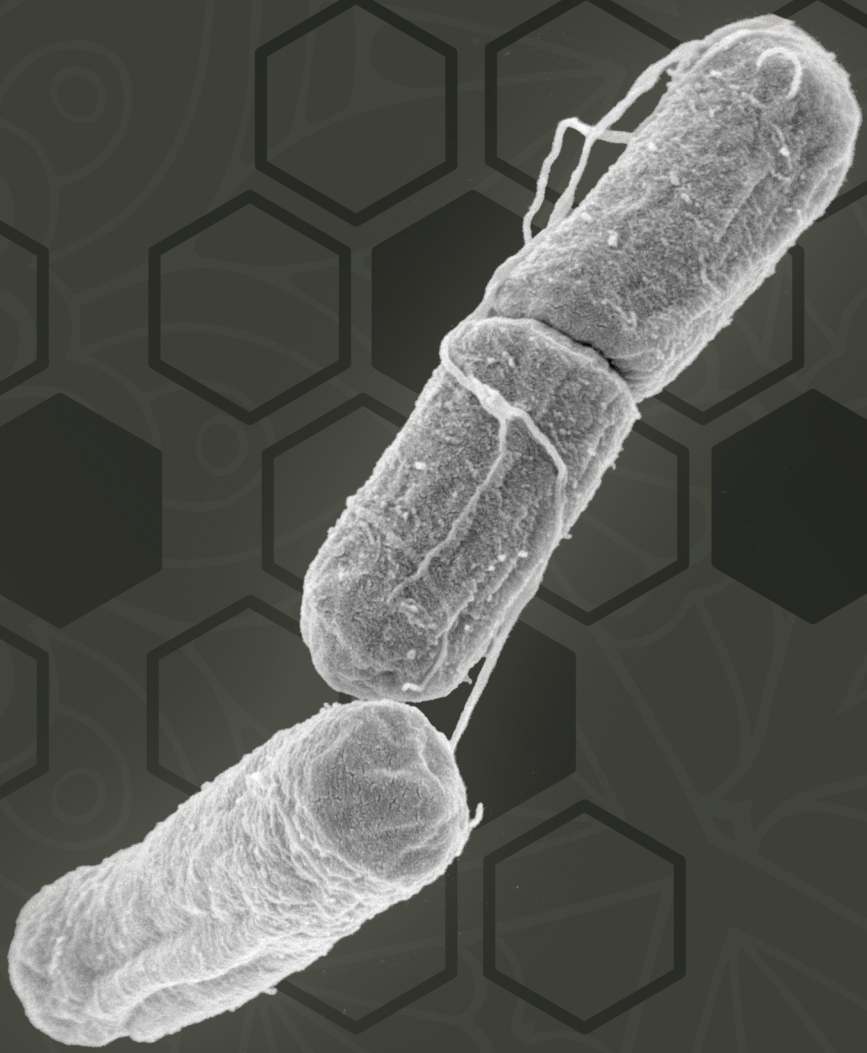
Lignocellulolytic capacities of *Geobacillus thermodenitrificans*

Towards consolidated bioprocessing



Martinus J.A. Daas

Lignocellulolytic capacities of *Geobacillus thermodenitrificans*:

Towards consolidated bioprocessing

Martinus J. A. Daas

Thesis committee

Promotors
Prof. Dr J. van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Prof. Dr R. van Kranenburg
Special professor Bacterial Cell Factories
Wageningen University & Research

Other members
Prof. Dr R. H. Wijffels, Wageningen University & Research
Prof. Dr L. Dijkhuizen, University of Groningen
Prof. Dr D. J. Leak, University of Bath, UK
Dr T. van Rij, DSM Biotechnology Centre, Delft

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

Lignocellulolytic capacities of *Geobacillus thermodenitrificans*:

Towards consolidated bioprocessing

Martinus J.A. Daas

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 Friday 23 June 2017
at 4 p.m. in the Aula.

Martinus J.A. Daas
Lignocellulolytic capacities of *Geobacillus thermodenitrificans*:
Towards consolidated bioprocessing,
180 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)
With references, with summary in English

ISBN 978-94-6343-164-4
DOI http://dx.doi.org/ 10.18174/411820

Propositions

1. Consolidated bioprocessing needs more scientific consolidation (this thesis)

2. HUS locus is a name that does not cover its content (this thesis)

3. Sequence identity based annotations lead to wrong interpretations of protein functions

4. The preference of SYBR safe over ethidium bromide shows that scientists not always go by facts

5. Referenda strengthen democracy

6. Feeling offended is a choice

7. Tall people have a better view on things

Propositions belonging to the thesis, entitled

Lignocellulolytic capacities of *Geobacillus thermodenitrificans:* Towards consolidated bioprocessing

