4.8(±0.5)	5.3(±0.8)	++	White
Geobacillus thermoglucosidans C56_YS93	-	4.1(±0.2)	2.8(±1.0)	+	White
Geobacillus thermoglucosidans Y4_1MC1	-	4.4(±0.5)	5.8(±1.6)	++	White

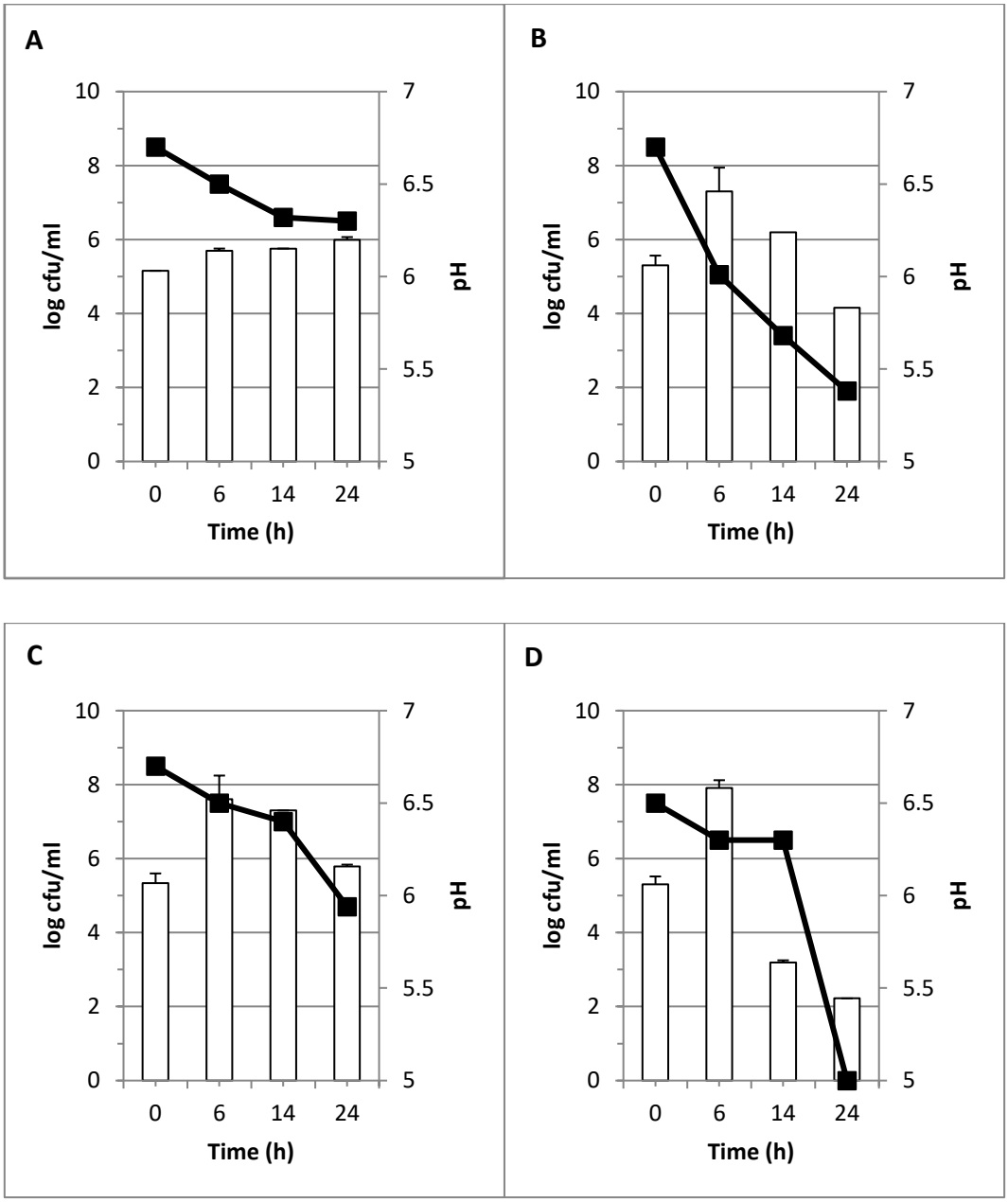
[1] – means no visible colony; + means visible colonies detected on skim milk agar plates;

[2] – means no halo; + means weak halo; ++ means large halo around the colony on MPCA plates indicative for proteolytic activity;

[3] blue colony on X-gal plates indicates positive for β-galactosidase activity, white colony indicates negative for β-galactosidase activity. X-gal stands for 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

To test the hypothesis that indeed only the lactose uptake and hydrolysing capacity were hampered, we conducted growth assays with *G. thermoglucosidans* in skim milk without and with supplementation of glucose or galactose, as well as in an assay with lactose-free milk, i.e., milk in which the lactose was hydrolysed by lactase resulting in the availability of glucose and galactose (Figure 3.1). *G. thermoglucosidans* exhibited limited growth in skim milk, having approximately 0.5 and 0.8 log unit increase in

viable counts in six and 24 hours, respectively (Figure 3.1A). The presence of glucose, galactose and hydrolysed lactose (glucose and galactose) enhanced growth of *G. thermoglucosidans* reaching approximately 2 log units increase within 6 hours (Figure 3.1B-D). The fact that glucose and galactose supported growth in skim milk of *G. thermoglucosidans* suggests that in a co-culture, hydrolysis of lactose by *A. flavithermus* may supply glucose, galactose or both to *G. thermoglucosidans*, supporting its growth in skim milk. We therefore analysed the sugars and organic acids present in cell free culture-filtrates of *A. flavithermus* TNO-09.006 and *G. thermoglucosidans* TNO-09.020 grown in skim milk by using HPLC, and this confirmed that lactose is degraded by *A. flavithermus* and that residual levels of glucose and galactose were present. Additionally, β-galactosidase activity was detected in the cell free culture-filtrate harvested after 24h growth of *A. flavithermus* in skim milk at 55° C (Y. Zhao, unpublished data). Subsequent studies in skim milk with added *A. flavithermus* TNO-09.006 culture filtrates, indeed showed stimulation of growth of *G. thermoglucosidans* TNO-09.020 (Figure 3.2). Above mentioned experiments support that the observed growth stimulation of *G. thermoglucosidans* can be attributed to supply of lactose degradation products via β-galactosidase activity of *A. flavithermus* TNO-09.006.


A
B
C
D
log cfu/ml
pH
Time (h)
0
6
14
24
10
8
6
4
2
0
7
6.5
6
5.5
5

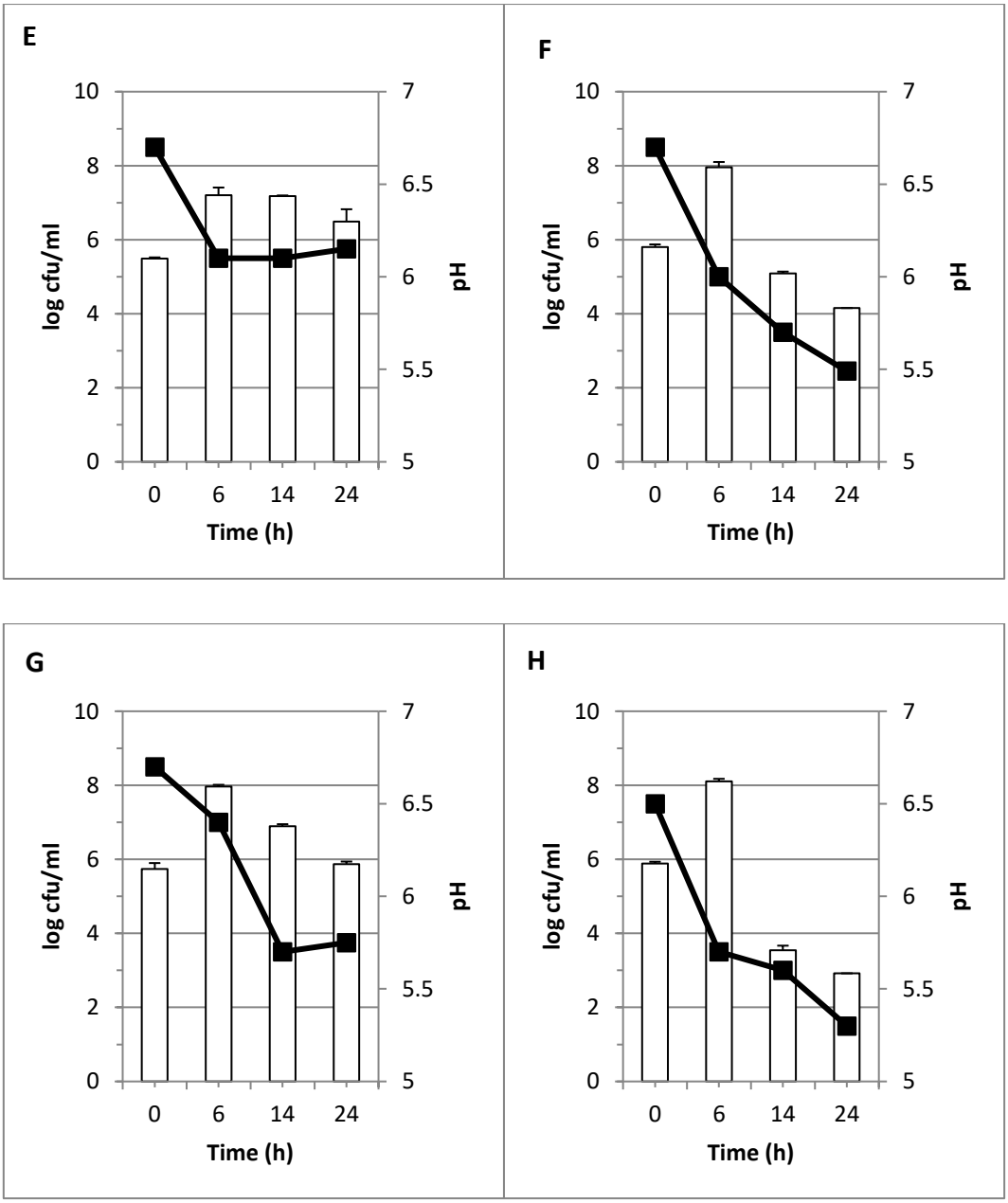


Figure 3.1 Growth of *Geobacillus thermoglucosidans* TNO-09.020 (A-D) and *Anoxybacillus flavithermus* TNO-09.006 (E-H) in skim milk without and with supplements, and in lactose-free milk. Growth in skim milk (A, E), skim milk + 0.25% glucose (B, F), skim milk + 0.25% galactose (C, G), and lactose-free (lactase treated) 2.5% glucose and 2.5% galactose containing milk (D, H). Total counts (log CFU/mL) are shown in bar graphs and pH measurements in filled squares. Error bars represent standard deviations of biological duplicates.

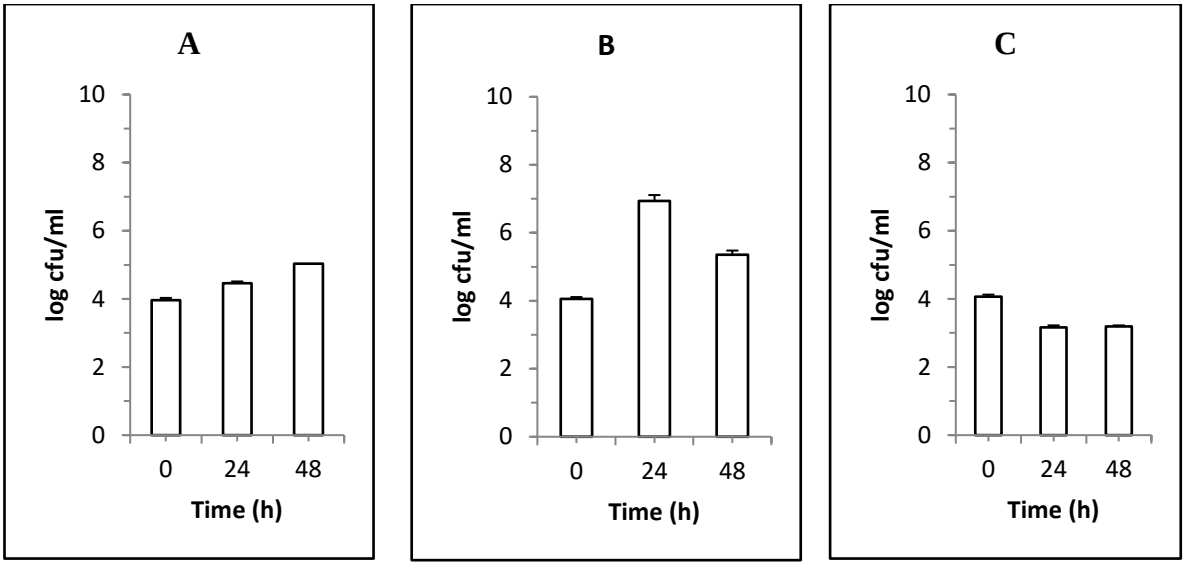


Figure 3.2 Effect of *Anoxybacillus flavithermus* TNO-09.006 cell free culture-filtrate (filtrate) on the growth of *Geobacillus thermoglucosidans* TNO-09.020 in milk. A, skim milk; B, skim milk with added filtrate; and C, filtrate only. Graph represents values of CFU measurements at the indicated sample times. Error bars represent standard deviations of biological duplicates.

Notably, when *G. thermoglucosidans* was grown in skim milk with addition of glucose, the viable counts after first growing, subsequently declined by almost 3 log units after 24h (Figure 3.1B). A significant drop of the pH value to pH 5.4 in this *G. thermoglucosidans* culture was also observed (Figure 3.1B). This acidification would explain the decrease in viability of the *G. thermoglucosidans* cells, and previous studies have reported the limited capacity of thermophilic bacilli to cope with low pH conditions where relatively high minimum pH values (around pH 6) for growth were documented (Tola and Ramaswamy, 2014). Notably, *G. thermoglucosidans* also showed efficient growth in the lactose-free (glucose and galactose containing) milk, likewise followed by a decline in cell numbers in 24h correlating with a significant decline in pH (Figure 3.1D). On the contrary, in skim milk with supplementation of galactose, the growth of *G. thermoglucosidans* was comparable to glucose added and lactose-free milk but there is no significant drop of cell density observed at 24h. In line with that, the drop in pH was also less significant (Figure 3.1C). Except for growth in skim milk (Figure 3.1E), similar growth and acidification behaviour was observed for *A. flavithermus* (Figure 3.1F-H).

Besides monitoring colony-forming unit (CFU) and pH change during growth, spore counts at different time points were also monitored. Notably, we could not visually

detect spores using qualitative microscopy in samples from the *G. thermoglucosidans* and *A. flavithermus* cultures in skim milk, and in skim milk supplemented with glucose and lactose-free milk (hydrolysed lactose). Nevertheless, we could observe spores of *G. thermoglucosidans* and *A. flavithermus* after the growth in the galactose-supplemented skim milk (final pH 6). The above observation combined with the extensive drop in pH of the growth medium observed only in skim milk supplemented with glucose or in glucose containing lactose-free skim milk, suggests that rapid acidification of the growth medium provides inhibitory conditions for sporulation of *G. thermoglucosidans* and *A. flavithermus.* The negative influence of low pH conditions on sporulation has also been described by Yazdany and Lashkari (1975). In their study, *G. stearothermophilus* ATCC 7953 and *G. stearothermophilus* NCIB 8919 produced few spores when cultured in a medium with final pH around 5.5, whereas when grown in a medium adjusted to pH 7.7 to 8.7, sporulation of the *G. stearothermophilus* occurred (Yazdany and Lashkari, 1975). It is conceivable that glucose metabolism-induced acidification leads to a reduction in the total CFU counts, hence preventing the formation of spores. Factors that contribute to the relatively high minimum pH for growth of a range of thermophilic spore formers, and loss of sporulation capacity at weak acidic conditions, remain to be elucidated.

In conclusion, this study indicated that the previously described growth stimulation of *G. thermoglucosidans* during co-culture in skim milk with *A. flavithermus* TNO-09.006 (Zhao et al., 2013) is not based on proteolytic activity of the latter, but rather is due to the supply of glucose and galactose following lactose degradation by β-galactosidase of *A. flavithermus* TNO-09.006. In addition, it is observed that glucose metabolism–induced acidification in skim milk led to a reduction in total count in the late growth stage of *A. flavithermus* and *G. thermoglucosidans* and, in turn, hampered sporulation. These observations shed new insights into the interre-lationship within microbiological communities of dairy sporeformers and could have implications for industrial hygiene operations. Future studies will be directed at studying whether these observations in *A. flavithermus* and *G. thermoglucosidans* hold true for other known dairy or nondairy thermophilic spore-forming species such as *G. stearothermophilus*. The latter suggestion is based on a recent paper by Burgess et al. (2017) that provided novel insights in diversity of *G. stearothermophilus* strains and presented genotypic and phenotypic evidences for lactose utilization capacity in dairy strains of this species. Greater understanding of

those thermophilic sporeformers is needed to develop better control measures in milk processing environments.

3.4 Experimental procedures

The genomes of the 8 thermophilic bacilli of the species *G. thermoglucosidans* and *A. flavithermus* (Table 3.2) have been compared to each other and to the reference non-thermophilic strain *Bacillus subtilis* 168. For the newly sequenced genomes from this study (Zhao et al., 2012; Caspers et al., 2013; Caspers et al., 2016), after assembly of sequence reads into contigs, RAST (Aziz et al., 2008) was used to predict ORFs and to do automatic annotation. Selected genes and protein sequences were manually curated using BLASTP and InterPro. Orthologous groups (OGs; i.e. gene families) in the 8 genomes were determined using OrthoMCL (Enright et al., 2002). This program uses all-against-all protein BLAST where it groups proteins with more homology within the species than homology with proteins outside the species. In this way, orthologs (genes in different species that evolved from a common ancestral gene by speciation) are separated from paralogs (genes related by duplication within a genome). When an OG contained more than 1 gene per strain, i.e. highly similar genes, this OG was manually split into separate OGs containing only 1 gene per strain. An exception was made for transposons and other mobile elements.

Table 3.2 Presence/absence of genes for lactose metabolism in selected *Anoxybacillus flavithermus* and *Geobacillus thermoglucosidans* strains

Strains	*NCBI Accession number and Reference*	*Origin*	*Genes*[1]		
			lactose ABC transporter, permease protein	*lactose ABC transporter, lactose-binding protein*	*Beta-galactosidase (EC 3.2.1.23)*
Anoxybacillus flavithermus TNO-09.006	AMCM00000000	Standard milk used for dairy processing plant, Netherlands	1	1	1
Anoxybacillus flavithermus TNO-09.014	LUFB00000000	Standard milk used for dairy processing plant, Netherlands	1	1	1

Anoxybacillus flavithermus TNO-09.016	LUCQ00000000	Standard milk used for dairy processing plant, Netherlands	1	1	1
Anoxybacillus flavithermus WK1	GCA_000019045.1	Hot spring, New Zealand	0	0	0
Geobacillus thermoglucosidans TNO-09.020	NZ_CM001483	Casein pipe fouling in the diary processing plant, Netherlands	0	0	0
Geobacillus thermoglucosidans TNO-09.023	LUCT00000000	Casein pipe fouling in the diary processing plant, Netherlands	0	0	0
Geobacillus thermoglucosidans C56_YS93	NC_015660.1	Obsidian Hot Spring, YNP, USA	0	0	0
Geobacillus thermoglucosidans Y4_1MC1	NC_014650.1	Bath Hot Spring, YNP,USA	0	0	0

[1] differences regarding the presence (1) or absence (0) of genes including lactose ABC transporters and β-galactosidase in *A. flavithermus* and *G. thermoglucosidans* are presented.

The strains used in this study were obtained from -80 °C stocks. Tryptone Soy Agar (TSA, plates) and Tryptone Soy Broth (TSB) (Tritium Microbiologie, The Netherlands) were used as basic culture media. Inoculum used in this study were overnight cultures with optical density (OD) standardized to one. OD was measured at the wavelength at 600nm using a spectrophotometer (Ultrospec 2100, Amersham Bioscience, UK). Inoculum density used in the growth experiments was 1% (vol/vol). All experiments throughout this study were performed as multiple independent duplicates or triplicates.

Skim milk agar plates for comparative growth tests contain 50% v/v UHT skim milk 0% fat (FrieslandCampina, The Netherlands), 1.5% agarose and distilled water. Upon mixing the ingredients, media was subsequently autoclaved at 20 min 121 °C. The lactose free milk plates were made as described above, instead of UHT skim milk, Lactose free milk (Konings Zuivel B.V., The Netherlands) was used. Milk plate count agar (MPCA) plates were made, containing tryptone (0.13g/l), yeast extract (2.7g/l), glucose (1.1g/l), skim milk powder (1.1 g/l) and agar (33.3g/l). X-gal plates with a final concentration of 40 µg/mL X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) were made as described previously (Zhao et al., 2013). Overnight

cultures (OD = 1) of the 8 selected strains were streaked separately to acquire single colonies on the prepared plates; subsequently, the plates were incubated overnight at 55°C for further observation. Growth experiments in liquid skim milk without and with added supplements were performed using the following procedures. D-glucose (Tritium Microbiologie, The Netherlands) or D-galactose (Tritium Microbiologie, The Netherlands) were added to freshly obtained UHT skim milk in concentration of 0.25g/L for the growth assays; and Lactose free milk with concentrations of 2.5g/L glucose and galactose was directly used for this experiment. These milk media were subsequently inoculated with *A. flavithermus* or *G. thermoglucosidans* and incubated for maximal 24h at 65°C, 100 rpm and selected time samples were subsequently diluted and spread plated on TSA plates for enumeration. The pH of the cultures was also measured using a pH meter (Mettler Toledo, Switzerland).

Finally, to study the effect of cell free culture supernatants of *A. flavithermus* on growth of *G. thermoglucosidans* in skim milk, cell free culture supernatants of 24h grown cultures of *A. flavithermus* in skim milk medium were prepared. The pH of the cultures (pH 5.9) was set to pH 4.7 using 4*N* HCl to precipitate milk caseins and subsequently filtered using a 0.45µm filter (Millipore, Merck, USA). The filtrate was centrifuged (Allegra bench top centrifuge, Beckman Counter, USA) at 4000 g for 10 min. The clear supernatant was collected, and the pH of the supernatant was readjusted to the initial pH 5.9 using 4*N* NaOH and stored at -20 °C for later use. The filtrates of the 24h *A. flavithermus* cultures in skim milk were added to the skim milk with the ratio 2:1 to assess the effect on growth of *G. thermoglucosidans*. These filtrates were also subjected to high-performance liquid chromatography (HPLC) assays in order to quantify sugars and organic acids present in cell free culture-filtrates of *A. flavithermus* TNO-09.006. HPLC was performed using an Aminex HPX-87H column (Biorad, USA) mounted in a Waters Alliance e2695 HPLC-apparatus. The column was eluted with 0.6 mL/min 5 m*M* H_2SO_4 at 60°C. Peaks were detected by a refractive-index detector and a dual-wavelength absorbance detector at 210nm and 270nm. Raw data were processed into Peak Areas and concentrations were calculated using Waters Empower™ 2 software (Waters®, USA).

3.5 Acknowledgments

We wish to thank Lars Wilms for assisting in conducting filtrate experiments, Karin Overkamp for assisting HPLC experiments, and Roland Siezen for bio-IT support. The project is funded by TI Food and Nutrition, a public-private partnership on precompetitive research in food and nutrition. The public partners are responsible for the study design, data collection and analysis, decision to publish, and preparation of the manuscript. The private partners have contributed to the project through regular discussion.

3.6 References

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke and O. Zagnitko. 2008. The RAST server: Rapid annotations using subsystems technology. BMC Genomics. 9:75.

Burgess, S. A., D. Lindsweay and S. H. Flint. 2010. Thermophilic bacilli and their importance in dairy processing. International Journal of Food Microbiology. 144:215-225.

Burgess, S. A., S. H. Flint, D. Lindsay, M. P. Cox and P. J. Biggs. 2017. Insights into the *Geobacillus stearothermophilus* species based on phylogenomic principles. BMC Microbiology. 17:140

Caspers, M. P. M., J. Boekhorst, T. Abee, R. J. Siezen and R. Kort. 2013. Complete genome sequence of *Anoxybacillus flavithermus* TNO-09.006, a thermophilic sporeformer associated with a dairy-processing environment. Genome Announcements. 1:e0010-13.

Caspers, M. P. M., J. Boekhorst, A. de Jong, R. Kort, M. N. Nierop Groot and T. Abee. 2016. Draft genome sequences of four thermophilic spore formers isolated from a dairy-processing environment. Genome Announcements. 4:e00757-16.

De Vos, W. M. and E. E. Vaughan. 1994. Genetics of lactose utilization in lactic acid bacteria. FEMS Microbiology Reviews. 15:217-237.

Enright, A. J., S. Van Dongen and C. A. Ouzounis. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Research. 30:1575-1584.

Jensen, R. G. 1995. Handbook of Milk Composition. Page 3 in Introduction. Academic Press, London, UK.

Sadiq, F. A., Y. Li, T. Liu, S. Flint, G. Zhang, L. Yuan, Z. Pei and G. He. 2016. The heat resistance and spoilage potential of aerobic mesophilic and thermophilic spore forming bacteria isolated from Chinese milk powders. International Journal of Food Microbiology. 238:193-201.

Seale, B., S. Burgess, S. Flint, J. Brooks, P. Bremer and S. Parkar. 2015. Thermophilic spore-forming bacilli in the dairy industry. Pages 112-137 in Biofilms in the Dairy Industry. John Wiley & Sons, Ltd, Chichester, UK.

Setlow, P. and E. A. Johnson. 2013. Spores and their significance. Pages 45-79 in Food Microbiology. Spores and their significance. ASM Press, Washington, DC.

Tola, Y. B. and H. S. Ramaswamy. 2014. Combined effects of high pressure, moderate heat and pH on the inactivation kinetics of *Bacillus licheniformis* spores in carrot juice. Food Research International. 62:50-58.

Watterson, M. J., D. J. Kent, K. J. Boor, M. Wiedmann and N. H. Martin. 2014. Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest. Journal of Dairy Science. 97:2487-2497.

Wells-Bennik, M. H. J., R. T. Eijlander, H. M. W. den Besten, E. M. Berendsen, A. K. Warda, A. O. Krawczyk, M. N. Nierop Groot, Y. Xiao, M. H. Zwietering, O. P. Kuipers and T. Abee. 2016. Bacterial spores in food: Survival, emergence, and outgrowth. Annual Review of Food Science and Technology. 7:457-482.

Yazdany, S. and K. B. Lashkari. 1975. Effect of pH on sporulation of *Bacillus stearothermophilus*. Applied Microbiology. 30:1-3.

Zhao, Y., M. P. M. Caspers, T. Abee, R. J. Siezen and R. Kort. 2012. Complete genome sequence of *Geobacillus thermoglucosidans* TNO-09.020, a thermophilic sporeformer associated with a dairy-processing environment. Journal of Bacteriology. 194:4118.

Zhao, Y., M. P. M. Caspers, K. I. Metselaar, P. de Boer, G. Roeselers, R. Moezelaar, M. N. Nierop Groot, R. C. Montijn, T. Abee and R. Kort. 2013. Abiotic and microbiotic factors controlling biofilm formation by thermophilic sporeformers. Applied and Environmental Microbiology. 79:5652-5660.

Chapter 4 Biofilm dynamics of *Geobacillus thermoglucosidans*, a dairy processing isolate

Y. Zhao

4.1 Summary

Contamination by thermophilic sporeformers in end-products is a primary concern for dairy concentrate-processing plants. Biofilms are known to be one of the major sources of persistent contaminations in these plants, but little is known about the biofilm-forming mechanism of thermophilic sporeformers in such environments. The contaminant *Geobacillus thermoglucosidans* has been isolated at different locations inside these plants. Compared to other common thermophilic contaminants – such as *Geobacillus stearothermophilus* and *Anoxybacillus flavithermus, G. thermoglucosidans* shows the best biofilm-forming capacity with high biofilm mass produced in a reproducible manner under lab conditions. In this study, *G. thermoglucosidans* was selected as a model organism to investigate biofilm-formation mechanisms.

In the present study, a comparison was made of the gene expression profiles of *G. thermoglucosidans* TNO-09.020 in the planktonic phase and the biofilm phase during biofilm development. It was found that cells in these two phases showed discernible behaviour in their decline stage and differences in expression profiles over the course of biofilm development. The gene categories, with functions such as methionine biosynthesis and phenylalanine degradation, were significantly upregulated in biofilm-phase cells compared to planktonic-phase cells. The potential roles of these newly identified pathways in biofilm formation were also investigated and provided insight in the biofilm development of this organism.

This study described the expression profile of *G. thermoglucosidans* TNO-09.020 during biofilm formation and discovered some novel factors which might influence biofilm development of this strain.

4.2 Introduction

Besides mesophilic spore-formers, thermophilic sporeformers are especially problematic in food-producing industrial facilities operating processes with temperatures in the range of 40 to 65 °C, since these temperatures support the growth and biofilm formation of these organisms (Burgess et al., 2010). The growth of these thermophiles in biofilms can result in bacteria and spores being released into the end-products, including whey and milk concentrates in numbers of up to 10^6 CFU/g (Scott et al., 2007; Watterson et al., 2014). These spores could germinate under favourable conditions, ultimately resulting in high numbers of bacteria and off-flavour in end-products (Scheldeman et al., 2005; Scott et al., 2007). In order to prevent the presence and outgrowth of the accumulated spores, costly precautionary measures are taken in dairy concentrate-processing environments, such as frequent cleaning, short production runs, and intensive microbial product testing.

Most of the thermophilic sporeformers that have been identified to-date in dairy concentrate-processing lines and products belong to the genera *Bacillus, Geobacillus*, and *Anoxybacillus* (Flint et al., 1997; Scott et al., 2007; Yuan et al., 2012). *Geobacillus* spp. and *A. flavithermus* are the most frequently reported species in thermophilic dairy biofilms (Burgess et al., 2010). The presence of spores of these thermophilic bacilli in the end-products most likely results from the detachment of spores from biofilms from stainless steel surfaces in dairy concentrate-processing plants (Scott et al., 2007). Moreover, it is also reported that biofilms render bacteria tolerance to antimicrobial factors, making those cells residing in biofilm even more difficult to be eradicated (Sadiq et al., 2017). This is why special interest is focused on understanding the development of biofilm by thermophilic bacilli.

To begin with, three isolates from a dairy concentrate-processing plant were selected according to their ability to grow at high temperatures and their biofilm-forming capacity, namely: *Anoxybacillus flavithermus* TNO-09.006, *Geobacillus stearothermophilus* TNO-09.008, and *Geobacillus thermoglucosidans* TNO-09.020. Of these three model strains, *G. thermoglucosidans* TNO-09.020 (or GT20), which was isolated in a bio-fouling sample from a dairy caseinate concentrate-processing plant, showed the strongest and most robust biofilm-forming capacity at high temperatures (60°C-70°C) in TSB (tryptone soy broth) under lab conditions (Zhao et

al., 2013). GT20 was therefore chosen in the current study aiming to understand its biofilm-forming mechanism. To this end, DNA microarray technology was applied to identify global gene expression profiles in biofilm-phase cells during their biofilm development and compare them to planktonic-phase cells in a static biofilm system. Our objective was to use the gene expression profiles to discover unique patterns of gene expression associated with biofilm development for thermophilic sporeformers; these organisms are rarely studied but are highly relevant to plants with high-temperature operating processes.

4.3 Results

4.3.1 Growth of *G. thermoglucosidans* TNO-09.020 in a static biofilm system

The biofilm development cycle of GT20 was monitored by counting both biofilm-phase and planktonic-phase cells by Colony Forming Units (CFU) (Figure 4.1) in a static biofilm system over a time span of 30h. The total viable cell counts reached 7.2 log CFU/cm^2 in the biofilm phase and 7.0 log CFU/ml in the planktonic phase both in about 10h. Subsequently, a stationary phase, with a constant total viable cell count, was observed for the planktonic-phase cells, while a significant drop in the total viable cell count was observed for the biofilm-phase cells. At 8h and 10h, spore counts remained similar to 4h, and from 16h onward the number reached a final count around 5 log CFU/ml, both in the biofilm and the planktonic phase. At time points 24h and 30h, total cell counts equalled spore counts in the biofilm phase. Two spore-selection treatments were used in this study (80°C 10 min for all spores, 100°C 30 min for heat-resistant spores). At 30h, the results of these two treatments did match, and before 30h, not all spores were very heat resistant, especially in the planktonic phase. Under the microscope, a monolayer of biofilm cells was observed from 8h onwards, and spores were observed in both the biofilm and the planktonic phase from 16h onwards (data not shown). Taken together, based on the information described above from CFU monitoring as well as microscopic observation, cells from

the biofilm and the planktonic phase have discernible development characteristics, especially in the late development phase.

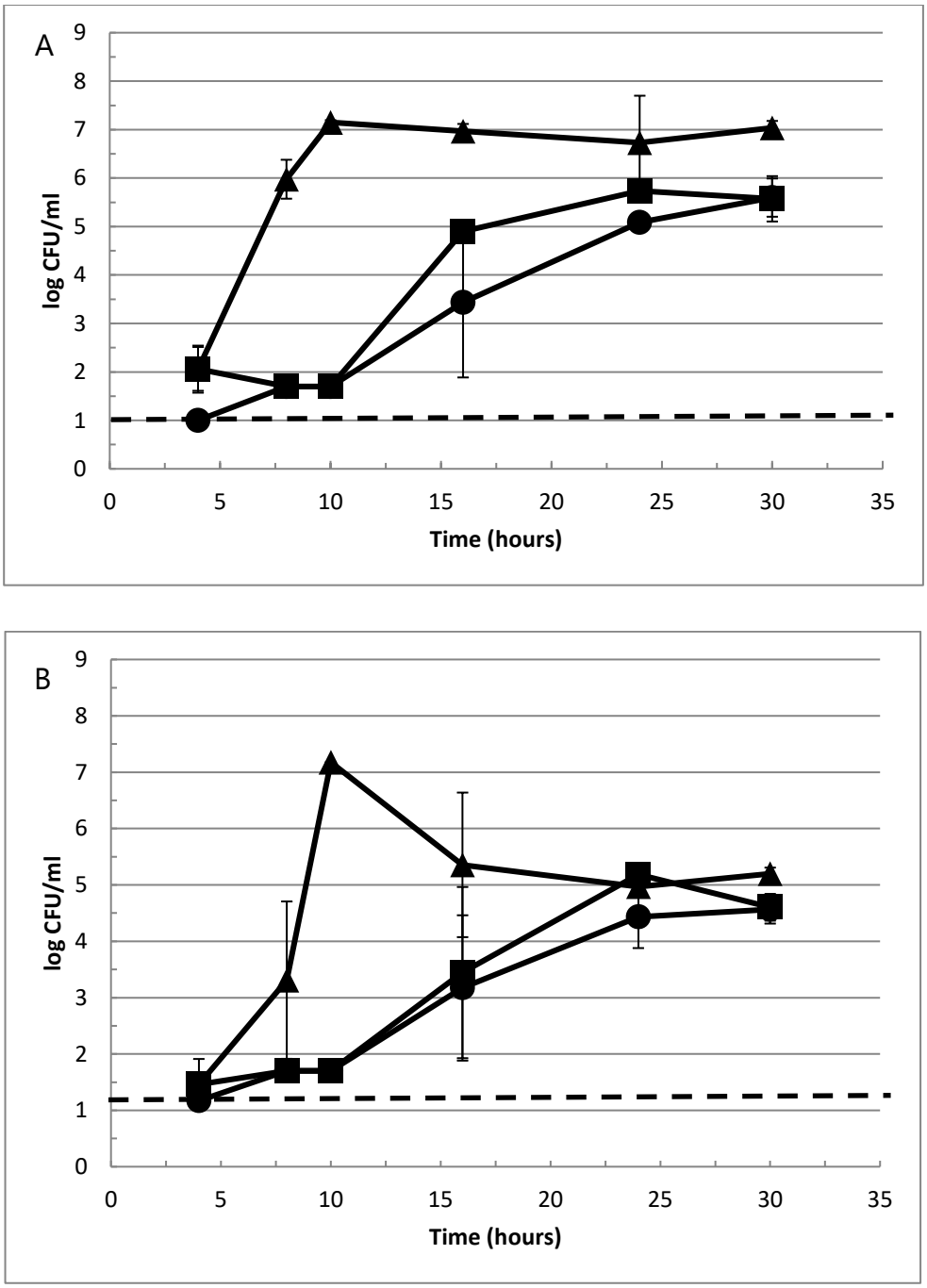


Figure 4.1 Cell counts of *G. thermoglucosidans* TNO-09.020 in 1.05% tryptone medium at 65 °C, in both the planktonic phase (A) and the biofilm phase (B). Each point with error bar represents the mean and standard deviation from three independent trials. Triangles = total cells, Squares = spores, Circles = heat-resistant spores. Biofilm phase cells (surface-attached cells) from each sample were dissolved in 6ml tryptone medium,

later the cell number per ml (CFU/ml) was enumerated[2]. Any point below the detection limit (dotted line) indicates that no viable colonies were detected on TSA plate.

4.3.2 Evaluation of early stage differential expression in *G. thermoglucosidans* TNO-09.020 (GT20)

In order to examine the mechanisms associated with biofilm formation during overall biofilm development, a whole genome transcriptomic study was conducted. First of all, to study the genes responsible for biofilm formation, without having to deal with noise from genes involved in sporulation, the differentially expressed genes in the early biofilm-development phase, namely at the 8h time point, were investigated. A list of gene clusters was selected in which the majority of gene candidates were over-expressed in the biofilm phase compared to planktonic phase (Table 4.1). The comparison of biofilm-phase and planktonic-phase cells showed that upregulation in biofilm-phase cells was mainly observed in categories such as methionine biosynthesis, fatty acid utilization, and peptide utilization. Downregulation in biofilm-phase cells was scattered in some functional groups in which at most five identified gene members in GT20 were present (data not shown). The highest downregulation percentage is in the category biosynthesis of glycogen, in which 4 out of 5 members of the functional group were more than two times downregulated. Some of these changes were initiated early and maintained throughout the process; others were restricted to the earliest stages of the biofilm formation.

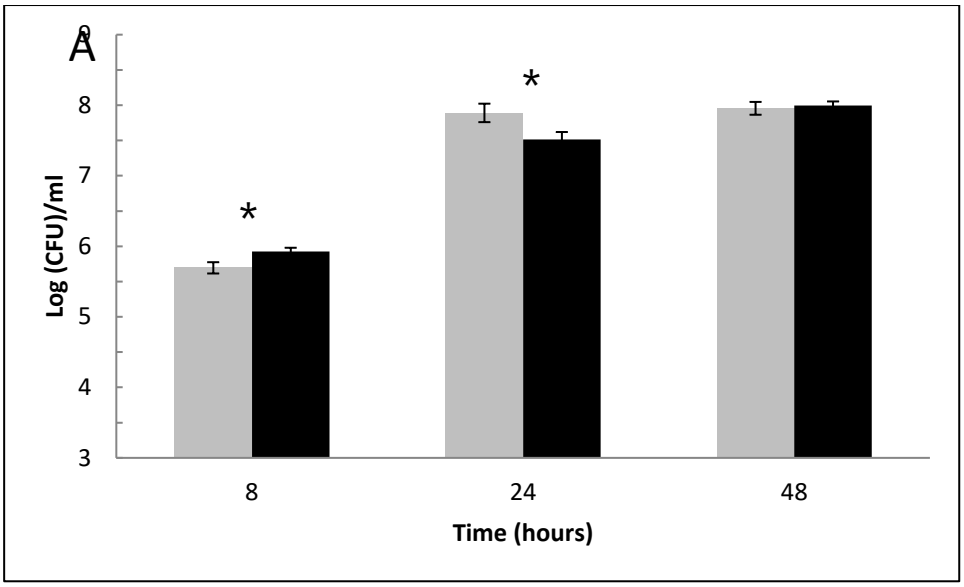
[2] Biofilm of 3.92 cm^2 (1.4 * 1.4 cm^2 *2 sides) was dissolved in 6 ml medium and then CFU/ml was determined. The conversion factor between CFU/ml and CFU/cm^2 is 1.5.

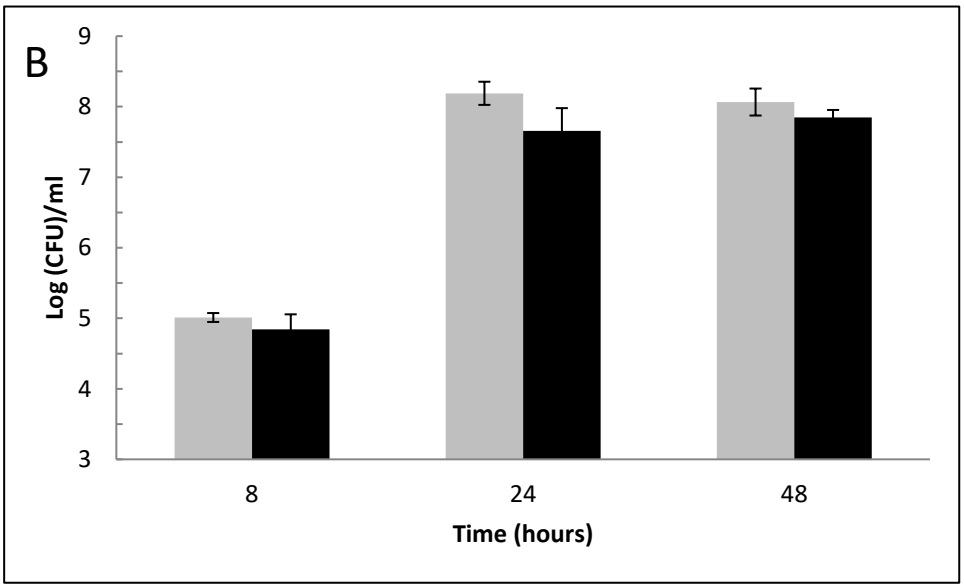
Table 4.1 Categories (Subtiwiki gene function categories and regulon categories which contain more than 5 genes) of genes in which more than 50% of the genes are uniquely upregulated at 8h in biofilm phase. The first column (genes/category) shows the total numbers in the specific category in the genome; the second column (upregulated genes) shows the number of genes that are more than doubly upregulated in the biofilm phase in comparison to the planktonic phase in the specific category; the third column (upregulated genes%) presents the percentages of upregulated genes in the gene category. No downregulated gene is presented in the table because all downregulated gene categories have no more than 5 genes.

Subtiwiki Categories	*Number of Genes*	*Number of Upregulated Genes*	*Upregulated Genes (%)*
Biosynthesis/ acquisition of methionine/ S-adenosylmethionine	19	14	74
Utilization of glutamine/ glutamate	9	6	67
Utilization of peptides	16	9	56
Utilization of fatty acids	16	9	56
FadR regulon	14	9	64
SdpR regulon	17	12	71
TnrA regulon	19	12	63

In the functional category biosynthesis/acquisition of methionine/S-adenosylmethionine (http://subtiwiki.uni-goettingen.de/wiki/), 74% of the genes are more than doubly upregulated in biofilm-phase cells compared to planktonic-phase cells at 8h (Table 4.1). Therefore, it can be hypothesized that the amino acids produced by the methionine synthesis pathway – namely, methionine, aspartate, and homocysteine – are needed for biofilm mass formation. If the hypothesis is true, then increased availability of, for example, methionine inside the cells should lead to more biofilm formation. This was tested by addition of the amino acids to the medium. Effect of methionine is presented in Figure 4.2 (results for aspartate and homocysteine not shown). In these tests, the addition of these amino acids in tryptone medium, showed a significant influence ($P<0.05$) on biofilm formation at 48 hour when compared to tryptone medium without the addition: total cell count 48h ($P=0.03$) (Figure 4.2C); spore count 48h ($P=0.005$)(Figure 4.2D); this was monitored by the standard plate count method. This observation suggests that the increased

availability of amino acid methionine has an influence on biofilm mass formation, which coincides with transcriptomic data.





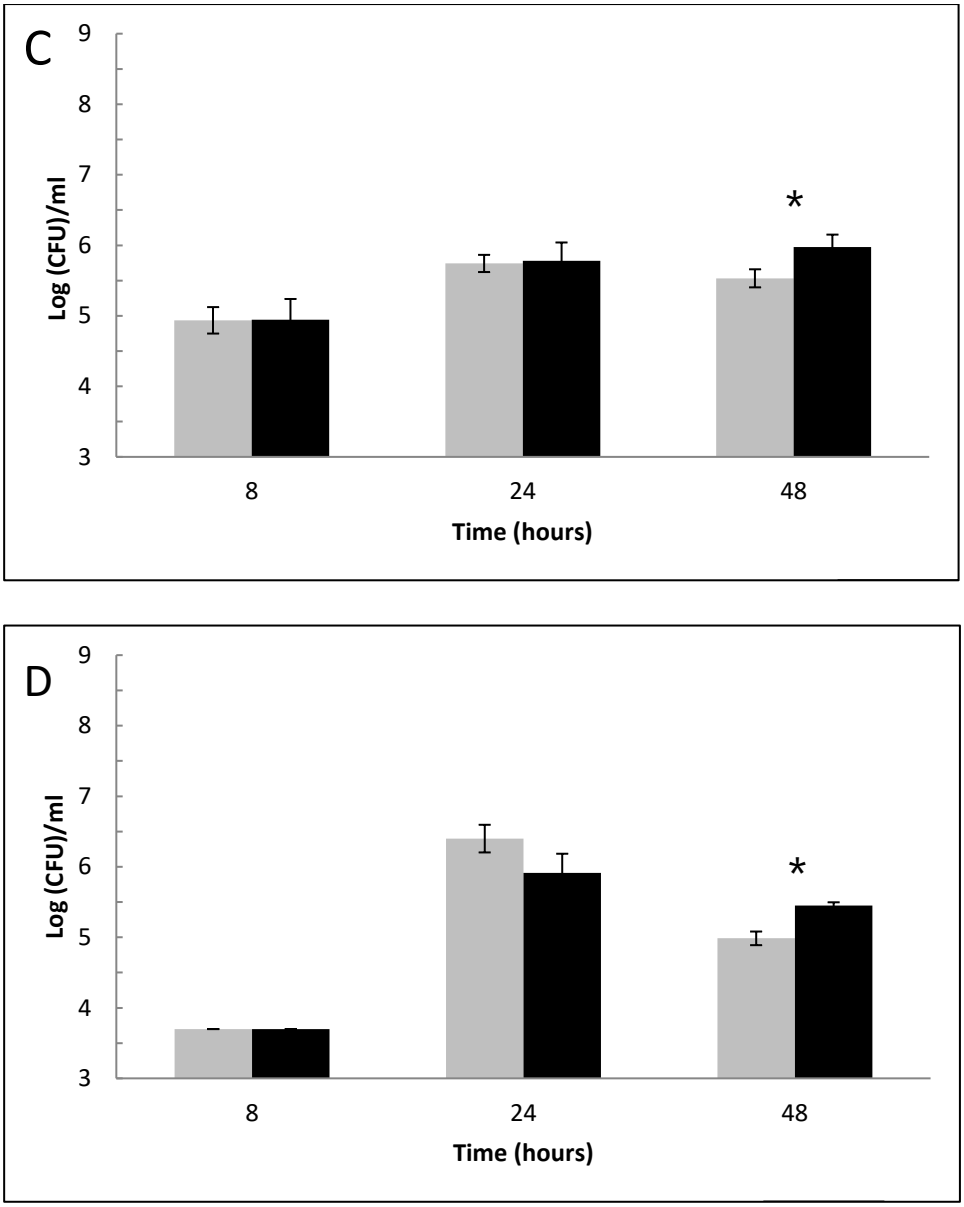


Figure 4.2 Effect of methionine on the growth of *G. thermoglucosidans* TNO-09.020 in the planktonic phase and the biofilm phase at 65°C. Data points represent CFU/ml in tryptone medium (grey) or in tryptone medium with addition of 200μM methionine (black). Each bar with error bar represents the mean and standard deviation, from three independent trials. An asterisk (*) indicates that there is a statistical difference between the two samples verified by t-test ($P<0.05$).

A. Total cell counts in planktonic phase.
B. Spore counts in planktonic phase
C. Total cell counts in biofilm phase.
D. Spore counts in biofilm phase.

Further, it is hypothesed that methionine biosynthesis accelerated the activities of the LuxS catalysed side pathway, which plays a role in signalling biofilm cell attachment by GT20 via AI-2 (AI-2 is a quorum sensing molecule). To test this hypothesis, the presence of AI-2 in *G. thermoglucosidans* culture was first checked through an AI-2 assay. If AI-2 is present in the culture medium, the *V. harveyi* reporter strain BB170 will generate luminescence which can be captured and measured by a spectrophotometer. This study observed that the luminescent signals generated from the AI-2 reporter strain in cell-free extracts from GT20 were in proximity range of the negative control, and even lower (Figure 4.3). Those signals indicated no (or low) secretion of AI-2 by GT20; therefore, it is unlikely that methionine biosynthesis influenced the biofilm formation via AI-2 generated from the LuxS catalysed side pathway. The reasons behind over-expression of methionine biosynthesis-related genes in biofilm-phase cells at 8h is not linked to AI-2 production, but it is linked to biofilm mass formation.

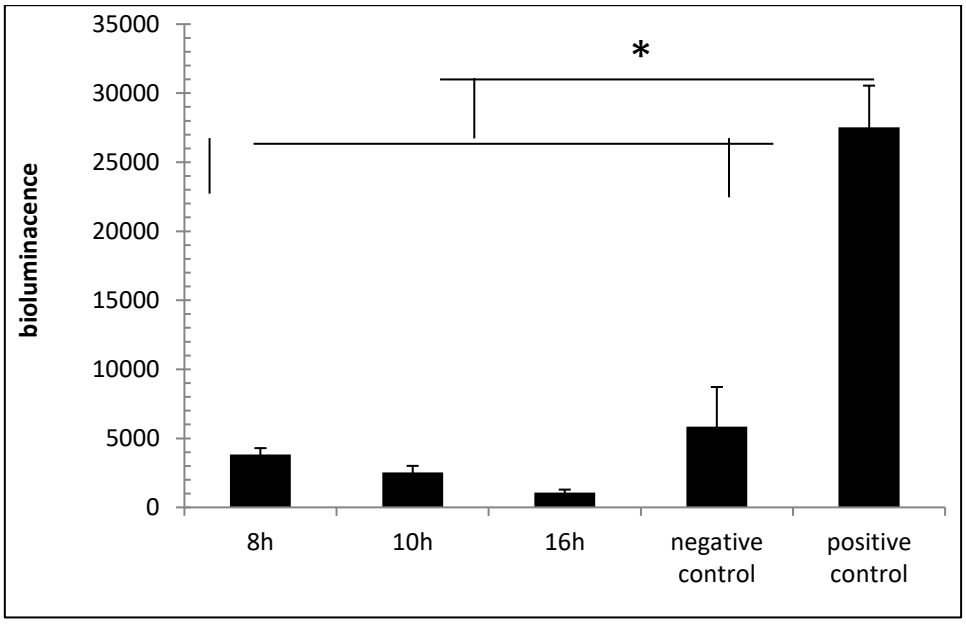


Figure 4.3 AI-2 assay response of *G. thermoglucosidans* TNO-09.020 cell-free culture fluids. The responses of *V. harveyi* reporter strains BB170 (sensor 12, sensor 21) to signalling substances present in cell-free culture fluids of *G. thermoglucosidans* at 8h, 10h, and 16h culture in tryptone medium and sterile tryptone medium (negative control), *V. harveyi* BB120 (positive control) are shown. An asterisk (*) indicates that there is a statistical difference between the two samples verified by t-test ($P<0.05$).