Martinus J.A. Daas

Wageningen, 23 June 2017

Table of Contents

mmary of the performance of several candidate CBP organisms

Species and strain	Genotype	(hemi) cellulolytic	T °C	pH	M
oteurocense					ac
cii	Δpta /ack	Xylanolytic	55	6.5	MS
	ΔldhA/adhE-				(g/
ocellum		Cellulolytic and xylanolytic	72	67	0.58 Y
ellum and	ev. Δldh, Δhpt, Δpta				
lyticum	ev. Δldh, Δhpt, Δpta	Cellulolytic	55	70	78 YE
osidasius	ΔldhL pdh up.Δpdf	Cellulolytic and xylanolytic	55	63	78 YE
um	wt	-			
		Xylanolytic	60	67	108 YE
			60	67	38 YE

Chapter 1

General introduction and Thesis outline

General introduction

A brief history

Microorganisms have been used by humans for centuries, even before their existence was demonstrated by Antonie van Leeuwenhoek in the year 1676. The production processes of wine, cheese and beer were no longer thought to be caused by spontaneous reactions, but were demonstrated to be catalyzed by microorganisms (Ligon 2002). The best known microbial process is that of fermentation, a metabolic process in which sugars are converted to other molecules such as alcohol, CO_2 and/or acids. The understanding of this process began in the 19th century, when Liebig and Pasteur both claimed to understand the driving force behind fermentation. Liebig proposed fermentation to be a chemical process, in which yeast played an incidental role by disturbing molecules and their arrangements (Brock 1995). Pasteur, however, was convinced that fermentation required viable yeast cells, in which so-called "ferments" catalyzed the fermentation process (Ligon 2002). It was finally discovered by Buchner, that even with cell free extracts from yeast, sugars could be converted to ethanol and CO_2 (Morange 2007). We now know that both Liebig and Pasteur were partially right, and that viable organisms are needed to produce the enzymes required for fermentation, which by itself is a chemical process. The name enzyme originates from the Greek ἐν "in" and ζύμη "yeast", referring to their discovery as catalysts inside yeast cells. Enzymes catalyze reactions in which substrates are converted into products. Nomenclature of enzymes is often based on the reaction that is catalyzed or the substrate that is converted, complemented with the suffix –ase (*e.g.* xylanase is an enzyme that cleaves xylan). Enzymes have been used in many applications, such as in detergents, the brewing industry and in food processing (Li et al. 2012). Ever since their discovery, the search for novel enzyme activities and applications has been ongoing (Hult & Berglund 2003).

Lignocellulosic biomass

One of the applications of enzymes is their use in the conversion of biomass into fermentable sugars. The biomass referred to comprises all plants and plant-derived materials which are most often called lignocellulosic biomass. This lignocellulosic biomass can be degraded using saccharolytic enzymes and subsequently converted into bulk products such as fuels, polymer building blocks as well as a broad spectrum of other, high valuable chemical compounds (Fitzpatrick et al. 2010). Lignocellulose consists in general of 40-50% cellulose, 25-35% hemicellulose, 15-20% lignin and residual components (Menon & Rao 2012; Murciano Martínez et al. 2015). Cellulose is a linear polymer of glucose units, linked to each other by β-1, 4-glycosidic bonds. The degree of polymerization (DP) depends on the substrate, but is typically found in the range of 5000-7500 glucose units in primary cell walls spanning up to 15.000 glucose units for cellulose derived from cotton (Wyman et al. 2004). Cellulose chains contain crystalline and amorphous regions, and are extremely insoluble in water, a necessary trait as cellulose provides the stability in plant cells (Figure 1). De degradation of cellulose requires the synergistic action of three classes of enzymes: exoglucanases, endoglucanases and β-glucosidases (Garvey et al. 2013; Van Dyk & Pletschke 2012). Exoglucanases, often called cellobiohydrolases (CBH), include both the β-1,4-glucan glucohydrolases (EC 3.2.1.74), which can release glucose units from the cellulose chain and β-1,4-cellobiohydrolases (EC 3.2.1.91), which release cellobiose units from the cellulose chain. Exoglucanases are further classified based on whether they act on the reducing or non-reducing end of the cellulose (Figure 2A) (Wyman et al. 2004). Endoglucanases (EC 3.2.1.4) act randomly on cellulose, although more amorphous regions are preferred due to better penetration of the substrate (Hasunuma et al. 2013). By cleaving the cellulose chain more reducing and non-reducing ends are generated which act as a substrate for exoglucanases. This synergism between endo and exoglucanases is shown to be important for complete and efficient degradation of cellulose (Vazana et al. 2010). The β-glucosidases (EC 3.2.1.21) are a class of enzymes that act on cellobiose units, generated