4.3.3 Clustering of expression profiles of biofilm-phase and planktonic-phase cells

Using the Neighbourhood Co-Regularized Spectral Clustering Algorithm, the expression profiles of GT20 cells of all samples taken from both the planktonic and the biofilm phase at time points 4h, 8h, 10h, 16h, 24h, and 30h were found to be clustered into four distinguishable groups (Figure 4.4). These groups were: group 1 (8h and 10h samples), group 2 (16h planktonic-phase samples), group 3 (4h, 24h, 30h planktonic-phase samples), and group 4 (16h, 24h, 30h biofilm-phase samples). Biofilm-phase samples at 4h were not processed in the transcriptomic study due to insufficient cell numbers. The clustering result of the samples shows early (8h and 10h) and late (16h, 24h, and 30h) biofilm development have discernible expression profiles. Moreover, the separation of samples in group 3 and group 4 suggested that there were important genes or gene clusters that differentiate planktonic cells and biofilm cells of GT20.

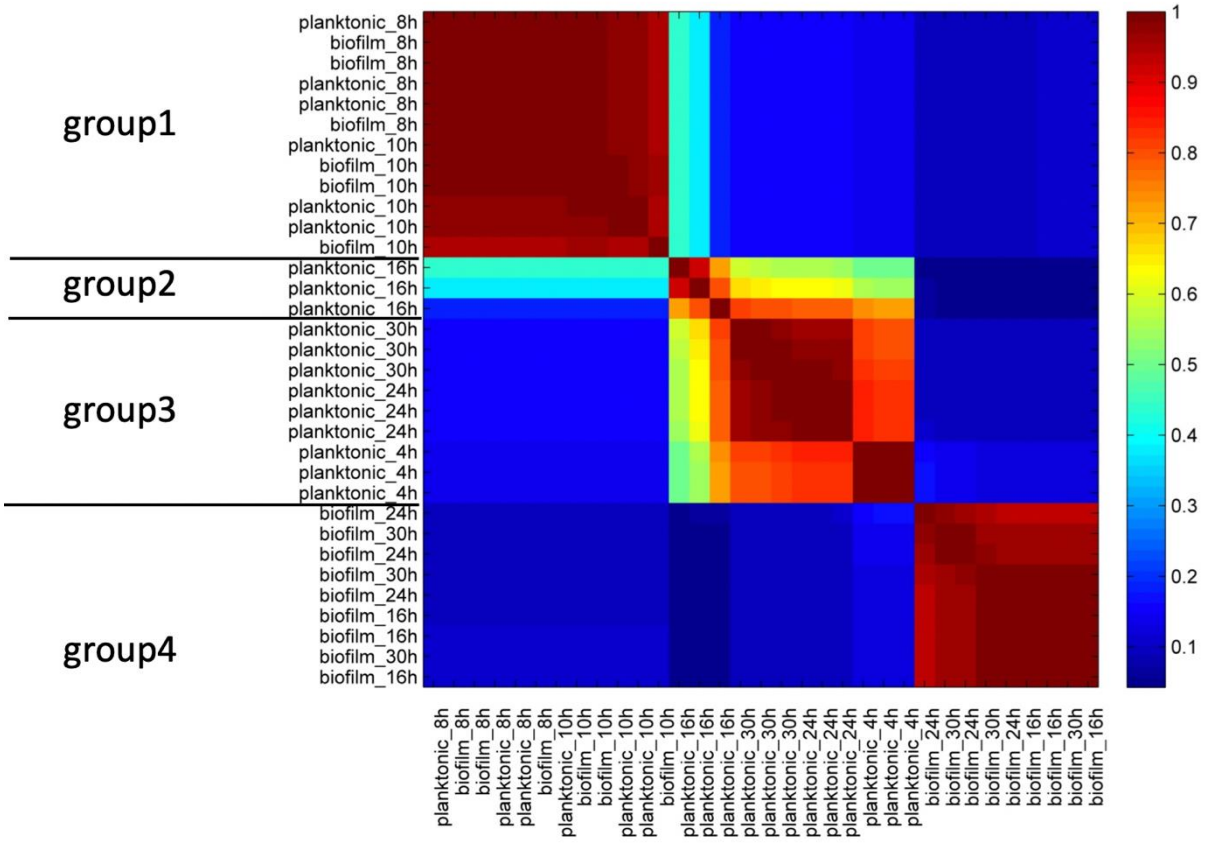


Figure 4.4 Consensus spectral clustering of expression profiles of biofilm- and planktonic-phase cells. Samples are symmetrically ordered along both the x- and the y-axis, according to the co-occurrence matrix (i.e., the more similar the gene expression profiles of any two samples, the higher their tendency to cluster together). Co-occurrence values range from 0.0 illustrated with colour blue (for samples that never cluster together) to 1.0 illustrated with red (for samples that always cluster together).

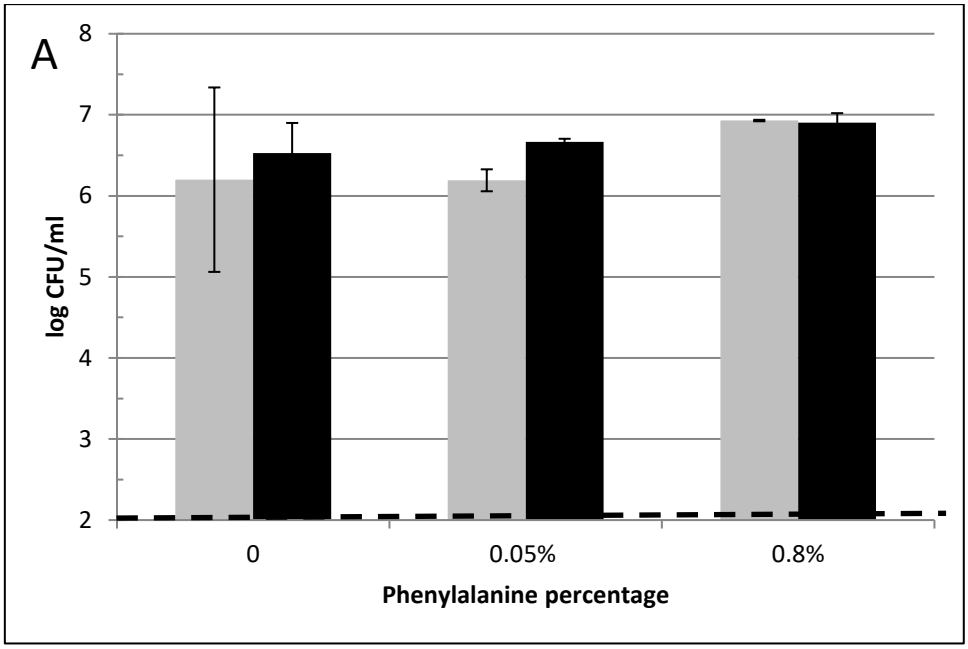
To investigate which dominant genes contributed to this clustering, Multi-View Unsupervised Feature Selection (MVUFS) was used. Among the 30 selected genes which contributed to the clustering, 8 genes with phenylalanine degradation function stood out. More than double upregulation of these 8 genes in biofilm-phase cells in comparison with planktonic-phase cells was observed (Table 4.2). It is hypothesized that the late phase of biofilm development of GT20 is influenced by the availability of phenylalanine.

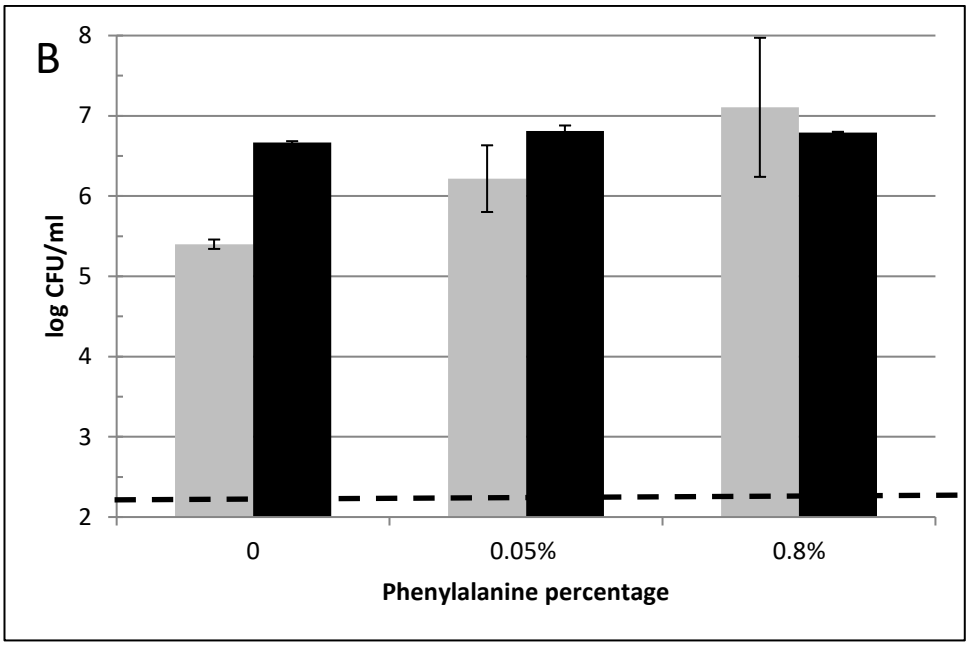
Table 4.2 Relative expression of 8 of the 30 most important genes contributing to clusters, in both planktonic and biofilm phase at different time points, based on Consensus Spectral Clustering. This is the output from the unsupervised feature selection method Multi-View Unsupervised Feature Selection (MVUFS). Maxexpr = maximum expression. The ratio numbers in the table, ascribed to each time point and each gene, are the actual expression divided by maximum expression (Maxexpr).

				Planktonic phase					*Biofilm phase*				
	Annotation	*Gene*	*MaxExpr*	*8h*	*10h*	*16h*	*24h*	*30h*	*8h*	*10h*	*16h*	*24h*	*30h*
GT20_3943	phenylacetic acid degradation protein paaI	paaL	741	0.4	0.4	0.4	0.6	0.4	0.6	0.4	0.8	1.0	0.7
GT20_2466	phenylacetate-CoA ligase	paaF	590	0.3	0.2	0.2	0.2	0.2	0.2	0.2	1.0	1.0	0.7
GT20_3362	phenylacetic acid degradation protein paaA	paaA	2011	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.9	1.0	0.8
GT20_1998	phenylacetic acid degradation B family protein	paaB	2364	0.0	0.0	0.1	0.3	0.3	0.1	0.1	1.0	1.0	0.8
GT20_2592	phenylacetic acid catabolic family protein	paaC	2139	0.1	0.1	0.2	0.4	0.4	0.1	0.1	1.0	1.0	0.8
GT20_3950	phenylacetate-CoA oxygenase, PaaJ subunit	paaJ	1854	0.1	0.1	0.2	0.4	0.5	0.1	0.1	1.0	1.0	0.8
GT20_1100	phenylacetic acid degradation protein PaaX	paaX	12202	0.3	0.5	0.5	0.4	0.4	0.5	0.8	1.0	0.8	0.8
GT20_3612	phenylalanine dehydrogenase	pah	427	0.2	0.2	0.2	0.4	0.5	0.2	0.2	0.7	0.9	1.0

To test this hypothesis, the growth of *G. thermoglucosidans* in tryptone medium was monitored, with and without the addition of phenylalanine. The results showed no detectable significant difference in growth, when measured in a crystal violet assay (data not shown). In addition, the influence of the added phenylalanine on the increase of vegetative or spore cell counts in the static biofilm system was monitored (Figure 4.5). No significant influence of phenylalanine on total cell numbers was observed. However, both in the planktonic and the biofilm phase, at 16h specifically, an increased number of spore counts in planktonic phase (P = 0.36) and spore counts in biofilm phase (P = 0.31) were observed with the addition of 0.8% phenylalanine (these differences were not significant). This coincides with the time point at which the differences in expression of the genes between planktonic and the biofilm phase were first observed (Table 4.2). There was an increase of around on average 1 to 1.5 log unit of spores in CFU/ml in tryptone medium with the addition of 0.8% of phenylalanine compared to the medium without this addition. However, the increase is not significant statistically (P>0.05), therefore, there is an observed

difference, but the relation between significantly higher expression of phenylalanine degradation-related genes in late developmental biofilm-phase cells and spore formation is not concluded.





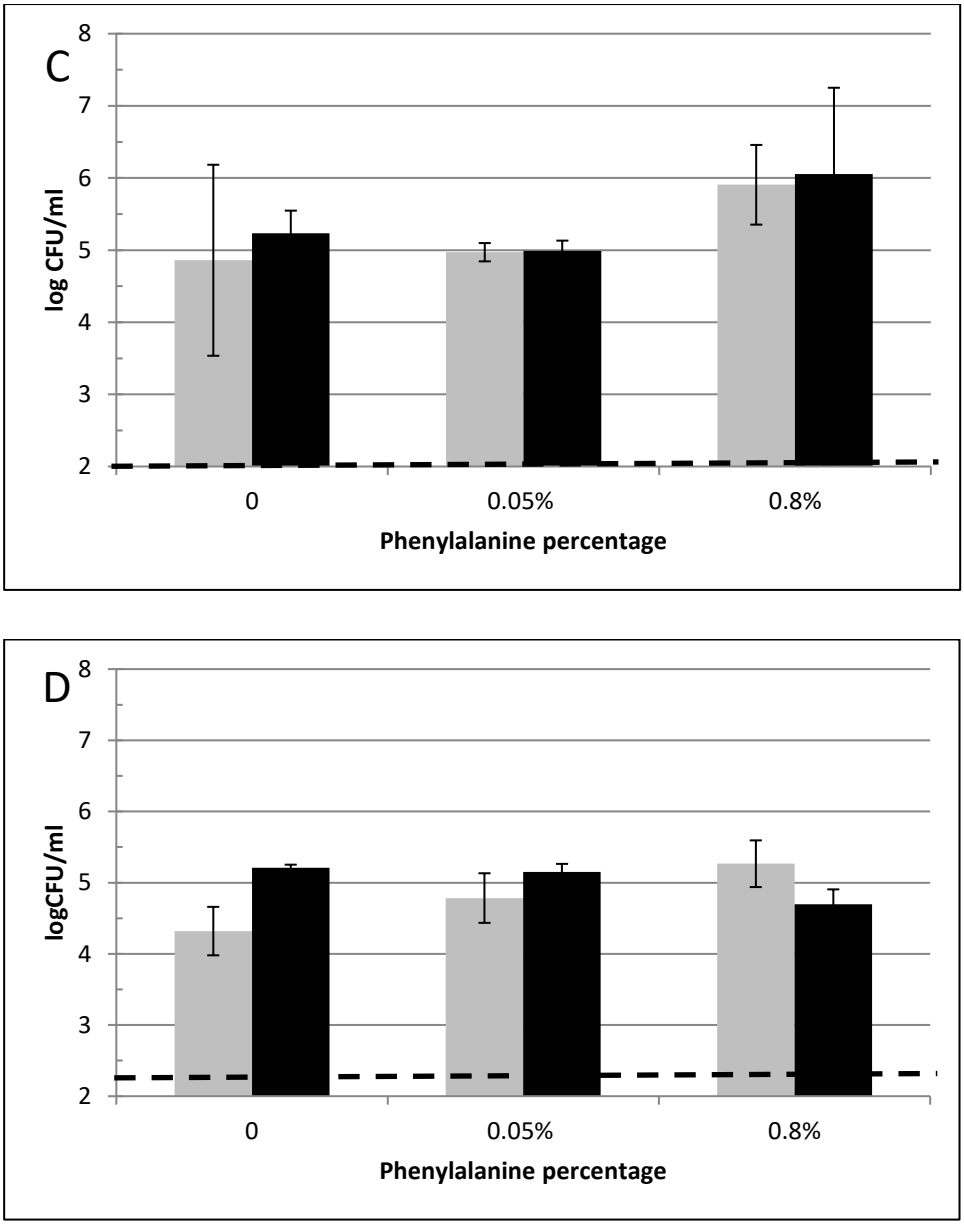


Figure 4.5. Influence of phenylalanine on growth of *G. thermoglucosidans* TNO-09.020 in both planktonic and biofilm phase. Data points represent CFU/ml in tryptone medium with no phenylalanine addition, 0.05% phenylalanine addition and 0.80% phenylalanine addition. Each bar with error bar represents the mean and standard deviation from two independent trials. Grey represents the result at 16h. Black represents the result at 24h. Biofilm phase cells (surface-attached cells) from each sample were dissolved in 6ml tryptone medium, later the cell number per ml (CFU/ml) was enumerated[3]. Any point below the detection limit (dotted line) indicates that

[3] Biofilm of 3.92 cm^2 (1.4 * 1.4 cm^2 *2 sides) was dissolved in 6 ml medium and then CFU/ml was determined. The conversion factor between CFU/ml and CFU/cm^2 is 1.5.

no viable colonies were detected on TSA plate. **A)** Total cell counts in planktonic phase. **B)** Spore counts in planktonic phase. **C)** Total cell counts in biofilm phase. **D)** Spore counts in biofilm phase.

Taken together, this study's results showed a discernible difference between biofilm-phase cells and planktonic-phase cells at different development phases of GT20, both in terms of growth patterns and expression profiles. Moreover, genes relating to methionine biosynthesis were upregulated in early biofilm development. This is associated with biofilm mass formation, but exact biological processes are not fully elucidated on the basis of this study. Lastly, phenylalanine degradation-related genes were significantly upregulated in the late biofilm development-phase cells. However, the link between phenylalanine availability and biofilm development is not concluded.

4.4 Discussion

The present study aimed at investigating the biofilm development mechanism of thermophilic sporeformers from dairy concentrate-processing environments. For this purpose, a whole genome transcriptomic study during biofilm development of a dairy concentrate-processing plant isolate was conducted.

The observation that total viable counts, in the late phase of biofilm development, dropped dramatically, and even reached the same level as the spore counts in the biofilm phase, can be explained in two ways. Firstly, it could be caused by the dispersal of the cells from the biofilm phase into the medium in the late biofilm-development phase. The same fluctuation of cell counts, with the resulting dispersal in the biofilm phase, has also been observed previously by Parkar et al. (2003). Secondly, the viable cells may not have been able to resist the environmental stress created by the over-crowding in the static environments in the biofilm where cells grew and therefore died in the late phase of biofilm development. Following cell death, the lysis of dead cells then could occur. It has been proven that cell lysis plays an essential role in intercellular adhesion and biofilm stability (Bayles, 2007), therefore, cell lysis is beneficial for the biofilm development. If these explanations hold true, it is essential for dairy concentrate-processing plants to set up timely cleaning practices to prevent the biofilm development proceeding into the late

development phase, since this is when both cell dispersal and cell lysis occur; the two phenomena could lead to either the recontamination of the product or to the persistence of the biofilm in the processing environment.

Two different spore measurement techniques were used to allow for a differentiation of the heat stability of the spores. The comparison of the outcomes of these two measurements indicated that the spores from the early growth phase (particularly in the planktonic phase) are less resistant to heat compared to the spores from the late phase. It would be worthwhile to further investigate the spore maturation process of thermophilic sporeformers. It is known that the matured spores are more resistant to heat or other environmental stresses compared to the spores in the early development stages (Sanchez-Salas et al., 2011). This suggests as a third reason that if the food processing industries intend to eradicate the spores, they should take hygienic measures to eliminate the possibility of the spores reaching a maturation stage during food processing.

The genome-wide transcriptome analysis points to the molecular processes involved in the biofilm development of the thermophilic sporeformers. The genes involved in methionine biosynthesis are upregulated in biofilm-phase cells at the early stage of biofilm formation. After observing an effect of the supplementation to the growth medium of amino acids from the methionine biosynthesis pathway on biofilm mass formation at late biofilm development phase, experiments were further done to examine whether methionine is linked to the secretion of AI-2, a quorum-sensing molecule side-product from this pathway and which has been reported to be related to biofilm attachment (Lebeer et al., 2007). However, according to the AI-2 assay, the presence of AI-2 in GT20 culture could not be detected. To avoid false negative judgment of the AI-2 assay, the possible constraints of the assay were considered. Other research indicates that low pH, low cell density, and the presence of glucose might influence the results (DeKeersmaecker and Vanderleyden, 2003). However, these factors were not relevant in the experiments conducted in this research. The samples taken had a pH value of 7, came from full-grown cultures with sufficient cells, and showed no measurable presence of glucose (data not shown). Nevertheless, it cannot yet be excluded that GT20 produced very low amounts of AI-2 which were below the detection limit of the current method. The upregulation of methionine biosynthesis at early stage is now only proven in this study to be linked

to the increased biofilm mass formation measured at late stage of biofilm development phase. Future experiments could be directed at exploring the mechanism behind this link.

A comparison of the results of this transcriptomic study with those of similar studies on different thermophilic bacilli revealed that the gene categories described as being associated with biofilm formation in other studies overlap very little with those found in this study. Despite the difficulty in comparing different transcriptomic studies on biofilm development, which make use of different methods and model organisms, the discrepancy of the results of varies studies does suggest that there is no one specific single pattern of gene expression relating to biofilm development that is independent of the biofilm-forming organisms and circumstances where biofilms are formed. This opinion has also been previously expressed in several papers (Lazazzera, 2005; Beloin and Ghigo, 2005; Kjelleberg and Givskov, 2007).

Nevertheless, the upregulation of some of the genes or gene clusters found in this study are supported by existing knowledge. For example, upregulation of genes related to fatty acid synthesis indicates that cells in biofilm are undergoing stressful situations and need to maintain the membrane fluidity by enhancing fatty acid synthesis. The importance of lipid metabolism in matrix production has also been found by Sadiq (2017) for *B. licheniformis*, a facultative thermophilic sporeformer. SdpR regulon encodes an auto-repressor which is associated with the regulation of protection against SdpC (excretes extracellular toxic peptide). In the biofilm, it can be extrapolated that *sdpR* is upregulated to protect neighbouring cells against SdpC toxin. Moreover, *sdpC* is shut off once the global regulator Spo0A, the master regulator for entry into sporulation in *Bacillus subtilis*, reaches a critical concentration (Stragier, 2006). This agrees with the observation that only the upregulation of *sdpR,* and not of *sdpC* was observed in the expression profile of biofilm-phase cells.

In conclusion, this global gene expression study provides insights into the characteristics of growth and gene expression of the dairy processing isolate GT20 during biofilm development. Some of these insights highlight the importance of earlier cleaning for the practical prevention of biofilms or spores in dairy concentrate-processing environments. Moreover, other observations arising from this study, in particular the upregulation of genes involved in methionine acquisition in the early biofilm formation phase, and the upregulation of phenylalanine

degradation-related genes in the late biofilm formation phase, provide interesting leads for further investigation.

4.5 Experimental procedures

Bacterial strains, media, and growth conditions

The strains were obtained from the -80 °C stock. Tryptone Soy Broth (TSB), Tryptone Soy Agar (TSA, plates) (Tritium Microbiologie, the Netherlands), and Tryptone (Tritium Microbiologie, the Netherlands) were used as basic growth medium. Demi water (Tritium Microbiologie, the Netherlands) was used for dilution of media in growth assays.

RNA isolation

Biofilm samples were collected at 5 time points in triplicate and immediately frozen in nitrogen. Samples were stored at −80°C until they were further processed as described previously (Zoetendal et al., 2006).

Design of the *Geobacillus thermoglucosidans* TNO-09.020 (GT20) microarray

The probe sets covering the bacterial genome were designed by Roche NimbleGen (US) according to the company's protocols and algorithms, using the custom array design (Prok Expr 12x135K Custom Arr Del = 12 arrays x 135,000 probes per slide). Max 10 probes per transcript were designed and were randomly spread over the entire transcript length, with 3 copies per probe at randomized positions on the arrays. All probes were manufactured directly onto the slides via photolithographic synthesis of 60-mer oligonucleotides. Probe sets are based on the following whole genome sequences deposited at NCBI Genbank (accession: NZ_CM001483 NZ_AJJN01000000). The probes had an average GC content of 47.1 ± 8.4%.

Labelling of cDNA, hybridization, and scanning

cDNA was synthesized and labelled as described by Ávila-Pérez et al. (2010). Probes were hybridized to the oligos on the array using the NimbleGen hybridization system

(Roche NimbleGen, US). Image analysis and data processing was performed as described previously (Yzerman et al., 2010).

Labelling of cDNA

Fluorescently labelled cDNA was prepared from 12.5 µg of total RNA by random hexamer pd$(N)_6$ primer (Roche, Mannheim, Germany) polymerization using Superscript II reverse transcriptase (Thermo Fisher Scientific, US). Concentrations of nucleotides in labelling reaction mixture were 0.4 m*M* dATP, dGTP, dCTP and 0.2 m*M* dTTP. The final concentration of Cy3-dUTP or Cy5-dUTP (GE Healthcare, US) was 0.1 m*M*. Unincorporated Cy-dye-labeled dUTP, dNTPs primers and salts were removed by purification with AutoSeq G50 columns (GE Healthcare, US).

Hybridization and image analysis

Microarray slides were incubated for 45 min at 42°C with pre-hybridization solution (1% bovine serum albumin, 5× SCC and 0.1% SDS, filtered), washed three times with milliQ water, and dried using a nitrogen flow. The Cy3-labelled cDNA from untreated cells sampled at 30°C, was mixed with Cy5-labelled cDNA from heat-treated cells for all hybridizations. After 2 min of denaturation at 95°C, hybridizations of microarray slides were performed overnight at 42°C in 40 µl of EasyHyb buffer (BioCat, Heidelberg, Germany) with both labelled cDNAs and 2.5 mg/ml yeast tRNA (final concentration). Microarray slides were washed at room temperature for 10 s in 1× SSC/0.2% SDS at 37°C, 10 s in 0.5× SSC at 37°C, and twice for 10 min in 0.2× SSC at room temperature. Slides were dried using a nitrogen flow and scanned with a ScanArray 5000 laser scanner (Perkin Elmer Life Sciences, US). The TIFF images were quantified with the software package ImaGene version 5.6.1 (BioDiscovery, El Segundo, US).

Data processing and normalization

The customized oligo-nucleotide microarrays (Roche NimbleGen, US) contain135,000 features (10 oligo's per gene, each gene in triplicate) covering 4334 annotated genes of *G. thermoglucosidans* TNO.09-020 (Zhao et al., 2012). The spot-intensity data of the set of hybridized and scanned arrays was normalized by log scale robust multi-array analysis (RMA) (Irizarry et al., 2003) resulting in data matrix with normalized

expression values. Average of the triplicate values for each gene was used for further analysis.

Statistical analysis

The data analysis was performed in two sequential steps. For the first step, the neighbourhood co-regularized spectral clustering algorithm developed by Tsivtsivadze et al. (2013) was used. Spectral clustering was applied on the data multiple times with varying parameters and results were aggregated in co-occurrence matrix visualized as a heat map (Figure 4.4). Clustering labels were retrieved based on the co-occurrence matrix using a probabilistic decomposition algorithm developed by Ter Braak et al. (2009).

The first step revealed the clustering structure of the data. The most important clustering features were discovered using Joint Embedding Learning and Sparse Regression algorithm developed by Hou et al. (2011). This method uses regularized spectral regression to calculate the importance of the weighting coefficients and ranks them based on their importance with respect to clustering labels. The method is unsupervised because, unlike traditional supervised methods, it does not need labels. It uses information retrieved from the first step as an indication of number of clusters, instead of using labels in a supervised way.

Growth assay in static biofilm system

To study thermophilic dairy biofilms on a laboratory scale, a standing steel biofilm model system was developed. This biofilm system included a sterile, vertically standing, 14 x 14 mm stainless steel coupon (Grade 316) in a well of a sterile 12-well plate (Falcon, Becton Dickinson, France). The plate was incubated in a tight plastic bag containing a wet paper towel to limit evaporation of the culture media. The wells were half filled with 3ml of 1.05% tryptone medium, which was inoculated with 1% (v/v) overnight culture of GT20. The plates were wrapped in a plastic bag and incubated for 4h, 8h, 10h, 16h, 24h, and 30h at 65°C. The total number of bacterial cells present in the tryptone medium or attached to the surface of the stainless-steel coupon, was determined by CFU counts. The coupons were washed in sterile 1.05% tryptone 3 times, the fluid was discarded. Then they were placed in 50 ml tubes filled with 6 ml 1.05% tryptone and 0.5 g of glass beads (100 µm diameter). Tubes were

vortexed for 1 min to detach the cells from the stainless-steel coupon to collect the biofilm phase cells. Serial dilutions were made and plated on TSA plates for counting after 24h of incubation at 55°C. Biofilm formation was assessed in biological duplicates and technical triplicates. For enumeration, 10 µl culture was taken from either 3ml medium or the 6ml medium in which the biofilm was dissolved of the 3.92 cm^2 surface area. Serial dilutions were made and 10 µl was plated. Detection limit is 2.0 log CFU/ml (100 cells per ml) in the 3 ml medium samples. For the dissolved biofilm this detection limit was also 2.0 log CFU/ml, which equals 2.18 log CFU/cm^2 (600 cells in 6 ml fluid coming from 3.92 cm^2 surface which equates to 152 cells per cm^2). In the experiments reported in figure 4.1, due to the low cell counts, 100 µl culture was taken instead of 10 µl, there the detection limit was 1 log CFU/ml in planktonic phase, 1.17 log CFU/ cm^2 in biolfilm phase.

Preparation of cell-free culture fluid

GT20 was grown at 55°C in tryptone broth 1.05%. Cell-free culture fluids were prepared by removing the cells from the growth medium by centrifugation at 15,000 rpm for 5 min in a micro-centrifuge. The cleared culture fluids were passed through 0.2 mm Millipore (Merck, US) and stored at -20°C. Cell-free culture fluids containing *V. harveyi* autoinducer-2 were prepared from *V. harveyi* strain BB152 (autoinducer 12, autoinducer 21). The *V. harveyi* strains were grown overnight at 30°C with aeration in AB (autoinducer bioassay) medium (Greenberg et al., 1979). Cell-free culture fluids from *V. harveyi* were prepared from the overnight culture exactly as described above for GT20.

AI-2 assay

The method is adopted from the AI-2 assay described by Surette and Bassler (1998). Cell-free culture fluids from GT20 were tested for the presence of signalling substances that could induce luminescence in the *V. harveyi* reporter strain BB170 or BB886. In the assays, 10 µl of cell-free culture fluids from GT20 harvested as described above were added to 96-well microtiter dishes. The *V. harveyi* reporter strain BB170 was grown for 16h at 30°C with aeration in AB medium and diluted 1:5,000 into fresh AB medium, and 90 µl of the diluted cells was added to the wells containing the GT20 cell-free culture fluids. Positive control wells contained 10 µl of cell-free culture fluid from strain *V. harveyi* BB152 (autoinducer-12, autoinducer-21).

Negative control wells contained 10 µl of sterile growth medium. The microtiter dishes were shaken in a rotary shaker at 175 rpm at 30°C. Every hour, luminescence light production was measured by using a Tecan infinite 520 (Thermo Fisher Scientific, US). Cell density of *V. harveyi* was measured at OD600. The measurements of luminescent light production at 6h time point was taken for further analysis.

4.6 Acknowledgments

The project was funded by TI Food and Nutrition, a public-private partnership on precompetitive research in food and nutrition. The founders had no role in study design, data collection and analysis, or preparation of the manuscript. Microarray slides were kindly provided by Nimblegen (Roche NimbleGen, US). We would like to extend our special thanks to Sultan Imangaliyev for his contribution to the microarray data analysis. We would also like to thank Jos Boekhorst for his contribution to discussions about microarray data analysis.

4.7 References

Abee, T., Á. T. Kovács, O. P. Kuipers and S. van der Veen. 2011. Biofilm formation and dispersal in Gram-positive bacteria. Current Opinion in Biotechnology. 22:172-179.

Ávila-Pérez, M., J. B. van der Steen, R. Kort and K. J. Hellingwerf. 2010. Red light activates the sigmaB-mediated general stress response of *Bacillus subtilis* via the energy branch of the upstream signaling cascade. Journal of Bacteriology. 192:755-762.

Bayles, K. W. 2007. The biological role of death and lysis in biofilm development. Nature Reviews Microbiology. 5:721-726.

Beloin, C. and J. Ghigo. 2005. Finding gene-expression patterns in bacterial biofilms. Trends in Microbiology. 13:16-19.

Burgess, S. A., D. Lindsay and S. H. Flint. 2010. Thermophilic bacilli and their importance in dairy processing. International Journal of Food Microbiology. 144:215-225.

DeKeersmaecker, S. C. J. and J. Vanderleyden. 2003. Constraints on detection of autoinducer-2 (AI-2) signaling molecules using *Vibrio harveyi* as a reporter. Microbiology. 149:1953-1956.

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

Greenberg, E. P., J. W. Hastings and S. Ulitzur. 1979. Induction of luciferase synthesis in *Beneckea harveyi* by other marine bacteria. Archives of Microbiology. 120:87-91.

Hou, C., F. Nie, D. Yi and Y. Wu. 2011. Feature selection via joint embedding learning and sparse regression. Page 1324 in IJCAI proceedings-international joint conference on artificial intelligence.

Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs and T. P. Speed. 2003. Summaries of affymetrix GeneChip probe level data. Nucleic Acids Research. 31:e15.

Kjelleberg, S. and M. Givskov. 2007. The Biofilm Mode of Life: Mechanisms and Adaptations. Horizon Bioscience, UK.

Lazazzera, B. A. 2005. Lessons from DNA microarray analysis: The gene expression profile of biofilms. Current Opinion in Microbiology. 8:222-227.

Lebeer, S., S. C. J. De Keersmaecker, T. L. A. Verhoeven, A. A. Fadda, K. Marchal and J. Vanderleyden. 2007. Functional analysis of luxS in the probiotic strain *Lactobacillus rhamnosus* GG reveals a central metabolic role important for growth and biofilm formation. Journal of Bacteriology. 189:860-871.

Parkar, S. G., S. H. Flint and J. D. Brooks. 2003. Physiology of biofilms of thermophilic bacilli potential consequences for cleaning. Journal of Industrial Microbiology & Biotechnology. 30:553-560.

Sadiq, F. A., S. Flint, Y. Li, K. Ou, L. Yuan and G. Q. He. 2017. Phenotypic and genetic heterogeneity within biofilms with particular emphasis on persistence and antimicrobial tolerance. Future Microbiology. 12:1087-1107.

Sanchez-Salas, J. L., B. Setlow, P. Zhang, Y. Q. Li and P. Setlow. 2011. Maturation of released spores is necessary for acquisition of full spore heat resistance during *Bacillus subtilis* sporulation. Applied and Environmental Microbiology. 77:6746-6754.

Scheldeman, P., A. Pil, L. Herman, P. De Vos and M. Heyndrickx. 2005. Incidence and diversity of potentially highly heat-resistant spores isolated at dairy farms. Applied and Environmental Microbiology. 71:1480-1494.

Scott, S. A., J. D. Brooks, J. Rakonjac, K. M. R. Walker and S. H. Flint. 2007. The formation of thermophilic spores during the manufacture of whole milk powder. International Journal of Dairy Technology. 60:109-117.

Stragier, P. 2006. To kill but not be killed: A delicate balance. Cell. 124:461-463.

Surette, M. G. and B. L. Bassler. 1998. Quorum sensing in *Escherichia coli* and *Salmonella typhimurium*. Proceedings of the National Academy of Sciences. 95:7046-7050.

Ter Braak, C. J. F., Y. Kourmpetis, H. A. L. Kiers and M. C. A. M. Bink. 2009. Approximating a similarity matrix by a latent class model: A reappraisal of additive fuzzy clustering. Computational Statistics & Data Analysis. 53:3183-3193.

Tsivtsivadze, E., H. Borgdorff, J. van de Wijgert, F. Schuren, R. Verhelst and T. Heskes. 2013. Neighborhood co-regularized multi-view spectral clustering of microbiome data. Page 80-90 in: Zhou, Z.H. and F. Schwenker. (eds) Partially Supervised Learning. PSL 2013. Lecture Notes in Computer Science, vol 8183. Springer, Berlin, Heidelberg.

Waite, R. D., A. Papakonstantinopoulou, E. Littler and M. A. Curtis. 2005. Transcriptome analysis of *Pseudomonas aeruginosa* growth: Comparison of gene expression in planktonic cultures and developing and mature biofilms. Journal of Bacteriology. 187:571-6576.

Wang, L., R. Grau, M. Perego and J. A. Hoch. 1997. A novel histidine kinase inhibitor regulating development in *Bacillus subtilis*. Genes & Development. 11:2569-2579.

Watterson, M. J., D. J. Kent, K. J. Boor, M. Wiedmann and N. H. Martin. 2014. Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest. Journal of Dairy Science. 97:2487-2497.

Wray, L. V. Jr, A. E. Ferson, K. Rohrer and S. H. Fisher. 1996. TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. Proceedings of the National Academy of Sciences. 93:8841-8845.

Yuan, D., G. Liu, D. Ren, D. Zhang, L. Zhao, C. Kan, Y. Yang, W. Ma, Y. Li and L. Zhang. 2012. A survey on occurrence of thermophilic bacilli in commercial milk powders in China. Food Control. 25:752-757.

Yzerman, E., J. W. den Boer, M. P. M. Caspers, A. Almal, B. Worzel, W. van der Meer, R. Montijn and F. Schuren. 2010. Comparative genome analysis of a large dutch *Legionella pneumophila* strain collection identifies five markers highly correlated with clinical strains. BMC Genomics. 11:433.

Zhang, W., D. E. Culley, L. Nie and J. C. M. Scholten. 2007. Comparative transcriptome analysis of *Desulfovibrio vulgaris* grown in planktonic culture and mature biofilm on a steel surface. Applied Microbiology and Biotechnology. 76:447-457.

Zhao, Y., M. P. M. Caspers, T. Abee, R. J. Siezen and R. Kort. 2012. Complete genome sequence of *Geobacillus thermoglucosidans* TNO-09.020, a thermophilic sporeformer associated with a dairy-processing environment. Journal of Bacteriology. 194:4118-12.

Zhao, Y., M. P. M. Caspers, K. I. Metselaar, P. de Boer, G. Roeselers, R. Moezelaar, M. N. Nierop Groot, R. C. Montijn, T. Abee and R. Kort. 2013. Abiotic and microbiotic factors controlling biofilm formation by thermophilic sporeformers. Applied and Environmental Microbiology. 79:5652-5660.

Zoetendal, E. G., C. C. G. M. Booijink, E. S. Klaassens, G. H. J. Heilig, M. Kleerebezem, H. Smidt and W. M. de Vos. 2006. Isolation of RNA from bacterial samples of the human gastrointestinal tract. Nature Protocols. 1:954-959.

Chapter 5 Genomic comparison of dairy and non-dairy associated thermophilic sporeformers

Yu Zhao

5.1 Summary

To gain a better understanding of genomic features of thermophilic sporeformers that may facilitate spoilage in dairy processing environments, a comparative genomic analysis was performed on lactose utilization, proteolysis, and biofilm formation. The genomes of the 22 thermophilic bacilli (16 *Geobacillus* spp., 4 *Anoxybacillus* spp. and 2 *Caldibacillus* spp.) isolated from both dairy associated and non-dairy associated environments were analysed and compared. Whole genome comparison distributed the 22 genomes in five clades, including two clearly distinctive and well supported clades representing *Geobacillus* spp. and *Anoxybacillus* spp. Most of the thermophilic, sporeforming isolates (except *Caldibacillus debilis* DSM16016) that were able to utilize lactose could also grow on skim milk plates. The presence of genes involved in lactose-utilization correlated well with growth on lactose media for *Anoxybacillus flavithermus* and *Geobacillus thermoglucosidans* strains, but this correlation was not observed for all thermophilic sporeformer species. For example, 3 out of 5 *Geobacillus stearothermophilus* had the ability to utilize lactose, but in their genomes, genes encoding the lactose hydrolysing β-galactosidase appeared absent. This suggests that lactose utilization by some *Geobacillus stearothermophilus* strains may involve a pathway that includes a different β-galactosidase. Although all thermophilic sporeformers analysed appeared to encode a complete pathway for proteolysis, not all strains could grow on skim milk plates. On the other hand, biofilm formation is predominantly determined by the environmental conditions, rather than strain specific genetics. Random forest-based gene-trait matching using the differential growth on skim milk plates, led to the identification of 100 candidate genes of which presence and absence is associated with this phenotype. The gene with the strongest phenotype association encodes a niacin transporter, which was shown to be relevant for growth in the dairy environment. These results revealed several links between genomic signatures of thermophilic bacilli and their spoilage potential in dairy industry. Moreover, they exemplify the genomic and phenotypic diversity of this group of bacteria that plays an important role in the dairy industry.

5.2 Introduction

Thermophilic sporeformers are not pathogenic but can proliferate in dairy processing environments with high temperatures and have the potential to spoil dairy products. Therefore, strict specifications for thermophilic sporeformers in dairy powder products are applied. For example, sporeformer specifications of international customers for dairy powders have been reported as follows: aerobic mesophilic and thermophilic spore counts < 500 to < 1,000 CFU/g for skim milk powder, non-fat dry milk, and whole milk powder destined for infant formula; and < 500 to < 2,000 CFU/g for aerobic thermophilic spores in skim milk powder and whole milk powder destined for recombined products or UHT products (Watterson et al., 2014). These strict sporeformer specifications are very difficult to achieve, presenting an important challenge to the dairy industry worldwide. Among the sporeformers able to survive in dairy-powder processing environments, thermophilic sporeformers that are able to grow at high temperatures are a particular concern with regard to spoilage of dairy powder products processed at high temperatures (Watterson et al., 2014). Thermophilic sporeformers have a number of properties that contribute to spoilage in dairy processing environments, four of the main ones are the following. First, there are multiple sources of contamination of thermophilic sporeformers in dairy-powder processing environments such as soil, animal feed, and other ingredients (Carlin, 2011). Second, thermophilic sporeformers are fast growers (doubling time ranging from 15 to 20 minutes under optimal growth conditions), indicating that upon initiation of their proliferation in dairy-processing environment they can rapidly surpass the contamination specifications for dairy powders. Third, thermophilic sporeformers can persist in the dairy-powder processing environments in biofilms (Burgess et al., 2011). Lastly, spores of thermophilic bacteria can survive the harsh conditions encountered in dairy manufacturing processes, including heat and chemical exposure, and mechanical disruption, and are thereby hard to remove and can end-up in dairy products, like milk powder. Since numerous factors support contamination by thermophilic sporeformers, to meet strict customer specifications the reduction of their counts in finished dairy powder products requires a systematic approach, which tackles adaptation, survival and accumulation of sporeforming organisms, at the farm, during processing and during distribution.

The species *Anoxybacillus* spp. and *Geobacillus* spp. are the predominant aerobic thermophilic sporeformers isolated from dairy powder products (Flint et al., 1997; Zhao et al., 2013). *Anoxybacillus* spp. is mainly isolated - in the early processing line like at heat exchangers, while *Geobacillus* spp. predominates in the late processing line like at evaporators and in end-products (Zhao et al., 2013). This suggests that *Geobacillus* spp. may, on a population basis, have a better ability to survive in dairy-powder processing environments. Such differences in adaptation, survival and growth abilities are highly relevant for measures to address the spoilage caused by thermophilic sporeformers. Our knowledge of these differences is however relatively limited, since those thermophilic sporeformers are rarely studied, especially when compared to mesophilic sporeformers (Switt et al., 2014). Comparative genomic analysis is an effective method to study the genomic features of thermophilic sporeformers in terms of adaptation, survival- and growth abilities in dairy-processing environments (Goh et al., 2014). Recently, a number of *Anoxybacillus* spp. and *Geobacillus* spp. genomes, representing strains isolated from dairy and non-dairy associated environments, were sequenced and their genomes were published (Berendsen et al., 2016; Caspers et al., 2016).

Here we applied comparative genomics approaches to study the genomic features of thermophilic sporeformers related to their spoilage capability in dairy-powder processing environments. We analysed the relation between lactose utilization and proteolytic activity with the ability to grow on skim milk plates. Moreover, we discuss the predictive value of the presence of known lactose-utilization and proteolysis associated genes for the ability of thermophilic sporeformers to grow on skim milk plates. In addition, biofilm forming capability of the selected thermophilic sporeformers was also assessed using comparative genomic approaches and biofilm assay. Finally, gene trait matching identified candidate genomic features of thermophilic sporeformers that imply potential relevance for spoilage in skim milk, but further isogenic mutant controlled experiments are needed to further assess the importance of those candidate genomic features.

5.3 Results and discussion

5.3.1 Phylogenetic analyses divide strains into 5 major clades

A maximum likelihood phylogeny tree was created based on multiple sequence alignments of the core genes (as determined using OrthoMCL) of the 22 strains (Figure 5.1). The 22 thermophilic strains analysed here can be clustered into five clades, representing the species *Caldibacillus debilis*, *Geobacillus vulcani*, *Anoxybacillus flavithermus*, *Geobacillus thermoglucosidans* and *Geobacillus stearothermophilus,* respectively. The genome of *G. vulcani* clustered with the genome of the reference strain *Bacillus subtilis* 168. *Geobacillus caldoxulosilyticus* and *Geobacillus toebii* are clustered together with *G. thermoglucosidans* strains. One notable observation is that *G. stearothermophilus* 10 did not cluster with other *G. stearothermophilus* strains, but clustered with *Geobacillus thermoleovorans*. This observation is in apparent agreement that in API tests, *G. stearothermophilus* 10 displays a distinct sugar utilization pattern relative to other *G. stearothermophilus* strains (Figure 5.2). It is suspected that *G. stearothermophilus* 10 does not belong to the *G. stearothermophilus* species and its taxonomic classification should be readdressed. Notably, the difficulty in identifying *G. stearothermophilus* has been reported before and we would recommend employing the high-resolution core genome approach to refine the classification of this species (Burgess et al., 2017).

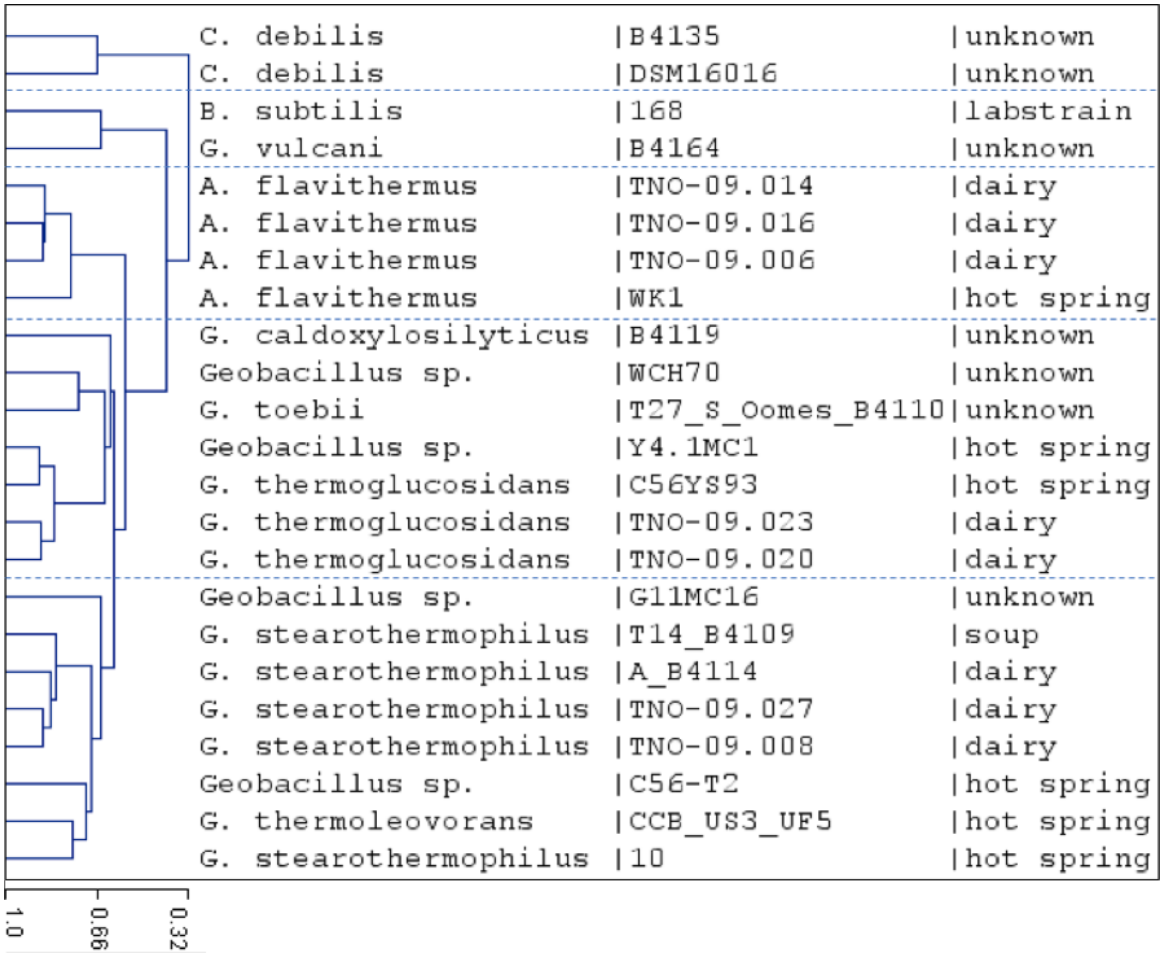


Figure 5.1 Maximum likelihood phylogeny from presence or absence of orthologous genes from the 22 thermophilic strains. The clade containing *Bacillus subtilis* 168 was selected to root the tree.

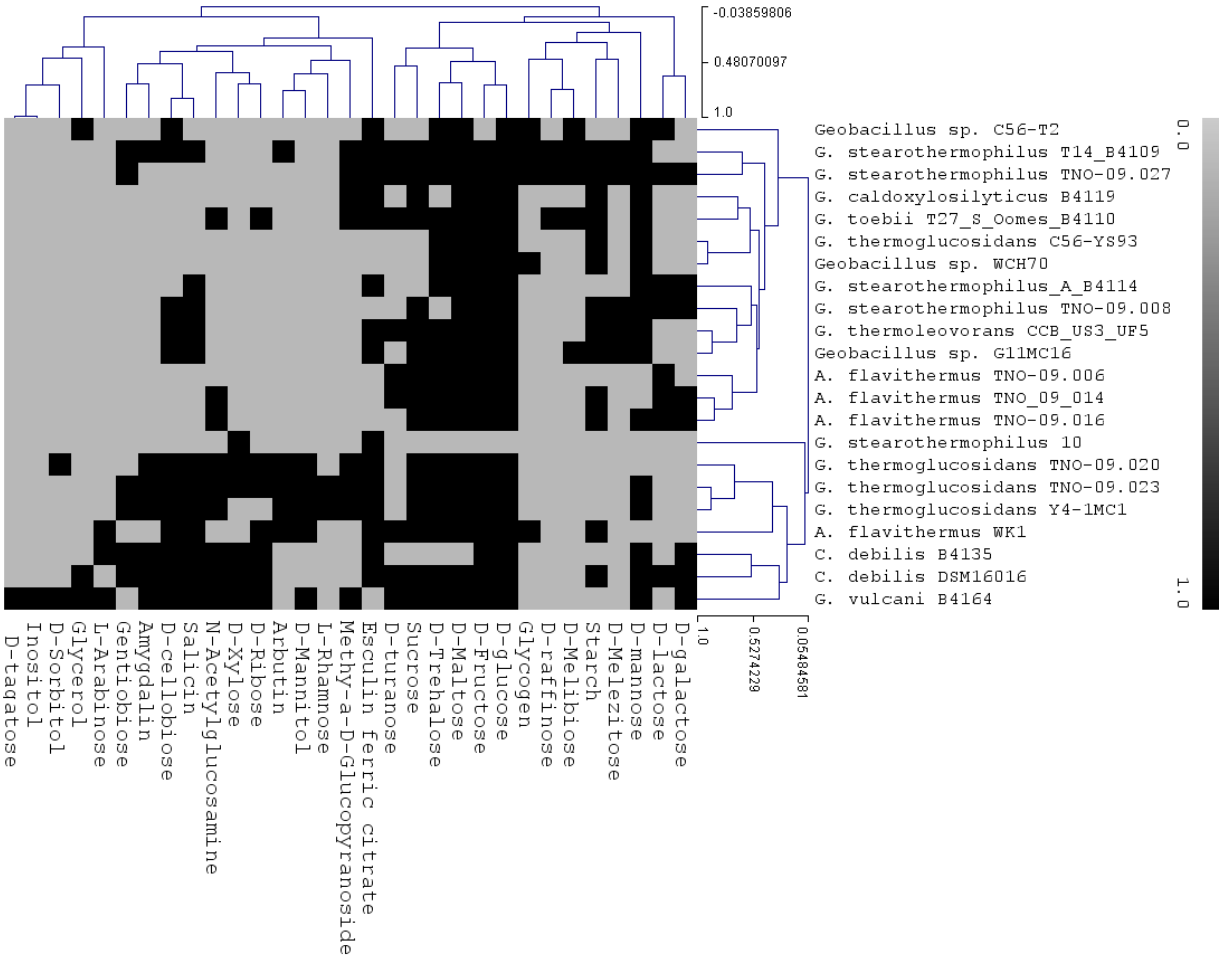


Figure 5.2 hierarchical clustering of the strains and their carbohydrate utilization. The tree with the highest log likelihood is shown in the figure. High to low correlation of the strain and sugar is illustrated in the dendrograms as value from 1 to 0.

It is notable that the strains belonging to the same species and with dairy origins are clustered in one clade (Figure 5.1). For example, the three *A. flavithermus* and two *G. thermoglucosidans* strains, originating from the same dairy environment, are clustered together, suggesting that the specific genomic characteristics in these strains would reflect their adaptation to the dairy environment (Rohmer et al., 2011; Ceapa et al., 2015). To confirm this hypothesis, future research could expand the number of total number of isolates as well as the number of dairy-processing locations from which these isolates are collected.

5.3.2 Lactose utilization capability is an effective indication of the ability for growth of thermophilic sporeformers on skim milk plates

A growth assay on skim milk plates demonstrated that of the 22 strains, 9 showed growth and 13 did not (Table 5.1). Lactose is the major carbohydrate in milk. Zhao et al. (2017) describes how the *G. thermoglucosidans* strain depends on lactose utilization capability of *A. flavithermus* for rapid growth in skim milk. In order to study the relation between lactose utilization capability and growth in skim milk for the selected 22 thermophilic sporeformers, their lactose utilization capacity was examined using an API test. Moreover, the production of active β-galactosidase, an important enzyme which cleaves the glycosidic bond in D-lactose present in milk, can hydrolyse X-gal (an analogue of lactose). Except for *Geobacillus vacani* B4164, the results of the API test and X-gal assay (Table 5.1) showed a good match. Lactose can be imported not only by a permease but also by PTS import system (Solopova et al., 2012). In this study, after searching in all the genomes, no PTS cassettes related to lactose importation were found (Data not shown). β-galactosidase is the primary enzyme which is used to indicate lactose utilization for the selected strains.

Table 5.1 Phenotypic characterization of the strains used in this study.

Strain	*Growth on skim milk agar plate (growth +/no growth -)*	*Growth on MPCA (growth +/no growth -)*	*On MPCA (halo +/ no halo -)*	*On X-gal plate (blue +/white -) on TSA* [1]	*Lactose (API test)*	*Lactose utilization and transportation related genes*[2]
Geobacillus thermoglucosidans C56YS93	-	+	+	-	-	-
Anoxybacillus flavithermus WK1	-	+	+	-	-	-
Geobacillus thermoglucosidans TNO-09.020	-	+	++	-	-	-
Geobacillus thermoglucosidans TNO-09.023	-	+	++	-	-	-
Geobacillus thermoglucosidans Y4.1MC1	-	+	++	-	-	-
Anoxybacillus flavithermus TNO-09.006	+	+	+	++	+	+
Anoxybacillus flavithermus TNO-09.014	+	+	+	++	+	+

Anoxybacillus flavithermus TNO-09.016	+	+	+	++	+	+
Geobacillus sp. C56-T2	+	+	-	+	+	+
Geobacillus sp. G11MC16	-	+	-	-	-	_
Geobacillus thermoleovorans CCB_US3_UF5	-	+	+	-	-	_
Geobacillus sp. WCH70	-	+	+	-	-	_
Geobacillus caldoxylosilyticus B4119	-	+	-	-	-	+
Caldibacillus debilis B4135	-	+	-	-	-	_
Geobacillus stearothermophilus A_B4114	+	+	-	+	+	_
Geobacillus stearothermophilus T14_B4109	-	+	++	-	-	_
Geobacillus toebii T27_S_Oomes_B4110	-	+	+	-	-	_
Geobacillus vulcani B4164	+	+	++	++	_	_
Geobacillus stearothermophilus TNO-09.027	+	+	-	+	+	_
Geobacillus stearothermophilus TNO-09.008	+	+	-	+	+	_
Caldibacillus debilis DSM16016	-	+	++	+	+	_
Geobacillus stearothermophilus 10	+	+	++	-	-	+

[1] "-"colony appears as white colour, "+" blue colour only appears in the middle of the colony, "++" the whole colony is blue

[2] presence (+) or absence (-) of lactose utilization and transportation related genes including lactose ABC transporters and β-galactosidase

Of the 9 strains that grew on skim milk plates, 7 showed positive results in the API test with exception for *G. stearothermophilus* 10 and *G. vulcani* B4164, and 8 showed positive results in X-gal assay with exception for *G. stearothermophilus* 10. In case of the 13 strains that did not grow on milk plates, 12 showed negative results in the API test and X-gal assay, with exception for *C. debilis* DSM16016. Taken together, these results indicate that not all thermophilic sporeformers are able to grow in milk, but

the strains that grow on milk quite consistently have the ability to utilize lactose and produce active β-galactosidase. Thereby, the characterization of lactose utilization and the production of active β-galactosidase can be employed as an indicator for the ability of thermophilic sporeformers to grow on (skim) milk (plates).

5.3.3 The discrepancy between the presence of lactose-utilization related genes and lactose utilization capability

Lactose fermentation contains three stages, namely: vectoral translocation and phosphorylation of the disaccharide by the multi-component, lactose-specific PTS; catabolism of lactose-6'P by energy-yielding pathways, and efflux of lactic acid from the cell (Thompson et al., 1987). All 22 genomes analysed in this study were searched for genes encoding these enzymes which can exert those functions, establishing strain specific presence or absence of putative lactose utilization associated genes (Table 5.1). There were three genes included in the analysis, namely, β-galactosidase (EC 3.2.1.23), a permease protein (lactose ABC transporter) and a lactose-binding protein (lactose ABC transporter). No PTS cassettes related to lactose importation were found, and tagatose pathway known in lactic acid bacteria was not included in this analysis. The three genes were always either all present, or all absent. API test results (Table 5.1) appeared to show some discrepancies with the results of the presence or absence analysis for lactose utilization genes. The hypothesis of those discrepancies is discussed below: First, comparative genome analysis on the 22 strains revealed that *G. stearothermophilus* 10 and *G. caldoxylo*silyticus B4119 contain the set of lactose-metabolism related genes but did not demonstrate X-gal nor lactose utilization activity. The comparative genome analysis was done with algorithm OrthoMCL, a sequence comparison tool that performs bi-directional best hit analyses to identify orthologous genes. In this analysis, identification of present but inactive pseudogenes might occur due to the fact that in itself the OrthoMCL tool does not consider domain presence, absence or intactness, and may tolerate small sequence variations (e.g. truncations or mutations). Therefore, the occurrence of errors should be checked for these 2 genomes like: (i) abnormal gene organization (e.g. genomic reorganizations or loss of promoter- or cistronic- context), (ii) dysfunctional promotors (e.g. missing essential promoter motives like TATA-boxes), (iii) lacking or mutated protein domains like signal peptides, binding- and active

domains, (iv) N-, C- terminal truncations, (v) codon mutations leading to inactivity or misfolding. To perform this analysis in full depth for all the reported lactose-metabolism related orthologous genes is a massive task. Therefore, for this paper, only the OrthoMCL generated orthologue alignments were examined. From this examination, it is observed that *G. caldoxylo*silyticus B4119, one of two strains with inactive lactose-metabolism genes, has only percentage identity ranging from 6% to 13% in the amino acid sequences of the putative lactose-metabolism related genes ((β-galactosidase (EC 3.2.1.23/OG_3249), a permease protein (lactose ABC transporter/ OG_2883) and a lactose-binding protein (lactose ABC transporter/OG_3549)) compared to the strains which had the consistence between the presence of the genes and lactose utilization activity (e.g. *A. flavithermus* TNO-09.006 and *A. flavithermus* TNO-09.016) (Table S5.1). This low percentage identity explains the absence of lactose utilization ability. On the other hand, although the sequences of *G. stearothermophilus* 10 showed percentage identity around 90% compared with the sequences of consistent strains (e.g. *A. flavithermus* TNO-09.006 and *A. flavithermus* TNO-09.016) (Table S5.1), it is shown in the phenotypic experiments that *G. stearothermophilus* 10 showed distinct sugar metabolism patterns compared to other strains (Figure 5.2). It is suspected that the inactivity of lactose utilization related genes of *G. stearothermophilus* 10 is related to environmental reasons: for example, the effect of environmental conditions such as oxygen availability and temperature, or carbon catabolite repression (CCR) on the expression and regulation of sugar utilization systems (Mekalanos, 1992; Wassarman, 2002; Rohmer et al., 2011). Second, 3 *G. stearothermophilus* strains (*G. stearothermophilus* TNO-09.027, *G. stearothermophilus* TNO-09.008, *G. stearothermophilus* A_B4114), and *C. debilis* DSM16016 showed β-galactosidase activity and lactose utilization capability. However, in the genome of those strains, neither the lactose transporting system (including an ABC transporter or lactose permease) and β-galactosidase gene, nor the functional domains of these genes could be found. This observation suggests that the three *G. stearothermophilus* strains and *C. debilis* DSM16016 may use different but equivalent cassettes for the utilization of lactose.

The phenomenon that thermophilic bacteria have more variable lactose degradation architectures and strategies is also suggested in the paper by Brumm et al. (2015). Also, within the same species, there are different genes that are responsible for

lactose metabolism for lactic acid bacteria (Thompson et al., 1987). Those alternative pathways can be used as a clue to search for alternatives in those above mentioned thermophilic sporeformers.

Contrary, among mesophilic bacteria, the carbohydrate utilization genes were found to be more conserved, particularly in case of lactose utilization (Warda et al., 2016). As described above, the thermophilic bacteria have more variable carbohydrate degradation architectures and strategies when compared to mesophilic bacteria. Therefore, when using presence of lactose-utilization related genes as a predictive factor for growth of thermophilic sporeformers in skim milk, more detailed knowledge of these alternative pathways should be acquired to accurately detect lactose-utilization genes in these species.

5.3.4 All thermophilic sporeformers sequenced encode proteolytic systems with a putative role in casein breakdown

Casein, which makes up 80% of the proteins in cow's milk, is the main source of nitrogen of microorganisms in milk (Mills and Thomas, 1981; Exterkate and de Veer, 1987). It is reported that *Bifidobacterium*'s ability to break down casein is linked to its rapid growth in milk; conversely, the lack of ability to break down casein will retard its growth (Klaver et al., 1993). To study this relationship for thermophilic sporeformers, proteolytic activity of 22 thermophilic sporeformers was analysed using MPCA (milk plate counting agar plates). The initial consideration in choosing MPCA with the addition of yeast extract for testing proteolytic activity is due to the fact that some thermophilic sporeformers cannot grow on complex milk protein. Using only the milk plates, many strains would not grow due to the inability to use lactose. Nevertheless, there is the presence of free peptide and amino acid in the MPCA which can potentially inhibit proteolytic activity since proteins are broken down in response to carbon, nitrogen or sulfur limitation (Sims, 2006; Sims and Wander, 2002). When the environment is rich in small peptides and free amino acid, the gene regulation of the overall proteolytic pathways will suppress protease and peptidase expression levels in bacteria. In *Bacillus subtilus*, for example, CodY can suppress exoprotease when cells are growing exponentially in a medium containing abundant quantities of proteins or their degradation products (Barbieri et al., 2016).

In this study, 15 out of 22 collected thermophilic sporeformers showed halo formation on MPCA plates, which indicates proteolytic ability is not suppressed. On the contrary, the strains which do not appear a halo can also not be excluded for having the potential for proteolytic activity, but the ability is likely to be limited due to the ample presence of nitrogen source in the environment. With the knowledge of this study, an improved version of MPCA plate to test proteolytic activity would be without yeast extract but with the addition of monosaccharide (e.g. glucose or galactose) to support initial growth but with the limitation of the nitrogen source.

Therefore, as for the link between growth on milk plates and the positive results on MPCA plates, analysis of the proteolytic activity did show no clear correlation. This indicates that the ability to grow on skim milk plates seems to have no clear correlation with the ability of strains in casein breakdown as measured by MPCA plates. This conflicts with the findings for dairy *Bifidobacterium* (Klaver et al., 1993). However, it is observed that *G. thermoglucosidans* strains which presented biggest halos on MPCA plates (data not shown) indicating a strong ability of casein breakdown, could outcompete other species in growth in skim milk, as long as the growth was initiated (Zhao et al., 2013). Since also a time-dependent phenomenon has been mentioned for dairy lactic acid bacteria (Donkor et al., 2007), it can be hypothesized that the extent of proteolysis varies among strains and appears to be stage-dependent for thermophilic sporeformers, meaning lactose utilization is essential for the initiation of growth in milk, and proteolytic activity is essential for the continued growth of bacteria in milk (Bachmann et al., 2010).

In order to see the link between the proteolytic activity and the genomic features in relation to proteolysis in these thermophilic sporeformers, searches for orthologous genes in all 22 genomes sequenced were performed. The objective was to screen for the presence of genes encoding proteolytic systems and proteins previously reported as facilitating casein breakdown in *Bacillus subtilis* (Michna et al., 2014). Peptidase encoding genes were identified in all 22 genomes, albeit it at highly different levels of redundancy, ranging from 38 peptidase genes in *A. flavithermus* TNO-09.006 to 67 peptidase genes in *Geobacillus vulcani* B4164. Genes unique to each species were also identified (Table S5.2). For the species *G. thermoglucosidans*, in which all strains showed the most apparent clearing zones; the species-specific genes encoding proteases encoded subtilisin-like serine protease and hydrogenase

maturation protease. Overall, even though some of the 22 strains characterized here were negative for proteolysis on MPCA agar, all strains appear to encode proteins for all major steps required for casein breakdown (Table S5.2) (i.e., cell-wall proteinase activity, peptide transport, and intracellular peptidases).

5.3.5 Biofilm forming capacity of thermophilic sporeformers

The biofilm forming capacity of strains in dairy-powder processing environments was also assessed, since biofilm formation strongly contributes to the persistence of certain spoilage organisms residing in the dairy processing line (Scott et al., 2007); therefore being also of concern in respect to contamination of dairy powder products. Among the 22 collected thermophilic sporeformers, only *G. vulcani* B4164 formed pellicles. The majority of the strains could form biofilm under at least one of the conditions tested (Table 5.2). In fact, only *A. flavithermus* TNO-09.014 did not form a biofilm under any condition tested. Although *A. flavithermus* TNO-09.014 showed turbidity in the medium, which means growth of the strain after overnight culture, under test conditions (Table 5.2), biofilm formation could not be detected in the crystal violet assay. In contrast, for *G. stearothermophilus* TNO-09.008 biofilm was detected, but the turbidity in the medium was not very apparent. This could be due to the fact that the bacteria were aggregated to the surface to form a biofilm, which resulted in a clearance of the medium. This phenomenon, whereby cells aggregate at the surface for biofilm formation, has also been described by Branda et al. (2001). It is considered that biofilm is the surface associated structure, and crystal violet assay is a good approach to evaluate this. However, the crystal violet assay cannot directly indicate the amount of bacteria population aggregating to the surface (Ramirez et al., 2015). Therefore, one should also be careful with the interpretation of the results from the crystal violet assay.

According to the genome comparison analysis, every strain has a certain number of genes involved in biofilm formation (*Bacillus subtilis* was used as a reference for selecting genes involved in biofilm formation) (Figure S5.1). The presence of the genes is an indication of the strains' potential to form biofilm. Nevertheless, biofilm formation can be influenced by different factors (Sadiq et al., 2017), such as surface for attachment (Hinton et al., 2002), time of exposure to a surface (Jimenez-Flores,

2014), and accessible nutrients (Zhao et al., 2013). This study underpins that biofilm formation is condition dependent, and the different strains did not form biofilms under all conditions tested.

Table 5.2 A. Overview of biofilm-forming capacities of all 22 thermophilic sporeformers using microtiter plate assay ("0" biofilm formation was not detectable, "1" biofilm formation was detected). **B. Overview of growth in the medium of all 22 strains using microtiter plate assay** ("0" growth was not observed, "1" growth was observed).

	A. Biofilm formation capacity of strains under the following conditions:				*B. Growth of strains in static biofilm system under the following conditions:*			
Thermophilic sporeformers	55°C		65°C		55°C		65°C	
	16h	25h	16h	25h	16h	25h	16h	25h
Geobacillus sp. C56-T2	1	0	1	1	1	1	1	1
Geobacillus sp. G11MC16	1	0	0	0	1	0	1	1
Geobacillus thermoglucosidans Y4-1MC1	1	1	1	1	1	1	1	1
Geobacillus thermoglucosidans C56-YS93	1	1	1	1	1	0	1	1
Geobacillus thermoleovorans CCB_US3_UF5	0	1	0	1	1	1	1	1
Geobacillus sp. WCH70	1	1	0	0	1	0	1	1
Geobacillus caldoxylosilyticus B4119	1	1	0	1	1	0	1	0
Caldibacillus debilis B4135	1	0	0	0	1	1	1	1
Geobacillus stearothermophilus A_B4114	1	1	1	1	1	0	1	0
Geobacillus stearothermophilus T14_B4109	0	1	1	1	0	0	1	0
Geobacillus toebii T27_S_Oomes_B4110	0	0	1	1	1	1	1	1
Geobacillus vulcani B4164	1	1	0	0	1	1	0	0
Anoxybacillus flavithermus TNO-09.014	0	0	0	0	1	1	1	1
Anoxybacillus flavithermus TNO-09.016	1	1	1	1	1	0	1	0
Geobacillus thermoglucosidans TNO-09.023	1	1	1	1	1	1	1	1
Geobacillus stearothermophilus TNO-09.027	0	0	1	1	1	0	1	0
Anoxybacillus flavithermus TNO-09.006	1	1	1	1	1	1	1	0
Geobacillus stearothermophilus TNO-09.008	1	1	1	1	0	0	0	0
Geobacillus thermoglucosidans TNO-09.020	0	1	1	1	1	1	1	1
Caldibacillus debilis DSM16016	0	1	0	1	1	1	1	1
Anoxybacillus flavithermus WK1	1	1	0	0	1	1	1	1
Geobacillus stearothermophilus 10	1	0	1	0	1	1	1	1

5.3.6 Specific genomic content relating to growth in milk plates in thermophilic sporeformers

In order to find other thermophile specific genomic features relating to milk spoilage, a gene-trait matching analysis (GTM) was conducted for the relevant phenotypes, including ability to use different sugars, production of active β-galactosidase, and ability to grow on skim milk plates etc. GTM identified 100 genes that specifically help to differentiate strains by the ability to grow on skim milk plates (Table S5.3). The gene annotated as "transporter, major facilitator superfamily (MFS)", predicted to be involved in the transport of niacin (niacin, known as nicotinic acid; together with nicotinamide it composes vitamin B3 complex) weighted the heaviest in contributing to the differentiation of growth from non-growth group indicated by the MDA (mean decrease in accuracy) score. The Random Forest algorithm uses supervised classification for discriminant selection, and "importance" is an arbitrary indicator for the contribution of the discriminant to the classification. This gene appeared present in the strains *G. stearothermophilus* TNO-09.027, TNO-09.008, A-B4114, and *A. flavithermus* TNO-09.006, TNO-09.014, TNO-09.016 that could grow on skim milk plates. However, there is not a complete association of this gene with this phenotype, since several other skim-milk positive strains lacked this gene, i.e., *Geobacillus* sp. C56-T2, *Geobacillus stearothermophilus* 10, and *Geobacillus vulcani* B4164. Interestingly, niacin has been reported as an essential growth factor for lactic and propionic acid bacteria (Sneli et al., 1939). Moreover, strains included in this study in which the niacin transporter was identified were all isolated from diary-processing environments and demonstrated rapid growth on skim milk plates. This observation supports that the niacin transporter might be relevant for the growth of thermophilic sporeformers in milk and may reflect an adaptation to the dairy environment. Further investigation is required to understand this relation.

5.3.7 Concluding remarks

In conclusion, this study demonstrated that all 22 thermophilic sporeformers contain casein breakdown related genes and biofilm formation related genes. A correlation between lactose utilization and growth in skim milk was found. This suggests that the ability to utilize lactose in skim milk is an essential factor for growth in milk. On one

hand, dairy powder-processing facilities can bear in mind this possibility of enhanced growth capability for thermophilic sporeformers, when there is a presence of free monomeric sugars in the ingredients. If stable product specifications are desired, preventive measures to limit the presence of monomeric sugars can be taken into account during the product formulation, raw material selection, or processing. For example, the factory can prevent the addition of monomeric sugars in the product formulation, or the factory could use a mild processing method to prevent sugar break down in the products. On the other hand, if specific lactose-utilization related genes for thermophilic sporeformers can be identified, together with the known lactose-utilization related genes, the information can be used as a biomarker for the prediction of growth potential of thermophilic sporeformers in skim milk. This message might not be directly linked to spoilage control of the thermophilic sporeformer in dairy processing environments, however, the knowledge which is acquired in this study has shown the basis over what it is to be found through the comparative genomic approach for the thermophilic sporeformers.

5.4 Experimental procedures

Whole genome comparison

Previously sequenced genomes representing *Anoxybacillus* spp. (n = 4), *Geobacillus* spp. (n = 16) and *Caldibacillus* spp. (n = 2) were retrieved from NCBI (Table 5.3) and used for comparative analyses. All genomes were (re)annotated using RAST (Aziz et al., 2008) to allow a better comparison of all annotations. Orthologous groups (OGs;) in the genomes were determined using OrthoMCL analysis, the settings used in the analysis are default settings described in the paper from Li et al. (2003) (e-value threshold used in the protein blast is 0.01). When OG (orthologous groups) contained more than one gene per strain (i.e. highly similar genes), an effort was made to manually split OG into separate OGs containing 1 gene per strain (except for transposases or mobile elements). The analysis resulted in 16.455 OGs over all the genomes. When an OG contained more than one gene per strain (i.e. highly similar genes), an effort was made to manually split this OG into separate OGs containing only one gene per strain (except for transposases/mobile elements). Phylogenetic trees, created by MEGA5 (http://www.megasoftware.net/), were used to help indicate

how the OGs with questionable alignments should be split into subgroups. Moreover, for every OG, a multiple sequence alignment was made on the amino acid level using MUSCLE (Edgar, 2004) to facilitate identification of pseudogenes (encoding incomplete proteins).

The protease, peptidase, and carbohydrate utilization systems of *B. subtilis* 168 listed in the Subtiwiki database (Florez et al., 2009) were initially used to search for orthologous systems in the selected thermophilic sporeformers OG table. Additional genes and systems were found with keyword searches. Besides Subtiwiki database (Florez et al., 2009), biofilm formation related genes are selected according to Vlamakis et al. (2013). LocateP (http://www.cmbi.ru.nl/locatep-db/cgi-bin/locatepdb.py/) was used to determine the subcellular locations of *B. subtilis* protease and carbohydrate utilization systems.

Subsequently, selected genes and proteins were manually curated by comparison with the reference *Geobacillus* genomes using databases (e.g. NCBI-BLASTP http://blast.ncbi.nlm.nih.gov/), family and domain databases (e.g. Interpro (Mitchell et al., 2015)), enzyme databases (e.g. Brenda (Schomburg et al., 2004) and pathways databases (e.g. KEGG (Kanehisa and Goto, 2000)). Details of the protease system cassette, biofilm formation related genes in the OG database can be found in Table S5.1, Table S5.2.

Table 5.3 Name and genome description of strains used in this study.

Strain	*Species*[1]	*Environment/ isolation source*	*Sequence statues*	*NCBI Acc #*	*Replicons / contigs*	*Size (Mb)*	*G+C %*[2]
C56-T2	*Geobacillus* sp.	Double Hot Springs, Nevada	complete + plasmids	SAMN0017395	3	3.55	52.39
G11MC16	*Geobacillus* sp.	No record	scaffolds	ABVH00000000	31	3.55	48.8
Y4.1MC1	*Geobacillus thermo-glucosidans*	Bath Hot spring, YNP, USA	complete + plasmids	NC_014650.1	2	3.84	44.02

C56YS93	*Geobacillus thermo-glucosidans*	Obsidian Hot Spring, YNP, USA	complete + plasmids	NC_015660.1	3	4	43.93
CCB_US3_UF5	*Geobacillus thermo-leovorans*	Ulu Slim hot spring, Malaysia.	complete	NC_016593.1	1	3.6	52.28
WCH70	*Geobacillus* sp.	Middleton, USA	complete	NC_012793.1	3	3.51	42.8
B4119	*Geobacillus caldo-xylosilyticus*	Dairy	scaffolds	LQYS00000000.1	122	3.93	44
B4135	*Caldibacillus debilis*	Dairy	scaffolds	LQYT00000000.1	129	3.22	50.9
A_B4114	*Geobacillus stearothermo-philus*	Buttermilk powder, Netherlands	scaffolds	LQYY00000000.1	134	2.74	52.9
T14_B4109	*Geobacillus stearothermo-philus*	Pea soup, Netherlands	scaffolds	LQYV00000000.1	146	2.76	52.5
T27_S_Oomes_B4110	*Geobacillus toebii*	Pea soup, Netherlands	scaffolds	LQYW00000000.1	170	3.5	42.1
B4164	*Geobacillus vulcani*	No record	scaffolds	unpublished data	31	4.09	46.2
TNO-09.014	*Anoxybacillus flavithermus*	Standard milk used for dairy processing plant, Netherlands	scaffolds	LUFB00000000	180	2.56	41.8
TNO-09.016	*Anoxybacillus flavithermus*	Evaporator, Netherlands	scaffolds	LUCQ00000000	187	2.65	41.1
TNO-09.023	*Geobacillus thermo-glucosidans*	Casein pipe fouling in a diary processing plant, Netherlands	scaffolds	LUCT00000000	52	3.7	43.8
TNO-09.027	*Geobacillus stearothermo-philus*	Casein pipe fouling in a diary processing plant, Netherlands	scaffolds	LUCR00000000	226	2.7	52.5

TNO-09.006	*Anoxybacillus flavithermus*	Standard milk used for dairy processing plant, Netherlands	scaffolds	AMCM00000000.	68	2.65	42
TNO-09.008	*Geobacillus stearothermo-philus*	Dairy powder end product, Netherlands	scaffolds	unpublished data	43	2.93	52.4
TNO-09.020	*Geobacillus thermo-glucosidans*	Casein pipe fouling Diary processing plant, Netherlands	scaffolds	NZ_CM001483	1	3.74	43.82
DSM16016	*Caldibacillus debilis*	Sugar beet juice from extraction installations, Austria	scaffolds	GCA_000383875.1	40	3.09	51.6
WK1	*Anoxybacillus flavithermus*	Hot spring, New Zealand,	complete	GCA_000019045.1	1	2.85	41.8
10	*Geobacillus stearothermo-philus*	Hot springs, YNP, USA,	complete + plasmids	GCA_001274575.1	120	3.59	52.61

[1] Putative species are confirmed using phylogenetic analysis of presence or absence of orthologue genes from the 22 thermophilic strains (Figure 5.1).
[2] Inferred from genome sequence

Phylogenetic analysis

The OrthoMCL analysis of all the genes of the 22 strains, together with manual curation, resulted in a matrix of 22 strains (columns) 16,224 Orthologous Groups (OGs, rows). An OG is defined by OrthoMCL as a set of orthologous genes of different strains with a high degree of protein homology and possibly the same or similar function. In this matrix, for each of the OGs, the absence/presence of a gene-orthologue for each strain (identified by OrthoMCL) was marked by 0.1/1[4]. This matrix was used to perform hierarchical clustering (Pearson correlation, average Linkage) the with MeV software (http://www.tm4.org/mev.html/) resulting in a

4. note that 0.1 was used instead of 0 since the subsequent MeV software can't process 0 properly.

heatplot with dendrograms for strains and/or OGs based on Pearson correlations among either strains or OGs.

Bacterial strains and growth media

The strains (Table 5.3) were obtained from the -80 °C stock. Tryptone Soy Broth (TSB), Tryptone Soy Agar (TSA plates, Tritium Microbiologie, the Netherlands) and Tryptone (Tritium Microbiologie, the Netherlands) were used as basic growth medium. UHT milk Friesche Vlag® Langlekker 0% fat was used for the growth experiments in milk. Demi water (Tritium Microbiologie) was used for dilution of media in growth assays.

Milk plate counting agar (MPCA) plates for testing protease activity

2.5g Tryptone, 1.25g yeast extract, 0.5g glucose, and 15g agar were dissolved in 450 ml demi water and autoclaved for 20 minutes at 120°C. Separately, 0.5g skim milk powder was dissolved in 50 ml demi water and was autoclaved for 5 minutes at 120°C. The two above-prepared contents were poured together in a 47°C water bath. MPCA Plates were made and stored in the cold room (-4°C). Strains were inoculated on the MPCA plates and incubated for 24h at 55°C. Grown strains were observed for halo formation on MPCA plates.

Plate Assay for testing β-galactosidase activity

200 mg X-gal was dissolved in 10 ml DMSO. TSA was melted and cooled down to lower than 65°C. 500 µl X-gal stock solution was added into the TSA (final concentration 40 µg/ml). Plates were poured, dried, and stored in the cold room (-4°C) and in the dark (covered with aluminium foil) for following inoculation. Overnight grown strains were streaked on the previous made plates and incubated for 24h at 55°C. The strains which showed blue colonies are strains which produced active β-galactosidase.

API growth test

API50CH (BioMerieux, France) test was used in combination with API 50CHB/E medium (BioMerieux, France) according to the manufacturer's instruction. Single colonies were used to inoculate API 50CHB/E medium to a turbidity equivalent of 2McFarland. Test strips were filled with inoculated medium, incubated at 55°C, and

checked for media colour change after 24 and 48h. For each strain, two independent repetitions were performed.

Gene trait matching (GTM)

GTM was performed using Phenolink, a web tool that associates bacterial phenotypes with genomic data (Bayjanov et al., 2012), in order to correlate observed phenotypes with the presence or absence of particular genes. For example, in relation to spoilage in the dairy processing line, phenotypic data were divided into strains that could grow on skim milk plates versus strains that could not grow on skim milk plates. Growth (Yes) or non-growth (No) of 22 thermophilic sporeformers strains on skim milk plates was measured. A gene that is found to be important to distinguish strains of different phenotype is assumed as important. Selected genes are ranked based on the mean decrease in accuracy (MDA). The higher the number is, the more relevant statistically a gene is in differentiating the grower or non-grower on skim milk plates.

In vitro biofilm assays for screening biofilm formation ability

Biofilm formation on polystyrene was measured as described previously (Wijman et al., 2007), with some modifications. In short, polystyrene microtiter plates (Greiner Bio-one, Germany) were filled with 200 µl TSB, and for each strain, six wells were inoculated with 1.5% (vol/vol) suspended overnight cultures, grown on TSA plates, respectively. The microtiter plates were incubated at 55 and 65°C. After incubation, for either 16 or 25 h, the wells were gently washed three times with 220 µl of sterile distilled water, and subsequently biofilm cells were stained with 220 µl of 0.1% (wt/vol) crystal violet for 30 min. After this 30 min, the wells were washed twice with 220 µl sterile deionized water to remove unbound crystal violet. The remaining crystal violet was dissolved in 220 µl 96% ethanol, and the absorbance was measured at 595 nm (Quant, Bio-Tek Instruments). Strains with optical density at 595 nm (OD595) values above two times the background signal, or higher were considered positive for biofilm formation. The average biofilm formation and standard deviation for six wells were calculated for each strain in the conditions tested. The turbidity of medium measured by spectrophotometer at OD600 is used as an indication of growth.

5.5 Acknowledgments

The project is funded by TI Food and Nutrition, a public-private partnership on pre-competitive research in food and nutrition. We wish to thank Lars Wims for technical assistance, Alicja Warda from the food microbiology group at Wageningen University and other members of TNO microbiology and system biology group for stimulating discussions on this topic.

5.6 References

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vontein, A. Wilke and O. Zagnitko. 2008. The RAST server: Rapid annotations using subsystems technology. BMC Genomics. 9:75.

Bachmann, H., D. Molenaar, M. Kleerebezem and J. E. van Hylckama Vlieg. 2011. High local substrate availability stabilizes a cooperative trait. The ISME journal. 5:929.

Barbieri, G., A. M. Albertini, E. Ferrari, A. L. Sonenshein, B. R. Belitsky. 2016. Interplay of CodY and ScoC in the regulation of major extracellular protease genes of *Bacillus subtilis*. Journal of Bacteriology. 1986:907-920.

Bayjanov, J. R., D. Molenaar, V. Tzeneva, R. J. Siezen and S. A. van Hijum. 2012. PhenoLink a web-tool for linking phenotype to ~omics data for bacteria: Application to gene-trait matching for *Lactobacillus plantarum* strains. BMC Genomics. 13:170.

Berendsen, E. M., M. H. Wells-Bennik, A. O. Krawczyk, A. de Jong, A. van Heel, S. Holsappel, R. T. Eijlander and O. P. Kuipers. 2016. Draft genome sequences of seven thermophilic spore-forming bacteria isolated from foods that produce highly heat-resistant spores, comprising *Geobacillus* spp., *Caldibacillus debilis*, and *Anoxybacillus flavithermus*. Genome Announcement. 4:e00105-16.

Branda, S. S., J. E. Gonzalez-Pastor, S. Ben-Yehuda, R. Losick and R. Kolter. 2001. Fruiting body formation by *Bacillus subtilis*. Proceedings of the National Academy of Sciences. 98:11621-11626.

Brumm, P. J., P. De Maayer, D. A. Mead and D. A. Cowan. 2015. Genomic analysis of six new *Geobacillus* strains reveals highly conserved carbohydrate degradation architectures a nd strategies. Frontiers in Microbiology. 6:430.

Burgess, S. A., S. H. Flint, D. Lindsay, M. P. Cox and P. J. Biggs. 2017. Insights into the *Geobacillus stearothermophilus* species based on phylogenomic principles. BMC Microbiology. 17:140.

Carlin, F. 2011. Origin of bacterial spores contaminating foods. Food Microbiology. 28:177-182.

Caspers, M. P. M., J. Boekhorst, A. de Jong, R. Kort, M. N. Nierop Groot and T. Abee. 2016. Draft genome sequences of four thermophilic spore formers isolated from a dairy-processing environment. Genome Announcement. 4: e00757-16.

Ceapa, C., J. Lambert, K. van Limpt, M. Wels, T. Smokvina, J. Knol and M. Kleerebezem. 2015. Correlation of *Lactobacillus rhamnosus* genotypes and carbohydrate utilization signatures determined by phenotype profiling. Applied and Environmental Microbiology. 81:5458-5470.

Donkor, O. N., A. Henriksson, T. Vasiljevic and N. P. Shah. 2007. Proteolytic activity of dairy lactic acid bacteria and probiotics as determinant of growth and in vitro angiotensin converting enzyme inhibitory activity in fermented milk. Le Lait, INRA Editions. 87:21-38.

Edgar, R. C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:113.

Exterkate, F. A. and G. J. C. M. de Veer. 1987. Optimal growth of *Streptococcus cremoris* HP in milk is related to β-and x-casein degradation. Applied Microbiology and Biotechnology. 25:471-475.

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

Flint, S., J. Palmer, K. Bloemen, J. Brooks and R. Crawford. 2001. The growth of *Bacillus stearothermophilus* on stainless steel. Journal of Applied Microbiology. 90:151-157.

Florez, L. A., S. F. Roppel, A. G. Schmeisky, C. R. Lammers and J. Stulke. 2009. A community-curated consensual annotation that is continuously updated: The *Bacillus subtilis* centered wiki SubtiWiki. Database (Oxford). 2009:bap012.

Goh, K. M., H. M. Gan, K. Chan, G. F. Chan, S. Shahar, C. S. Chong, U. M. Kahar and K. P. Chai. 2014. Analysis of *Anoxybacillus* genomes from the aspects of lifestyle adaptations, prophage diversity, and carbohydrate metabolism. PLoS One. 9: e90549.

Griffiths, A., W. Gelbart and J. Miller. 1999. Regulation of the lactose system. Modern Genetic Analysis. W. H. Freeman, US, New York.

Hinton, A. R., K. T. Trinh, J. D. Brooks and G. J. Manderson. 2002. Thermophile survival in milk fouling and on stainless steel during cleaning. Food and Bio-products Processing Publication of the Institution of Chemical Engineers Part C. 80:299-304.

Kanehisa, M. and S. Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 28:27-30.

Klaver, F. A., F. Kingma and A. H. Weerkamp. 1993. Growth and survival of *Bifidobacteria* in milk. Nederlands Melk en Zuiveltijdschrift. 47:151-164.

Li, L., C. J. Stoeckert and D. S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research. 13:2178-2189.

Mekalanos, J. J. 1992. Environmental signals controlling expression of virulence determinants in bacteria. Journal of Bacteriology. 174:1-7.

Michna, R. H., F. M. Commichau, D. Todter, C. P. Zschiedrich and J. Stulke. 2014. SubtiWiki-a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information. Nucleic Acids Resources. 42:D692-698.

Mills, O. E., and T. D. Thomas. 1981. Nitrogen sources for growth of lactic streptococci in milk. New Zealand Journal of Dairy Science and Technology. 16:43-45.

Mitchell, A., H. Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, M. Scheremetjew, C. Rato, S. Y. Yong, A. Bateman, M. Punta, T. K. Attwood, C. J. Sigrist, N. Redaschi, C. Rivoire, I. Xenarios, D. Kahn, D. Guyot, P. Bork, I. Letunic, J. Gough, M. Oates, D. Haft, H. Huang, D. A. Natale, C. H. Wu, C. Orengo, I. Sillitoe, H. Mi, P. D. Thomas and R. D. Finn. 2015. The InterPro protein families database: The classification resource after 15 years. Nucleic Acids Resources. 43:D213-221.

Ramírez, M. D. F., E. J. Smid, T. Abee and M. N. Nierop Groot. 2015. Characterisation of biofilms formed by *Lactobacillus plantarum* WCFS1 and food spoilage isolates. International Journal of Food Microbiology. 207:23-29.

Rohmer, L., D. Hocquet and S. Miller. 2011. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends in Microbiology. 19:341-348.

Sadiq, F. A., S. Flint, L. Yuan, Y. Li, T. Liu and G. He. 2017. Propensity for biofilm formation by aerobic mesophilic and thermophilic spore forming bacteria isolated from Chinese milk powders. International Journal of Food Microbiology. 262:89-98.

Schomburg, I., A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn and D. Schomburg. 2004. BRENDA, the enzyme database: Updates and major new developments. Nucleic Acids Resources. 32:D431- 433.

Scott, S. A., J. D. Brooks, J. Rakonjac, K. M. Walker and S. H. Flint. 2007. The formation of thermophilic spores during the manufacture of whole milk powder. International Journal of Dairy Technology. 60:109-117.

Siezen, R. J. and J. E. van Hylckama Vlieg. 2011. Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer. In Microbial cell factories (Vol.10, No. 1, p. S3). BioMed Central.

Sims, G. K. 2006. "Nitrogen starvation promotes biodegradation of N-Heterocyclic compounds in soil". Soil Biology & Biochemistry. 38:2478–2480.

Sims, G. K and M. M. Wander. 2002. "Proteolytic activity under nitrogen or sulfur limitation". Applied Soil Ecology. 568:1–5.

Solopova, A., H. Bachmann, B. Teusink, J. Kok, A. R. Neves, O. P. Kuipers. 2012. A specific mutation in the promoter region of the silent *cel* cluster accounts for the appearance of lactose-utilizing *Lactococcus lactis* MG1363. Applied and Environmental Microbiology. 78:5612-5621.

Switt, A. I. M., A. D. Andrus, M. L. Ranieri, R. H. Orsi, R. Ivy, H. C. den Bakker, N. H. Martin, M. Wiedmann and K. J. Boor. 2014. Genomic comparison of sporeforming bacilli isolated from milk. BMC Genomics. 15:26.

Thompson, J. 1987. Regulation of sugar transport and metabolism in lactic acid bacteria. FEMS Microbiology Reviews. 3:221-231.

Vlamakis, H., Y. Chai, P. Beauregard, R. Losick, and R. Kolter. 2013. Sticking together: building a biofilm the *Bacillus subtilis* way. Nature Reviews Microbiology. 11:157-168.

Warda, A. K., R. J. Siezen, J. Boekhorst, M. H. J. Wells-Bennik, A. de Jong, O. P. Kuipers, M. N. Nierop Groot and T. Abee. 2006. Linking *Bacillus cereus* genotypes and carbohydrate utilization capacity. PloS One. 11:e156796.

Wassarman, K. M. 2002. Small RNAs in bacteria: Diverse regulators of gene expression in response to environmental changes. Cell. 109:141-144.

Watterson, M. J., D. J. Kent, K. J. Boor, M. Wiedmann and N. H. Martin. 2014. Evaluation of dairy powder products implicates thermophilic sporeformers as the primary organisms of interest. Journal of Dairy Science. 97:2487-2497.

Wijman, J. G. E., P. P. L. A. de Leeuw, R. Moezelaar, M. H. Zwietering and T. Abee. 2007. Air-liquid interface biofilms of *Bacillus cereus*: Formation, sporulation, and dispersion. Applied and Environmental Microbiology. 73:1481-1488.

Zhao, Y., M. P. M. Caspers, K. I. Metselaar, P. de Boer, G. Roeselers, R. Moezelaar, M. N. Nierop Groot, R. C. Montijn, T. Abee and R. Kort. 2013. Abiotic and microbiotic factors controlling biofilm formation by thermophilic sporeformers. Applied and Environmental Microbiology. 79:5652-5660

Zhao, Y., M. Kumar, M. P. M. Caspers, M. N. Nierop Groot, J. M. B. M. van der Vossen, and T. Abee. 2017. Short communication: growth of dairy isolates of *Geobacillus thermoglucosidans* in skim milk depends on lactose degradation products supplied by *Anoxybacillus flavithermus* as secondary species. Journal of Dairy Science. 17:31071 – 31078

5.7 Supplemental figures and tables

5.7.1 Supplemental figures

Row	Gene/Protein	Genes	*Bacillus subtilis* 168	*Anoxybacillus flavithermus* TNO.09-006	*Anoxybacillus flavithermus* TNO.09-014	*Anoxybacillus flavithermus* TNO.09-016	*Anoxybacillus flavithermus* WK1	*Geobacillus thermoglucosidans* TNO.09-020	*Geobacillus thermoglucosidans* TNO.09-023	*Geobacillus thermoglucosidasius* C56_Y93	*Geobacillus stearothermophilus* TNO.09-008	*Geobacillus stearothermophilus* TNO.09-027	*Geobacillus stearothermophilus* 10	*Geobacillus stearothermophilus* A_B4114	Geobacillus_stearothermophilus_T14_B4109	*Geobacillus caldoxylosilyticus* B4119	*Geobacillus debilis* B4135	*Geobacillus debilis* DSM_16016	*Geobacillus thermoleovorans* CCB_US3_UF5	*Geobacillus toebii* T27_S_Oomes_B4110	*Geobacillus* sp. G11MC16	*Geobacillus* sp C56-T2	*Geobacillus* sp. WCH70	*Geobacillus* sp. Y4_1MC1	*Geobacillus vulcani* B4164
1	*YwcC*	BSU38220	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	*SlrA*	BSU38229	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3	sigM	BSU09520	1	1	1	1	0	1	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	1
4	sigW	BSU01730	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	sigX	BSU23100	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	Abh	BSU14480	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
7	SlrR	BSU34380	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8	SinR	BSU24610	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	SinI	BSU24600	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10	Spo0A	BSU24220	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	AbrB	BSU00370	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	KinA	BSU13990	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	KinB	BSU31450	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	KinC	BSU14490	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1
15	KinD	BSU13660	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	DegS	BSU35500	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	DegU	BSU35490	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	*epsA*	BSU34370	2	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	2
19	*epsB*	BSU34360	2	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1
20	*epsC*	BSU34350	1	1	1	0	0	1	2	0	0	0	1	0	0	0	1	1	1	0	0	1	0	1	1	2
21	*epsD*	BSU34340	1	1	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	2	0	0	0	1	0	1
22	*epsE*	BSU34330	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
23	*epsF*	BSU34320	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1
24	*epsG*	BSU34310	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
25	*epsH*	BSU34300	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1
26	*epsI*	BSU34290	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
27	*epsJ*	BSU34280	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
28	*epsK*	BSU34265	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
29	*epsL*	BSU34250	1	2	1	1	0	1	1	1	1	2	0	1	1	1	1	1	1	1	1	0	1	1	0	2
30	*epsM*	BSU34240	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
31	*epsN*	BSU34230	1	2	0	1	0	0	1	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1
32	*epsO*	BSU34220	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
33	tapA	BSU24640	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
34	*sipW*	BSU24630	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
35	*tasA*	BSU24620	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
36	bslA	BSU31080	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
37	*capB*	BSU35900	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
38	*capC*	BSU35890	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
39	*pgsA*	BSU16920	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	*capE*	BSU35870	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
41	remB	BSU00050	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
42	yaaO	BSU00270	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
43	veg	BSU00440	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
44	ftsH	BSU00690	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
47	gabT	BSU03900	1	0	0	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1
48	yisP	BSU10810	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	ampS	BSU14450	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
50	speA	BSU14630	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51	ylbF	BSU14990	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
52	ylmE	BSU15380	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
53	remA	BSU15670	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
54	rny	BSU16960	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
55	ymdB	BSU16970	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
56	ymcA	BSU17020	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
57	luxS	BSU30670	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
58	floT	BSU31010	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	0
59	degQ	BSU31720	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
60	racX	BSU34430	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	yvcA	BSU34850	1	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0
64	ptpZ	BSU36240	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1
65	ptkA	BSU36250	2	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	1
66	tkmA	BSU36260	2	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	2
67	galE	BSU38860	1	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0
68	yxaB	BSU40030	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	sfp/1	BSU03569																								
46	sfp/2	BSU03570																								
62	swrAA/1	BSU35230																								
63	swrAA/2	BSU35239																								

Figure S5.1 Presence and absence of genes involved in biofilm formation (*Bacillus subtilis* was used as a reference for selecting genes involved in biofilm formation). The first line from left to right are the name of the

Bacillus subtilis 164 and the 22 collected thermophilic sporeformers. To the left of the black square listed the selected genes involved in biofilm formation. Within the black square, "red square" indicates presence, and "white" indicates absence.

5.7.2 Supplemental tables

Table S5.1 Percentage identity of the lactose utilization related genes from each strain in comparison to the orthologue genes identified through OrthoMCL.

Percentage identity	*A permease protein/ lactose ABC transporter*	*β-galactosidase (EC 3.2.1.23)*	*a lactose binding protein/ lactose ABC transporter*
Anoxybacillus flavithermus TNO-09.006	100%	100%	100%
Anoxybacillus flavithermus TNO-09.016	100%	99%	100%
Geobacillus caldoxylosilyticus B4119	13%	6%	6%
Geobacillus stearothermophilus 10	97%	99%	90%

Table S5.2 Genomic features in relation to proteolysis in the 22 selected thermophilic sporeformer. (*Bacillus subtilis* was used as a reference for selecting genomic features involved in proteolytic activity). The first line from left to right are the name of 22 collected thermophilic sporeformers and *Bacillus subtilis* 168. "1" indicates presence, "0" indicates absence, "P" possible presence.

OG	*Anoxybacillus flavithermus* TNO-09.006	*Anoxybacillus flavithermus* TNO-09.014	*Anoxybacillus flavithermus* TNO-09.016	*Anoxybacillus flavithermus* WK1	*Geobacillus caldoxylosilyticus* B4119	*Geobacillus debilis* B4135	*Geobacillus debilis* DSM16016	*Geobacillus* sp C56-T2	*Geobacillus* sp G11MC16	*Geobacillus stearothermophilus* 10	*Geobacillus stearothermophilus* A_B4114	*Geobacillus stearothermophilus* T14_B4109	*Geobacillus stearothermophilus* TNO-09.008	*Geobacillus stearothermophilus* TNO-09.027	*Geobacillus thermoglucosidans* TNO-09.020	*Geobacillus thermoglucosidans* TNO-09.023	*Geobacillus thermoglucosidasius* C56YS93	*Geobacillus thermoglucosidans* Y4.1MC1	*Geobacillus thermoleovorans* CCB_US3_UF5	*Geobacillus toebii* T27_S_Oomes_B4110	*Geobacillus vulcani* B4164	*Geobacillus* sp. WCH70	*Bacillus subtilis* 168	improved annotation
OG_199	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU00430	Sporulation-specific protease YabG
OG_4365a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU02240	extracellular glutamyl endopeptidase
OG_311	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU06010	CAAX amino terminal protease, self-immunity protein
OG_3271	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	p	BSU10290	membrane-bound peptidase HtpX
OG_2682b	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	BSU10300	serine alkaline protease (subtilisin E)
OG_6727	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU10770	cell wall-associated subtilisin-like serine protease
OG_2833	0	0	0	0	1	0	0	1	0	1	p	p	p	p	0	0	0	0	1	0	0	0	BSU11100	extracellular neutral protease B
OG_1652a	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	BSU12900	Serine protease, DegP/HtrA, do-like

OG_2939a	1	1	1	0	1	0	0	0	1	0	0	0	0	0	1	1	1	1	0	0	1	0	BSU13190	Major intracellular serine protease precursor
OG_9928	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BSU14700	extracellular neutral metalloprotease
OG_457	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU15310	Sporulation sigma-E factor processing peptidase (SpoIIGA)
OG_511	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU16150	ATP-dependent protease HslV
OG_548	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU16560	membrane-associated zinc metalloprotease, peptidase
OG_561	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU16710	Zn-dependent peptidase
OG_570	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU16845	Zn-dependent peptidase
OG_571	1	1	1	1	1	1	1	1	1	1	1	1	p	1	1	1	1	1	1	1	1	1	BSU16860	Zn-dependent peptidase
OG_2456	0	0	0	0	1	0	0	1	1	1	0	0	0	0	1	1	1	1	1	0	1	0	BSU17260	subtilisin-like serine protease
OG_5457	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU19590	carboxy-terminal processing protease, peptidase S41 family
OG_676	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU22940	protease required for RsiW anti-sigma (W) degradation
OG_679	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU23010	CAAX amino terminal protease, self-immunity protein
OG_717	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU24230	Stage IV sporulation protein B
OG_798	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU25540	Endopeptidase spore protease Gpr
OG_854	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU27970	Stage IV sporulation pro-sigma-K processing enzyme (SpoIVFB)
OG_1247	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	BSU28070	Prepilin peptidase, membrane-bound
OG_873	1	1	1	1	1	1	1	1	1	1	1	1	p	1	1	1	1	1	1	1	1	1	BSU28210	ATP-dependent protease La, LonB Type I
OG_53a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU31310	membrane neutral zinc metallopeptidase
OG_1652b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU33000	HtrA-like serine protease
OG_1012	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU34540	ATP-dependent Clp protease proteolytic subunit
OG_1027	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU35240	ATP-dependent Clp protease proteolytic subunit
OG_4929a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU38090	Minor extracellular protease vpr
OG_153c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU38400	minor extracellular protease epr
OG_3174	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	0	0	0	BSU38780	membrane-bound peptidase HtpX
OG_9741	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BSU40160	membrane metalloprotease
OG_1112	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU40360	Serine protease, DegP/HtrA, do-like
OG_183	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU00100	D-alanyl-D-alanine carboxypeptidase
OG_221	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU00690	ATP-dependent metalloprotease FtsH
OG_9723	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BSU01140	proline iminopeptidase
OG_1298	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU01380	Methionine aminopeptidase
OG_2171b	0	0	0	0	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	1	0	BSU02650	Pyrrolidone-carboxylate peptidase
OG_3028	0	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1	0	1	BSU02810	D-alanyl-D-alanine carboxypeptidase VanY
OG_9612	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BSU04010	type I signal peptidase

OG_304	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU04790	Zn-dependent metalloprotease, SprT family
OG_1792	0	0	0	0	1	1	1	1	p	1	1	1	1	1	1	1	1	1	1	1	1	BSU06350	cysteine proteinase, membrane-bound
OG_6479	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU07690	methionine aminopeptidase
OG_4857	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU07850	intracellular cysteine peptidase (Pfp1 endopeptidase)
OG_4173	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	p	0	p	BSU08780	Peptidase E
OG_43a	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	0	BSU09200	sortase A
OG_10017	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BSU09370	gamma-D-glutamate-meso-diaminopimelate muropeptidase
OG_4866	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU10200	glutamyl aminopeptidase
OG_1969	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	p	1	1	BSU10490	Signal peptidase I
OG_1187	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU11350	Zn-dependent protease,DUF2268 family
OG_380	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU11540	Oligoendopeptidase F
OG_3114	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	0	BSU12920	zinc-dependent, D-specific aminopeptidase DppA
OG_6598	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU12970	LD-carboxypeptidase; Muramoyl-tetrapeptide carboxypeptidase
OG_3292	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	p	1	BSU12990	cell wall endopeptidase, NLP/P60 family
OG_5448a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU13490	membrane-bound peptidase HtpX
OG_1632	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU13860	Proline dipeptidase
OG_3213a	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1	0	0	1	0	BSU14410	Signal peptidase I
OG_628	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU14450	Aminopeptidase S (Leu, Val, Phe, Tyr preference)
OG_437	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU15050	Lon-like protease with PDZ domain
OG_4188a	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU15300	bacillopeptidase F
OG_466	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU15450	Lipoprotein signal peptidase
OG_567	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU16790	ATP-dependent Clp protease, protease subunit
OG_6390	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU16950	penicillin-binding endopeptidase X
OG_590	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU17850	SOS-response repressor and protease LexA
OG_6662	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU18350	D-alanyl-D-alanine carboxypeptidase
OG_6585	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU18410	membrane bound gamma-glutamyltranspeptidase
OG_2689a	0	0	0	0	0	p	p	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	BSU19170	LD-carboxypeptidase ;Muramoyl-tetrapeptide carboxypeptidase
OG_6695	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU19410	cell wall endopeptidase, NLP/P60 family, with LysM domains
OG_613	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU19620	D-alanyl-D-alanine carboxypeptidase
OG_608	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU22080	Thermostable carboxypeptidase 1

OG_688	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU23190	D-alanyl-D-alanine carboxypeptidase
OG_1196	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	BSU23310	Signal peptidase I
OG_699	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU23480	D-alanyl-D-alanine carboxypeptidase
OG_710	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU23910	Peptidase T
OG_737	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU24460	Xaa-Pro aminopeptidase, peptidase M24 family
OG_6752	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU24630	type I signal peptidase
OG_755	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU24830	Gamma-D-glutamyl-L-diamino acid endopeptidase I
OG_757	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU24870	membrane associated protease, rhomboid family
OG_3406b	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU25390	Membrane-bound serine protease (ClpP class)
OG_1211a	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU27020	intracellular cysteine peptidase
OG_1242	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU27340	peptidase, U32 family large subunit [C1]
OG_1243	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU27350	peptidase, U32 family small subunit [C1]
OG_855	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU27980	endopeptidase
OG_872	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU28200	ATP-dependent protease La, Type I
OG_1252	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU28820	glutamyl aminopeptidase PepA
OG_1865	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU29530	Signal peptide peptidase A (SppA)
OG_946	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	p	1	1	1	1	1	1	BSU29860	Glutamyl aminopeptidase
OG_1521	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU29980	Xaa-His dipeptidase (PepV)
OG_2022	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU30580	Dipeptidyl aminopeptidase
OG_2099	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	p	1	1	1	BSU32050	aminopeptidase PepA
OG_1670a	0	0	0	0	1	1	1	1	1	1	p	1	1	p	1	1	1	1	p	0	1	1	BSU32230	Dipeptidyl aminopeptidase
OG_989	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU32340	sporulation-specific L-Ala-D-Glu endopeptidase
OG_6405	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU34800	cell wall DL-endopeptidase, NPLC/P60 family
OG_1878	0	0	0	0	1	1	1	1	1	1	1	p	1	p	1	1	1	1	1	1	1	1	BSU36100	Gamma-glutamyltranspeptidase
OG_1061	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU36550	endopeptidase
OG_9559	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BSU37420	subtilosin production peptidase
OG_1099	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU37530	membrane-associated zinc metalloprotease
OG_6952	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	BSU38470	metallopeptidase
OG_1828	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU38920	Tripeptide aminopeptidase PepT
OG_1125	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	BSU40950	sporulation protein YyaC (spore protease Gpr family)
OG_4188b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	bacillopeptidase F
OG_153a	0	0	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	extracellular subtilisin-like serine protease
OG_153b	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	0	extracellular subtilisin-like serine protease

OG_4929b	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	extracellular subtilisin-like serine protease, peptidoglycan anchored
OG_53b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	1	0	membrane neutral zinc metallopeptidase
OG_4365b	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	extracellular glutamyl endopeptidase
OG_2171a	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	Pyrrolidone-carboxylate peptidase
OG_2689b	0	0	0	0	0	0	0	0	0	p	p	p	p	p	0	0	0	0	p	0	0	0	0	0	LD-carboxypeptidase ;Muramoyl-tetrapeptide carboxypeptidase
OG_3213b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Signal peptidase I
OG_2978	0	0	0	0	0	0	0	1	0	1	0	p	1	1	0	0	0	0	p	0	0	1	0	0	L/D-Ala aminopeptidase DmpA
OG_2682a	0	1	p	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	alkaline serine protease, with S-layer homology domains
OG_583a	1	0	0	0	1	0	0	1	0	1	p	p	1	1	1	1	1	1	1	1	0	1	0	0	cell wall-associated hydrolase (endopeptidase),NLPC/P60 family, with N-terminal LysM domains
OG_3870	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	p	0	0	membrane-associated zinc metalloprotease
OG_1646a	0	1	p	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	cell wall-associated hydrolase (endopeptidase),NLPC/P60 family
OG_1977	0	0	0	0	1	1	1	1	p	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	Oligopeptidase F (PepF)
OG_1209	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	Membrane dipeptidase
OG_2006	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	Methionine aminopeptidase
OG_2645	0	0	0	0	1	0	0	1	1	1	1	p	1	1	0	0	0	0	1	0	0	p	0	0	metalloendopeptidase InhA
OG_2173	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	Oligopeptidase F (PepF)
OG_2172	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	metallopeptidase
OG_2934	0	1	0	1	p	1	1	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	0	0	tripeptidase
OG_1271	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	endopeptidase
OG_4196a	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	metallopeptidase
OG_4196b	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	metallopeptidase
OG_4196c	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	metallopeptidase
OG_4000	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	peptidase family S1C (protease Do subfamily)
OG_3993a	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	SOS-response repressor and protease LexA
OG_3993b	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	SOS-response repressor and protease LexA
OG_3250a	0	0	1	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	Clp protease (caseinolytic protease; ClpP)
OG_3250b	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Clp protease (caseinolytic protease; ClpP)
OG_4448	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	subtilisin-like serine protease
OG_4411	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	Hydrogenase maturation protease
OG_4522	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	Hydrogenase maturation protease
OG_4523	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	Hydrogenase maturation protease

OG_3662	1	0	p	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0		Hydrogenase maturation protease
OG_3865	0	0	0	0	0	p	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0		membrane-associated zinc metalloprotease
OG_10880	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p	0	0	0	0	0	0		peptidase
OG_10922	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0		Xaa-Pro aminopeptidase
OG_5533	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0		peptidase
OG_5625	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0		CAAX amino terminal protease, self-immunity protein
OG_10447	0	0	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Oligopeptidase F (PepF)
OG_4696a	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		D-alanyl-D-alanine dipeptidase
OG_4696b	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		D-alanyl-D-alanine dipeptidase
OG_5862	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Xaa-Pro aminopeptidase PepP
OG_7818	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Xaa-Pro aminopeptidase PepP
OG_10384	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		serine protease (ClpP class)-like protein
OG_3921a	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0		signal peptidase I
OG_3921b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0		signal peptidase I
OG_7869	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		serine protease, trypsin-like
OG_10565	0	0	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0	0	0		glutamyl aminopeptidase
OG_10445	0	0	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0	0	0		peptidase
OG_5193	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		endopeptidase
OG_3801	0	0	0	1	0	0	0	p	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0		bacteriocin-processing endopeptidase /ABC transporter
OG_4301	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		aminopeptidase
OG_5365	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0		phage prohead protease, HK97 family
OG_5398	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		CAAX amino terminal protease, self-immunity protein
OG_4325	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		peptidase
OG_14144	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		metallopeptidase
OG_14348	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		Lantibiotic specific maturation protease, serine peptidase
OG_9124	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		peptidase
OG_8211	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Gamma-D-glutamyl-meso-diaminopimelate peptidase I
OG_6019	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		peptidase
OG_11196	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		subtilisin-like serine protease
OG_3150	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0		C-terminal processing peptidase
OG_16044	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p	0	0	0	0		prolyl oligopeptidase
OG_8837	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Clp protease subunit 2(caseinolytic protease; ClpP)
OG_7630	0	0	0	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0	0		endopeptidase
OG_8128	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Membrane dipeptidase
OG_8377	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		neutral zinc metallopeptidase, membrane bound
OG_11681	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		zinc metallopeptidase

OG_14554	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		Clp protease subunit 2(caseinolytic protease; ClpP)
OG_5438	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		Dipeptidyl aminopeptidase
OG_5475	0	0	0	0	0	p	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		D-alanyl-D-alanine carboxypeptidase
OG_5476	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		Oligopeptidase F (PepF)
OG_4692a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Gamma-D-glutamyl-meso-diaminopimelate peptidase I
OG_9112	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		extracellular serine protease, with S-layer homology domains
OG_6147a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		membrane-associated zinc metalloprotease
OG_6147b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		membrane-associated zinc metalloprotease
OG_7648	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0		membrane serine protease
OG_16017	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p	0	0	0	0		prolyl oligopeptidase
OG_15459	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p	0	0	0		Serine protease, DegP/HtrA, do-like
OG_9255	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Zn-Peptidase
OG_15045	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		Cell wall endopeptidase, with LysM domains
OG_5697	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0		subtilisin-like serine protease, with FlgD Ig-like domains
OG_8173	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		carboxypeptidase
OG_12053	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		CAAX amino terminal protease, self-immunity protein
OG_5169	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0		Zn-dependent peptidase
OG_8379	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		murein hydrolase activator NlpD
OG_5409	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		Oligopeptidase F (PepF)
OG_13708	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Ubiquitin specific protease
OG_8448	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		sortase C, membrane-bound cysteine transpeptidase
OG_8395	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		Xaa-Pro dipeptidase, Prolidase
OG_7981a	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		subtilisin-like serine protease
OG_7981b	0	0	0	0	0	0	0	0	0	0	0	p	0	0	0	0	0	0	0	0	0	0	0	0		subtilisin-like serine protease
OG_8197	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		CAAX amino terminal protease, self-immunity protein
OG_5448b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0		peptidase
OG_43b	1	1	1	1	1	0	0	0	0	0	p	0	0	0	1	1	1	1	1	1	1	1	0			sortase family protein
Sum	63	65	64	68	92	82	84	82	79	79	72	75	75	76	91	90	91	90	78	81	102	80	0			

Table S5.3 Presence and absence of GTM identified 100 genes that specifically help to differentiate strains by the ability to grow on skim milk plates in the selected thermophilic sporeformers. The number indicates the number of copies found in the genome.

Gene/OG	*importance (MDA)*	*Annotation*	*Anoxybacillus flavithermus WK1*	*Geobacillus caldoxylosilyticus B4119*	*Geobacillus debilis B4135*	*Geobacillus debilis DSM_16016*	*Geobacillus sp G11MC16*	*Geobacillus stearothermophilus T14_B4109*	*Geobacillus thermoglucosidans TNO-09.020*	*Geobacillus thermoglucosidans TNO-09.023*	*Geobacillus thermoglucosidasius C56_YS93*	*Geobacillus thermoleovorans CCB_US3_UF5*	*Geobacillus toebii T27_S_Oomes_B4110*	*Geobacillus* sp. *WCH70*	*Geobacillus thermoglucosidans Y4.1MC1*	*Anoxybacillus flavithermus TNO-09.006*	*Anoxybacillus flavithermus TNO-09.014*	*Anoxybacillus flavithermus TNO-09.016*	*Geobacillus* sp. *C56-T2*	*Geobacillus stearothermophilus 10*	*Geobacillus stearothermophilus A_B4114*	*Geobacillus stearothermophilus TNO-09.008*	*Geobacillus stearothermophilus TNO-09.027*	*Geobacillus vulcani B4164*
			no(#13)													yes (#9)								
OG_2621	0.009985	Outer surface protein of unknown function, cellobiose operon	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
OG_3740	0.008929	transporter, major facilitator superfamily	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0
OG_2238	0.005417	Xylose ABC transporter, substrate-binding component	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0
OG_2234	0.005417	Butyryl-CoA dehydrogenase (EC 1.3.99.2)	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0
OG_2216	0.005128	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit (EC 6.4.1.4)	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
OG_2215	0.005128	Biotin carboxyl carrier protein of methylcrotonyl-CoA carboxylase	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
OG_2214	0.005128	Biotin carboxylase of methylcrotonyl-CoA carboxylase (EC 6.3.4.14)	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
OG_2213	0.005128	Isovaleryl-CoA dehydrogenase (EC 1.3.99.10)	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
OG_2691	0.004706	FIG00672976: hypothetical protein	0	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
OG_3219	0.004357	toxin-antitoxin system, antitoxin	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	1	1	1	0

		component, Xrefamily																									
OG_3218	0.004357	toxin-antitoxin system, toxin component, RelEfamily	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	1	1	1	0
OG_2446	0.004064	Glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein (TC 3.A.1.1.3)	1	1	1	1	1	0	1	1	0	1	1	1	1	0	0	0	1	0	0	0	0	0
OG_3461	0.003926	urease accessory protein UreH	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0
OG_2212	0.00366	transcriptional regulator, MerR family protein	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
OG_2567	0.003394	SpoIID/LytB family protein	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1
OG_2665	0.003267	ABC transporter permease protein	0	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
OG_4533	0.002992	PAS domain S-box/diguanylate cyclase (GGDEF) domain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0
OG_4532	0.002992	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0
OG_1949	0.002819	Ribokinase (EC 2.7.1.15)	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	1	1
OG_4686	0.002793	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
OG_4096	0.002783	CRISPR-associated protein Cas7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	0
OG_2525	0.002655	Transcriptional regulator, TetR family	1	1	0	0	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0
OG_3570	0.00259	hypothetical protein	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	1
OG_3735	0.002576	hypothetical protein	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0
OG_3731	0.002576	hypothetical protein	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0
OG_3078	0.002343	sugar ABC transporter periplasmic protein	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
OG_1983	0.002317	hypothetical protein	2	0	1	0	1	2	1	1	3	1	0	1	1	0	1	1	0	0	0	0	0	0
OG_2322	0.00225	DNA-binding transcriptional regulator	0	1	1	1	1	0	1	1	1	2	1	1	1	0	0	0	0	1	0	0	0	0
OG_4599	0.002208	FMN-dependent NADH-azoreductase	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
OG_2526	0.00211	Transcriptional regulator, PadR family	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1

OG_2541	0.002068	3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157); 3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)	1	1	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0
OG_1839	0.002037	B12 binding domain / kinase domain / Methylmalonyl-CoA mutase (EC 5.4.99.2)	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0
OG_2394	0.001936	Dihydrodipicolinate synthase (EC 4.2.1.52)	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1	0	0	0	0
OG_2369	0.001895	Sialic acid utilization regulator, RpiR family	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1
OG_2829	0.001883	hypothetical protein	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
OG_2827	0.001883	hypothetical protein	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
OG_2814	0.001883	Beta-mannosidase (EC 3.2.1.25)	0	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
OG_4174	0.001833	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	1	0	0	0	0
OG_4564	0.001769	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0
OG_4787	0.001717	DNA translocase FtsK	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OG_1915	0.001702	Stage V sporulation protein AF (SpoVAF)	1	1	0	0	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1
OG_2521	0.001675	FIG00675231: hypothetical protein	0	1	1	1	0	0	1	1	1	0	2	2	1	0	0	0	0	0	0	0	0	0
OG_3795	0.001661	hypothetical protein	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0
OG_4528	0.001533	CRISPR-associated helicase Cas3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0
OG_4527	0.001533	CRISPR-associated protein Cas5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0
OG_4526	0.001533	CRISPR-associated protein Cas8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0
OG_1477	0.001479	CDS_ID OB1060	2	1	1	1	1	0	1	1	2	1	1	2	2	1	1	0	1	1	0	0	0	0
OG_4377	0.00146	Uncharacterized protein conserved in bacteria	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
OG_718	0.001357	3-ketoacyl-CoA thiolase (EC 2.3.1.16) @ Acetyl-CoA acetyltransferase (EC 2.3.1.9)	1	2	1	1	1	1	1	1	1	2	1	1	1	0	0	0	2	2	1	1	1	0
OG_4246	0.001292	FIG00673613: hypothetical protein	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0

OG_4245	0.001292	hypothetical protein	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
OG_4244	0.001292	Phage major tail protein	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
OG_4243	0.001292	hypothetical protein	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
OG_4242	0.001292	hypothetical protein	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
OG_4389	0.00125	transcriptional regulator, Xre family	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1
OG_4362	0.00125	potassium/proton antiporter	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1
OG_4332	0.00125	RNA polymerase sigma-70 factor	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1
OG_4331	0.00125	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1
OG_2574	0.001202	spore peptidoglycan hydrolase (N-acetylglucosaminidase) (EC 3.2.1.-)	0	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	0
OG_97	0.001176	Sulfur carrier protein adenylyltransferase ThiF	1	1	0	0	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1	2
OG_1808	0.001176	SPFH domain/band 7 family protein	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0
OG_2474	0.001175	N-acetylmuramic acid 6-phosphate etherase (EC 4.2.-.-)	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1
OG_2471	0.001175	PTS system, N-acetylmuramic acid-specific IIB component (EC 2.7.1.69) / PTS system, N-acetylmuramic acid-specific IIC component (EC 2.7.1.69)	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1
OG_2096	0.00116	Hypothetical protein, ydbS homolog	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1
OG_2088	0.00116	DNA-binding response regulator, AraC family	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	1
OG_1753	0.001158	HNH endonuclease	0	0	0	0	0	0	0	0	0	1	0	0	0	12	2	1	0	0	1	1	1	0
OG_2642	0.001114	nitric oxide synthase, oxygenase domain	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	1	1	1	1	1
OG_2491	0.001113	Putative regulatory protein	0	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0
OG_3032	0.0011	hypothetical protein	1	2	0	0	0	2	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
OG_4648	0.00106	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0

OG_2761	0.001035	sugar ABC transporter permease	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1
OG_3216	0.001033	CRISPR-associated RAMP protein, Cmr6 family	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	1	1	1	0
OG_3215	0.001033	CRISPR-associated RAMP protein, Cmr4 family	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	1	1	1	0
OG_2318	0.000993	Succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16)	1	3	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
OG_2880	0.000976	TOMM biosynthesis dehydrogenase (protein B)	1	1	0	0	0	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
OG_2649	0.000958	phosphate transport system regulatory proteinPhoU	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
OG_2648	0.000958	phosphate ABC transporter, substrate-bindingprotein PstS	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
OG_1816	0.00095	Transcriptional regulator, TetR family	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	0
OG_2449	0.000942	small, acid-soluble spore protein H	1	0	0	0	1	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1
OG_2448	0.000942	sulfite reductase [NADPH] flavoprotein, alpha-component	1	0	0	0	1	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1
OG_2674	0.000917	Nitrite reductase [NAD(P)H] small subunit (EC1.7.1.4)	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1
OG_2999	0.000902	Cytochrome c biogenesis protein	0	1	1	0	0	1	1	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0
OG_3418	0.00075	putative type II restriction enzyme methylase subunit	0	0	1	0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
OG_2637	0.0007	transposase, IS200/IS605 family	0	1	0	0	0	0	0	0	0	0	1	2	0	1	1	1	0	0	1	2	1	0
OG_3214	0.000693	CRISPR-associated protein, Csx1 family	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	1	1	1	0
OG_3213	0.000693	CRISPR-associated protein, Csm6 family	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	1	1	1	0
OG_3138	0.000664	Spore coat protein	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1
OG_3085	0.000664	cell envelope-related function transcriptional attenuator common domain	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1
OG_2063	0.000617	2-dehydropantoate 2-reductase (EC 1.1.1.169)	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0	1	0

OG_4	0.00061	Mobile element protein	1	1	1	1	8	1	2	2	8	7	2	14	10	1	1	1	0	0	1	2	1	0
OG_2418	0.00061	Methylglutaconyl-CoA hydratase (EC 4.2.1.18)	1	1	0	0	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
OG_3709	0.000583	tryptophan synthase subunit beta	0	1	0	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0
OG_2967	0.000583	ABC transporter substrate-binding protein	0	1	0	0	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
OG_2934	0.000581	FMN reductase (EC 1.5.1.29)	0	1	0	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
OG_2398	0.000576	Predicted beta-glucoside-regulated ABC transport system, permease component 1, COG1175	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0
OG_2517	0.00057	Late competence protein ComEC, DNA transport	0	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	2	0	0	0	0	0
OG_2434	0.000531	amino acid permease family protein	0	1	0	1	0	0	1	2	2	0	1	1	2	0	0	0	0	0	0	0	0	0
OG_2534	0.000526	FIG00674644: hypothetical protein	1	1	1	1	0	0	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0
OG_5237	0.00045	FIG00675263: hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0
OG_5204	0.00045	hypothetical protein	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0

Chapter 6 General discussion

6.1 Introduction

Thermophilic sporeformers are primary contaminants in the dairy concentrate-processing industry. The contamination of thermophilic sporeformers is not reported to cause human disease, but it can cause undesirable quality of the dairy products when temperature and humidity is favorable for them to germinate and grow. Therefore, this group of bacteria has been used as the hygienic indicator for the dairy concentrate-processing industry to ensure a desirable quality for these products. It is hypothesized that this group of bacteria is mostly originating from the dairy farm environment (e.g. silage, feed and soil) (Scheldeman et al., 2005; Pereira and Sant'Ana, 2018). During the processing of dairy-concentrates, while most bacteria cannot withstand harsh environmental conditions as well as treatment (e.g., heating, drying and the use of cleaning in place (CIP) etc.) commonly applied in dairy concentrate-processing facilities, those thermophilic sporeformers survive those conditions due to their capacity in spore formation. Moreover, biofilm formation and rapid growth at high temperature can help them to ultimately be present in the dairy concentrate-processing facilities. Due to the resistant properties of thermophilic sporeformers, more effective hygienic treatments are required to control their contamination of dairy powder products. Traditional methods aimed at controlling thermophilic sporeformers include shorter production runs, temperature alteration, reduction of surface areas at the optimal temperature zone, and the use of dual equipment (Burgess et al., 2010). However, because of the insufficient understanding of thermophilic sporeformers in dairy concentrate-processing environments, the development of more cost-effective and efficient hygienic treatments against these microorganisms is challenging, particularly considering the increased complexity of production plant design and product composition. With the emergence of high-throughput techniques, this exploration has allowed the gaining of insights at the molecular level. This study focused on the behavior of thermophilic sporeformers, including their origin, prevalence, and lifecycle, in dairy concentrate-processing environments with the help of the high-throughput molecular techniques.

Taken from the entire study, this chapter (General discussion) summarizes the biotic and abiotic factors influencing the growth and accumulation of thermophilic sporeformers; elucidates the characteristics of biofilm development of thermophilic sporeformers; addresses factors potentially involved in spore formation and heat

resistance of thermophilic sporeformers; lastly, elaborates current hurdles in using genomic approaches for predicting of microbial behavior.

6.2 Factors influencing the growth and accumulation of thermophilic sporeformers

Free monosaccharides

A carbon source is a critical requirement for bacterial growth, as this element is a principal component of biomolecules, which ultimately contribute to the bacterial structure and metabolism. In the context of milk, the most relevant carbon source is lactose ($C_{12}H_{22}O_{11}$), a disaccharide originated from the condensation of galactose and glucose. The metabolism of lactose requires two enzymes: a permease or an ABC transporter to transport lactose into the cell, and β-galactosidase to cleave the lactose molecule to yield glucose and galactose (Griffiths et al., 1999). On the basis of comparative genome analysis, this thesis showed that *G. thermoglucosidans* strains lack lactose-utilization related genes; in contrast, most *A. flavithermus* strains collected in this study are equipped with the genes involved in lactose uptake and utilization. Growth experiments showed that all *A. flavithermus* strains, except *A. flavithermus* WK1, could form visible colonies on skim milk plates, whereas the *G. thermoglucosidans* strains tested showed no visible growth on skim milk plates (like *A. flavithermus* WK1). Growth assays in liquid milk with the *A. flavithermus* and *G. thermoglucosidans* strains showed the same trend compared to growth on skim milk plates; nevertheless, with the addition of glucose and galactose to the skim milk, the growth of *G. thermoglucosidans* in skim milk is restored within a short time (Chapter 3). This suggests that free monosaccharide can promote growth of *G. thermoglucosidans* in skim milk when lactose is the sole energy source. Chapter 5 showed that other thermophilic sporeformers that do not have the lactose utilization capability were likely also unable to readily grow in milk. It is assumed that the supplementation of monosaccharides can restore their growth in skim milk. Further study is needed to confirm this assumption. From the recent literature, this variation of the ability of different thermophilic sporeformers to grow in skim milk was not

reported. This is the first study which discussed the relation between the capability of lactose utilization and the spoilage of thermophilic sporeformers in skim milk. This relation also suggests an enhanced growth capability for thermophilic sporeformers when there is a presence of free monomeric sugars in the ingredients. Therefore, limiting the presence of monomeric sugars during the product formulation, raw material selection, or processing in the dairy powder-processing facilities can be considered as a countermeasure for preventing growth of thermophilic sporeformers that cannot utilize lactose in milk ; along with this, other parameters including physicochemical and microbial properties of milk must be evaluated to ensure the product quality.

Other members in the dairy-associated microbiota

Dairy-associated microbiota show remarkable diversity, as previously reported and reflected in the study of dairy farm isolates on seven sporeforming genera, i.e., *Aneurinibacillus, Bacillus, Brevibacillus, Geobacillus, Paenibacillus, Ureibacillus,* and *Virgibacillus* (Scheldeman et al., 2005). The present study confirmed the presence of the sporeforming genera *Aneurinibacillus*, *Bacillus, Brevibacillus, Geobacillus,* and *Anoxybacillus* in the dairy concentrate-processing plant studied（Chapter 2). Moreover, in various samples taken from the dairy concentrate-processing plant, the co-existence of a diverse group of microorganisms was observed. This implies a potential dependence among these microorganisms for survival and growth. One such dependence which was discovered and studied in this thesis is that of *G. thermoglucosidans* on *A. flavithermus*.

This ecology and the interrelationship between *G. thermoglucosidans* and *A. flavithermus* were elucidated in Chapter 3. It is confirmed in this study that, in a co-culture, the hydrolysis of lactose via β-galactosidase activity by *A. flavithermus* supplies glucose, galactose, or both, to *G. thermoglucosidans*, thus supporting its growth in skim milk (Chapter 3). The most frequently described example of different species which have influence on each other in dairy environments relates to the yogurt consortium, where the proteolytic activity of *Lactobacillus delbrueckii* subsp. *bulgaricus* results in the supply of amino acids for *Streptococcus thermophiles* growth (Sieuwerts et al., 2008). Such information is relevant because it suggests that the presence of proteolytic microorganisms in the dairy concentrate production line may contribute to the diversity and spore load of specific thermophiles in end-

products. On the basis of this finding, it is suspected that by targeting at one type of less heat-resistance sporeformers, the growth of the more heat-resistance sporeformers can be prevented. However, how realistic this approach would be requiring further understanding of this dependence and further studying the characteristics of other thermophilic sporeformers.

6.3 Characteristics of biofilm development of thermophilic sporeformers in a static biofilm system

Biofilm formation at the air-liquid interface

In this study, a clear difference in the preferred environment for biofilm formation of microbial genera was identified: *Anoxybacillus* and *Geobacillus* preferentially reside at the air-liquid interface, whereas *Pseudomonas* accumulated at the surface of submerged steel (Chapter 2). It is hypothesized that the ability to grow in relation to the concentration of oxygen may play a crucial role in selective accumulation of bacteria and spores on the stainless-steel surface at the air-liquid interface. This suggests that biofilms of thermophilic sporeformers and associated spores may particularly develop at elevated temperatures and in industrial piping systems that are only partly filled and, as a result, are exposed to oxygen during operation causing aerotaxis (migration towards oxygen) of thermophilic sporeformers towards oxygen (Laszlo et al., 1984). This selective attachment of biofilm of different bacteria species have also been reported for mesophilic *Bacillus* spp. (Morikawa et al., 2006; Wijman et al., 2007). The observations from this study about the preferred formation of thermophilic sporeformers at air-liquid interface underlines the importance of the good closure of piping systems in preventing biofilm formation by thermophilic sporeformers in dairy-processing facilities. Moreover, when studying diversity in biofilm formation by thermophilic sporeformers, it is important to bear in mind that environmental conditions have a large impact on the results.

Biofilm developmental stages

The present study investigated the biofilm development mechanism of thermophilic sporeformers in dairy concentrate-processing environments. For this purpose, a whole genome transcriptomic study during biofilm development of one model strain, *G. thermoglucosidans* TNO-09.020, isolated from a dairy concentrate-processing plant, was conducted (Chapter 4). Throughout the study, based on the growth patterns, three distinct stages of biofilm development for *G. thermoglucosidans* in the static biofilm system were observed: the initiation, development, and maturation stages. In the initiation stage, the cells attach to abiotic surfaces. In the development stage the cells proliferate and aggregate themselves to the surface. In the maturation stage the cells disperse into the medium (Parkar et al., 2003), or lyse after their death caused by environmental stress, and lysis increases over time in the static biofilm system. In this context, it is important to remember that cell lysis has been shown to play an essential role in intercellular adhesion and biofilm stability (Bayles, 2007). Referring to Chapter 4, the initiation and development stages constitute the early biofilm phase; the maturation stage including cell dispersal and cell lysis is the late biofilm phase. In this chapter, a significant difference in the expression profiles of the cells was observed between cells in early biofilm phase and cells in late biofilm phase. Regarding the role of cell dispersal and cell lysis in biofilm resistance and stability, to prevent the biofilm development reaching the maturation stage is needed, for example by using a more frequent cleaning scheme to prevent the progression of biofilm development into the maturation stage, since the two phenomena (both cell dispersal and cell lysis) could contribute to the recontamination of the product and/or to the persistence of the biofilm in the processing environment.

6.4 Factors potentially involved in spore formation and heat resistance of thermophilic sporeformers

pH

Concerning the sporulation of thermophilic sporeformers, several studies have dealt with the locations where the spores were frequently detected in the manufacturing plants (e.g. direct steam injection [DSI] and evaporator) (Scott et al., 2007; Zhao et al., 2013). Only a few investigations have focused on the study on factors affecting the production of spores by thermophilic sporeformers. Spore formation by thermophilic sporeformers has been reported to be influenced by aeration (Long et al., 1959), manganese (Thompson et al., 1967), and pH (Yazdany and Lashkari., 1975). In this study, the inhibitory effect of rapid acidification of the growth medium on the sporulation of thermophilic sporeformers was also observed. In skim milk supplemented with glucose and in lactose-free milk (hydrolyzed lactose), no visible spores were detected using qualitative microscopy in samples from the *G. thermoglucosidans* and *A. flavithermus* cultures; and significant drop in the pH of the growth medium was observed (final pH < 5.5). Nevertheless, spores of *G. thermoglucosidans* and *A. flavithermus* were observed after the growth in the galactose-supplemented skim milk (final pH of 6). Similarly, in the study of Yazdany and Lashkari (1975), *G. stearothermophilus* ATCC 7953 and *G. stearothermophilus* NCIB 8919 produced few spores when cultured in a medium with final pH around 5.5, whereas when grown in a medium adjusted to pH 7.7 to 8.7, sporulation of the *G. stearothermophilus* occurred. It is conceivable that glucose metabolism-induced acidification leads to a reduction in the total CFU counts, hence preventing the formation of spores.

Development stages

In the present study, two different spore measurement techniques were used to allow for a differentiation of the heat stability of the spores (80°C, 10min and 100°C, 30min). Only at the late growth phase did the results of these two treatments match. At the early growth phase, not all spores were very heat resistant, especially in the planktonic phase. The comparison of the outcomes of these two measurements

indicated that the spores from the early growth phase are less resistant to heat compared to the spores from the late phase. It would be worth further investigating the spore maturation process of thermophilic sporeformers. It is known that the matured spores are more resistant to heat or other environmental stresses, compared to the spores in the early development stages (Sanchez-Salas et al., 2011). This suggests that if the food processing industries intend to eradicate the spores, they could take hygienic measures to eliminate the possibility of the spores reaching a maturation stage during food processing.

6.5 Using genomic analysis to study the behavior of thermophilic sporeformers in dairy-powder processing environments

Thanks to the emergence of various "omics" technologies such as metagenomics, proteomics, transcriptomics, metabolomics etc., researchers were able to produce genome-scale, or "omics" data sets[5] which reveal not only the static sequence of the genes and proteins, but also the biological function of the gene product in a high-throughput manner. For this study, using pyrosequencing, the location of the thermophilic sporeformers in a dairy concentrate-processing plant was determined. Transcriptomic analysis permitted the identification of genes that are correlated to the biofilm-development stages. Comparative genomics was used to explore the genomic features in those thermophilic sporeformers, related to contaminations in dairy concentrate-processing environments

This 'omics' data yields novel insights on the cellular inner workings of those not well-known thermophilic sporeformers. However, this abundance of information also presents many hurdles, the main one being the extraction of discernible biological

5 Omics data set: A generic term that describes the genome-scale data sets that are emerging from high-throughput technologies. Examples include whole genome sequencing data (genomics) and microarray-based genome-wide expression profiles (transcriptomics).

meaning from multiple omics data sets. Some hurdles encountered in this study are discussed below.

In order to explore biofilm-formation related genes, expression profiles of planktonic-phase cells were compared with those of biofilm-phase cells. Moreover, to investigate the dairy-contamination related genes, gene-trait matching analysis (GTM) was applied. The genome-wide transcriptome analysis points to the molecular processes involved in the biofilm development of the thermophilic sporeformers. In the late phase of biofilm development of *G. thermoglucosidans*, since there is a significant difference between the expression profiles of the planktonic cells and biofilm cells, we investigated the genes that contributed to this differentiation by using un-supervised mathematical algorithms and selected phenylalanine-degradation related genes which rank highest in determining the difference between biofilm phase cells and planktonic phase cells in the late phase. In one experiment it was observed that phenylalanine promoted spore formation for *G. thermoglucosidans*. In the literature, one reference to the effect of phenylalanine on biofilm formation in a crystal violet assay was found (Bernier et al., 2011), but no significant effect of phenylalanine on biofilm was observed in our study. In other words, it is clear that the mechanism behind the promotion of spore formation of *G. thermoglucosidans* by phenylalanine remains unknown and needs to be further studied. Meanwhile, in the early stage of biofilm formation, it was found that genes related to methionine biosynthesis are upregulated. After observing no effect of amino acids from the methionine biosynthesis pathway on biofilm mass formation, experiments were done to examine whether the effect of methionine on biofilm is linked to the secretion of AI-2, a quorum-sensing molecule which has been reported to be related to biofilm attachment (Lebeer et al., 2007; Kaur et al., 2018). However, in the AI-2 assay, the presence of AI-2 in *G. thermoglucosidans* culture could not be detected. Nevertheless, it is possible that the upregulation of methionine biosynthesis is related to early cell attachment. The above-mentioned example shows the relevance of the genomic information revealed by an unsupervised method, given our current knowledge, sometimes cannot be easily determined.

Moreover, when one refers to the already existing knowledge to interpret the data collected from high throughput “omics” analyses, the specific conditions and possible deviations need to be taken into careful consideration. For example,

following a transcriptomic study on biofilm development (Chapter 4), the results were compared with those of similar studies. This revealed that the gene categories described as being associated with biofilm formation in other studies coincide very little with those found in this one because different researchers made use of different methods, conditions and model organisms. It is proven that biofilm development is dependent on the biofilm-forming organisms and the circumstances under which biofilms form (Lazazzera, 2005; Beloin and Ghigo, 2005; Kjelleberg and Givskov, 2007). It indicates that either the thermophilic sporeformer or the conditions applied in the research required the cells to follow different strategies in developing biofilm, thus it is important to note that the conclusion made from a transcriptomic study is only a reflection of gene expression under the particular experiment conditions, unless proven otherwise.

All in all, 'omics' has triggered a paradigm shift in experimental study design, expanding beyond hypothesis-driven approaches to research that is basically explorative (Mayer et al., 2011). With the increasing amount of data and knowledge in the relevant fields, we should be able to make more accurate interpretations and deduce the data's underlying implications. This abundance of information also presents many hurdles, so as the extraction of discernable biological meaning from multiple omics data sets. However, investigators are making progress in identifying, extracting and interpreting biological insights from omics data sets. One successful approach to do so requires the integration of omics data (Joyce et al., 2006), which means the use of multiple sources of omics data to provide a better understanding of a process or a system. In short, this approach uses a range of experimental and statistical methods to quantitate and integrate intermediate phenotypes, such as transcript, protein and metabolite levels. For example, a process which is not well captured in transcriptomics in one study might be supplemented by proteomics data from another study to facilitate understanding. This research area attracts the attention of many researchers who indicate various advantages of such an integrated approach and, at the same time, the need to develop novel data integration methodologies (Gomez-Gabrero et al., 2014).

6.6 Final remarks

This study's observations of the dependence between *A. flavithermus* and *G. thermoglucosidans* shed new light on the inter-relationship within microbiological communities of dairy sporeformers. The new insights could have relevant implications for industrial hygiene operations. The comparisons of the physiology of cells in the early growth and late growth phases under various circumstances also produced some insights. These underline the importance of earlier cleaning for the practical prevention of biofilms or spores in dairy concentrate-processing environments. Moreover, the growth-promotion effect of free monomeric sugars was also emphasized in this study. This implies, on one hand, that dairy powder-processing facilities can consider the possibility of enhanced growth capability for thermophilic sporeformers when free monomeric sugars are present in the ingredients. Therefore, limiting the presence of monomeric sugars can be considered as a preventive measure against contamination by *G. thermoglucosidans* during the product formulation, raw material selection, or processing. On the other hand, if specific lactose-utilization related genes for thermophilic sporeformers can be identified, together with the known lactose-utilization related genes, this information can be used as a biomarker for the prediction of growth potential of thermophilic sporeformers in skim milk. In practice, this biomarker can be identified at either DNA or RNA level. With regard to the transcriptomic studies, the upregulation of genes involved in methionine acquisition in the early biofilm formation phase, and the upregulation of phenylalanine-degradation related genes in the late biofilm formation phase, provide interesting leads for further investigation. Lastly, the results of "omics" analyses can provide important clues about the key characteristics of thermophilic sporeformers with regard to the contamination of dairy concentrate-processing environments. A better understanding of these thermophilic sporeformers can lead to the development of more effective control measures to prevent contaminations from thermophilic sporeformers in dairy concentrate-processing environments.

6.7 References

Bayles, K. W. 2007. The biological role of death and lysis in biofilm development. Nature Reviews Microbiology. 5:721-726.

Beloin, C. and J. M. Ghigo. 2005. Finding gene-expression patterns in bacterial biofilms. Trends in Microbiology. 13:16-19.

Bernier, S. P., D. G. Ha, W. Khan, J. H. Merritt and G. A. O'Toole. 2011. Modulation of *Pseudomonas aeruginosa* surface-associated group behaviors by individual amino acids through c-di-GMP signaling. Research in Microbiology. 162:680-688.

Burgess, S. A., D. Lindsay and S. H. Flint. 2010. Thermophilic bacilli and their importance in dairy processing. International Journal of Food Microbiology. 144:215-225.

Gomez-Cabrero, D., I. Abugessaisa, D. Maier, A. Teschendorff, M. Merkenschlager, A. Gisel, E. Ballestar, E. Bongcam-Rudloff, A. Conesa, J. Tegnér. 2014. Data integration in the era of omics: current and future challenges. BMC Systems Biology. 8 (suppl2):I1

Griffiths, A. J. F., W. M. Gelbart and J. H. Miller. 1999. Regulation of the lactose system-Modern genetic analysis. W. H. Freeman, New York, US.

Joyce, A. R., B. Ø. Palsson. 2006. The model organism as a system: integrating 'omics' data sets. Nature Reviews Molecular Cell Biology. 7:198.

Kaur, A., N. Capalash and P. Sharma. 2018. Quorum sensing in thermophiles: prevalence of autoinducer-2 system. BMC Microbiology. 18:62.

Kjelleberg, S. and M. Givskov. 2007. The Biofilm Mode of Life: Mechanisms and Adaptations. Horizon Bioscience, Lyngby, Denmark.

Laszlo, D. J., Niwano, M., W. W. Goral and B. L. Taylor. 1984. *Bacillus cereus* electron transport and proton motive force during aerotaxis. Journal of Bacteriology.159:820-824.

Lazazzera, B. A. 2005. Lessons from DNA microarray analysis: The gene expression profile of biofilms. Current Opinion in Microbiology. 8:222-227.

Lebeer, S., S. C. J. De Keersmaecker, T. L. A. Verhoeven, A. A. Fadda, K. Marchal and J. Vanderleyden. 2007. Functional analysis of luxS in the probiotic strain *Lactobacillus rhamnosus* GG reveals a central metabolic role important for growth and biofilm formation. Journal of Bacteriology. 189:860-871.

Long, S. K. and O. B. Williams. 1960. Factors affecting growth and spore formation of *Bacillus stearothermophilus*. Journal of Bacteriology. 79:625.

Mayer, B. 2011. Bioinformatics for omics data: methods and protocols. Humana Press, Vienna, Austria.

Morikawa, M. 2006. Beneficial biofilm formation by industrial bacteria *Bacillus subtilis* and related species. Journal of Bioscience and Bioengineering. 101:1-8.

Parkar, S. G., S. H. Flint and J. D. Brooks. 2003. Physiology of biofilms of thermophilic bacilli - potential consequences for cleaning. Journal of Industrial Microbiology and Biotechnology. 30:553-560.

Pereira, A. P. M. and A. S. Sant'Ana. 2018. Diversity and fate of spore forming bacteria in cocoa powder, milk powder, starch and sugar during processing: A review. Trends in Food Science & Technology. 76:101-118.

Sanchez-Salas, J. L., B. Setlow, P. Zhang, Y. Q. Li and P. Setlow. 2011. Maturation of released spores is necessary for acquisition of full spore heat resistance during *Bacillus subtilis* sporulation. Applied and Environmental Microbiology. 77:6746-6754.

Scheldeman, P., A. Pil, L. Herman, P. De Vos and M. Heyndrickx. 2005. Incidence and diversity of potentially highly heat-resistant spores isolated at dairy farms. Applied and Environmental Microbiology. 71:1480-1494.

Scott, S. A., J. D. Brooks, J. Rakonjac, K. M. Walker and S. H. Flint. 2007. The formation of thermophilic spores during the manufacture of whole milk powder. International Journal of Dairy Technology. 60:109-117.

Sieuwerts, S., F. A. M. de Bok, J. Hugenholtz and J. E. T. van Hylckama Vlieg. 2008. Unraveling microbial interactions in food fermentations: From classical to genomics approaches. Applied and Environmental Microbiology. 74:4997-5007.

Tan, S. Y. E., S. C. Chew, S. Y. Y. Tan, M. Givskov and L. Yang. 2014. Emerging frontiers in detection and control of bacterial biofilms. Current Opinion in Biotechnology. 26:1-6.

Thompson, P. J. and O. A. Thames. 1967. Sporulation of *Bacillus stearothermophilus*. Applied Microbiology. 15:975-979.

Wijman, J. G., P. P. de Leeuw, R. Moezelaar, M. H. Zwietering and T. Abee. (2007). Air-liquid interface biofilms of *Bacillus cereus*: formation, sporulation, and dispersion. Applied and Environmental Microbiology. 73:1481-1488.

Yazdany, S. and K. B. Lashkari. 1975. Effect of pH on sporulation of *Bacillus stearothermophilus*. Applied Microbiology. 30:1-3.

Zhao, Y., M. P. M. Caspers, K. I. Metselaar, P. de Boer, G. Roeselers, R. Moezelaar, M. N. Nierop Groot, R. C. Montijn, T. Abee, R. Kort. 2013. Abiotic and microbiotic factors controlling biofilm formation by thermophilic sporeformers. Applied and Environmental Microbiology. 79:5652-5660.

Summary

Thermophilic sporeformers are a group of bacteria that proliferate at high temperatures (mostly above 50°C as optimum growth temperatures) and can form spores as a mean to survive in unfavourable environmental conditions. They are of primary concern for food processing plants that apply high temperatures. Even though they have been reported as not pathogenic, they could pose a threat to the quality of food products when they grow because their growth produces off flavour and changes the texture of the products. For this reason, they are sometimes used as a hygiene indicator in food processing environments. Customers of intermediate products also often set strict specifications for thermophilic sporeformers. Those specifications are difficult to achieve, thus present an important challenge to the dairy industry worldwide. To meet the strict specifications, a systematic approach to understanding the source of the contamination and the transmission routes of thermophilic bacilli in dairy-concentrate processing environments is needed.

In this thesis, the composition of the natural flora of thermophilic sporeformers present along an entire dairy-concentrate processing line is studied (Chapter 2), and the environmental and physiological factors contributing to their rapid growth in dairy-concentrate processing environments is explored (Chapter 2, Chapter 3). Furthermore, it has been reported that biofilms contribute to the persistence of thermophilic sporeformers in dairy-concentrate processing environments. For this reason, specific genes that are involved in biofilm formation and development in thermophilic sporeformers, were also subjects of study in this thesis (Chapter 4). This has produced a better understanding of the mechanism of biofilm formation in thermophilic sporeformers. Lastly, we applied comparative analysis to explore the genomic and physiological characteristics of a group of dairy and non-dairy associated thermophilic bacilli in relation to the contamination in the dairy-concentrate processing environments (Chapter 5). Altogether, the research described here has produced new insights into the life cycle of thermophilic sporeformers in dairy-concentrate processing environments. First of all, the presence of thermophilic sporeformers was observed in dairy-concentrate processing lines, particularly at locations where high temperatures were applied. Secondly, for the sporeformers to readily grow in dairy environments, the ability to utilize lactose seems essential. If this ability is absent, monomeric sugar such as glucose and galactose is required for

growth in such environments. Thirdly, genes involved in certain processes, such as methionine biosynthesis and phenylalanine utilization, seem to be relevant in the formation of thermophilic bacilli biofilms. Lastly, a niacin transporter, which is only present in thermophilic bacilli that can readily grow in skim milk, appears to be a relevant genetic marker for ability to grow in dairy-concentrate processing environments.

Acknowledgements

For the final finishing of this PhD work, I would like to thank all the people who have provided help and support. Without you, this PhD thesis would not be successful.

First, I would like to thank Marcel and Michel. Even though it was difficult, and you were also very busy, you did not give up on me and provided me the last chance to finish up this piece of work on biofilm thermophilic sporeformers. I also like to thank Remco, Jos, Tjakko and Masja for your supervision. I have learned a lot about conducting research from you.

Second, I would like to thank TIFN and TNO who provided me the financial and infrastructural support for my PhD. Without those support, it would not be possible for the curious minds to look into the behaviour of those thermophiles. Of course, the opportunities provided by TIFN to periodically talk about our progress with other scientists were also very meaningful. I often could talk about my hurdles during my research with other scientists and experts. The exchange of ideas enriched my research and also myself as a young scientist.

Further, I like to thank the technical support, with special thanks to Martien. Martien was always there when I needed some form of support, technical, bioinformatic and even sometimes a bit of coaching when the PhD project got difficult. Besides Martien, Marian, Trifa, Lars, and Jos were really supportive. Jos was really smart with data analysis and really a professional support, although sometimes, I needed extra explanation from Martien for me to understand all the ins and outs of the analysis. Lars, on the other hand, was very creative in designing and operating experiments. In the last phase of the PhD work, without his creativity, the high-throughput phenotypic assay would not have succeeded.

Hereby, I also would like to thank all the colleagues at both Food microbiology group and also Microbiology and system biology group. Those coffee break conversations and also the afterwork parties have definitely made the stressful and sometimes lonely PhD period social and also entertaining.

And lastly, I like to thank my family and friends for your support and trust in me. I am extremely thankful that because of my ambition to finish my PhD, my family in China

did not force me to follow the “normal” life journey of a Chinese woman. On the contrary, in the midst of difficulties, my grandfather, who passed away a couple of years ago, always encouraged me. I am also very thankful towards my dear friend Hanne and my Dutch mother Jul. During the most difficult period, you were there to support me and hold me. Those loving messages and words sometimes really could lift me from a moment of frustration. Last couple of months was a stressful period. I also like to thank Jan for your love, patient listening and support which gave me so much confidence for bring my PhD work towards its finish line.

List of publications

Zhao, Y., M. P. M. Caspers, T. Abee, R. J. Siezen and R. Kort. (2012). Complete genome sequence of *Geobacillus thermoglucosidans* TNO-09.020, a thermophilic sporeformer associated with a dairy-processing environment. Journal of Bacteriology. 194(15), 4118-4118.

Zhao, Y., M. P. M. Caspers, K. I. Metselaar, P. de Boer, G. Roeselers, R. Moezelaar and R. Kort. (2013). Abiotic and microbiotic factors controlling biofilm formation by thermophilic sporeformers. Applied and Environmental Microbiology. 79(18), 5652-5660.

Zhao, Y., M. Kumar, M. P. M. Caspers, M. N. Nierop Groot, J. M. B. M. van der Vossen and T. Abee. (2018). Growth of dairy isolates of *Geobacillus thermoglucosidans* in skim milk depends on lactose degradation products supplied by *Anoxybacillus flavithermus* as secondary species. Journal of Dairy Science. 101(2), 1013-1019.

Curriculum vitae

Yu Zhao was born on May 5th,1986 in Changsha, Hunan, China. In 2004, she graduated from Yali secondary school. She continued her bachelor study in bio-safety at Hunan Agricultural University. After her bachelor, she came to the Netherlands for her master study in Food safety at Wageningen University. She conducted her master thesis at NIZO studying the stress responses of probiotics. She did her second internship at DSM with a focus on high-throughput test methods for yogurt culture collection. In the end of the master study, she received the offer to continue a study in the area of food microbiology towards a PhD degree. Her PhD work is titled "Thermophilic sporeformers from dairy processing environments". The results of this work are described in this thesis. Currently, Yu is working as an integrated science teacher at International school Laren.

VLAG graduate school – Overview of completed training activities

Discipline- specific activities

- Systems biology course " statistics of ~omics data analysis" VLAG
- Conference: Biofilm 5 INRA
- Course "BioIT for biologist" TIFN
- Conference: China International food safety & quality conference China FDA
- Course "Genetics and physiology of food-associated micro-organisms" VLAG
- Lab Training in microphages handling food microbiology group (WUR)
- TIFN Experts meetings TIFN
- Symposium: TIFN we day TIFN
- Course "Food fermentation" VLAG

General courses

- VLAG PhD week VLAG
- Expatriate Training TNO
- Project and time management WGS
- Career perspectives WGS
- Teaching supervising MSc thesis students WU
- IP workshop TIFN
- English writing course for publication Babel
- Effective behavior in your professional surroundings WGS
- PhD peer consultation WGS

Optional courses

- Preparation of research proposal TIFN
- Weekly group meetings Microbiology and system biology (TNO)
- PhD study tour. Food microbiology
- PhD study tour. Food microbiology

The studies presented in this thesis were performed within the framework of TI food and Nutrition.

Cover & Layout design: Jan Zoutendijk || zoutendijk.co
Cover photo: Mae Mu || unsplash.com/@picoftasty
Printing: ProefschriftMaken || proefschrifmaken.nl