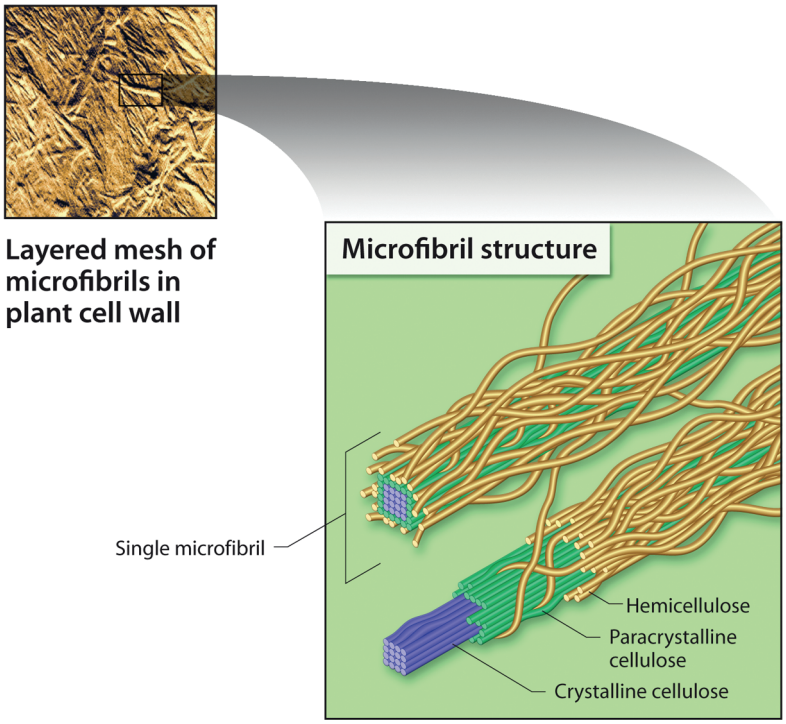


Figure 1 Cellulose, the main structural component of plant cell walls, is a linear polymer consisting of thousands of glucose residues arranged in a rigid, crystalline structure. These structures are arranged in multiple parallel layers, encased within a complex outer layer of amorphous cellulose and hemicellulose (U.S. DOE, 2005).

by cellobiohydrolases, by cleaving the cellobiose into two monomeric glucose units. β-glucosidases have been reported to also act on soluble cellodextrins, as well as an array of glycosides (Wyman et al. 2004). In natural substrates, cellulose is always present in a complex matrix with hemicellulose, thereby providing more recalcitrance to enzymatic breakdown. The structure of hemicellulose is more complex than that of cellulose and most often consists of a xylose backbone (xylans) branched with sugars like arabinose, mannose and other substituents like glucuronic and acetic acid groups. The chemical composition and structure of hemicellulose varies greatly depending on the source of biomass it is derived from. Xylans derived from grasses and annual plants often contain mostly arabinose and acetyl side groups, while xylans derived from hard woods contain mostly glucuronoxylans; a xylan rich in glucuronic-

and acetic acid substituents (Menon & Rao 2012; Losordo et al. 2016). In contrast to cellulose, an enzymatic diverse machinery is needed to break down hemicellulose. Complete degradation of this heteropolymer requires the synergistic action of endoxylanases (EC 3.2.1.8), β-xylosidases (EC 3.2.1.37), acetylxylan esterases (EC 3.1.1.72), L-arabinose releasing enzymes such as α-L arabinofuranosidase (EC 3.2.1.55) and arabinoxylan arabinofuranohydrolases, α-glucuronidases (EC 3.2.1.139), feruloyl esterases, and p-coumaroyl esterases (Figure 2B) (Schuster & Chinn 2013). Whereas the cellulose and hemicellulose fraction are polymers of sugars and are therefore an interesting source for fermentation purposes (Sun & Cheng 2002). The lignin fraction is a complex structure which is recalcitrant to enzymatic degradation and therefore often used to create energy by combustion, although efforts on its valorization are ongoing (Palazzolo & Kurina-Sanz 2016; Chen & Wan 2017).

Biorefinery

To satisfy the growing demand for energy and consumables in the future, sustainable production of these goods needs to be realized. Sustainable alternatives for energy production have already been developed such as solar power, wind energy and hydroelectricity. As not only the demand for energy increases but also the demand for chemical products, biomass is likely to become one of the industrially most important feedstocks. It is generally accepted as a sustainable feedstock with a low CO_2 footprint and is locally available. These aspects make it an ideal substrate for future production of fuels and chemicals as building blocks for a wide range of products.

The most sustainable way to convert lignocellulosic feedstocks into value added products is envisioned by the biorefinery concept. The biorefinery concept traditionally consists of five distinct steps: harvest and transport of feedstock, pre-treatment, enzymatic hydrolysis, fermentation and DSP. The first step is the harvest and transport of the feedstock, which can be of various origins such as the paper and textile industry, grasslands, agriculture or hard- and softwoods. While currently most biorefineries make use of the first generation feedstocks such as glucose and starch (Abdel-Rahman et

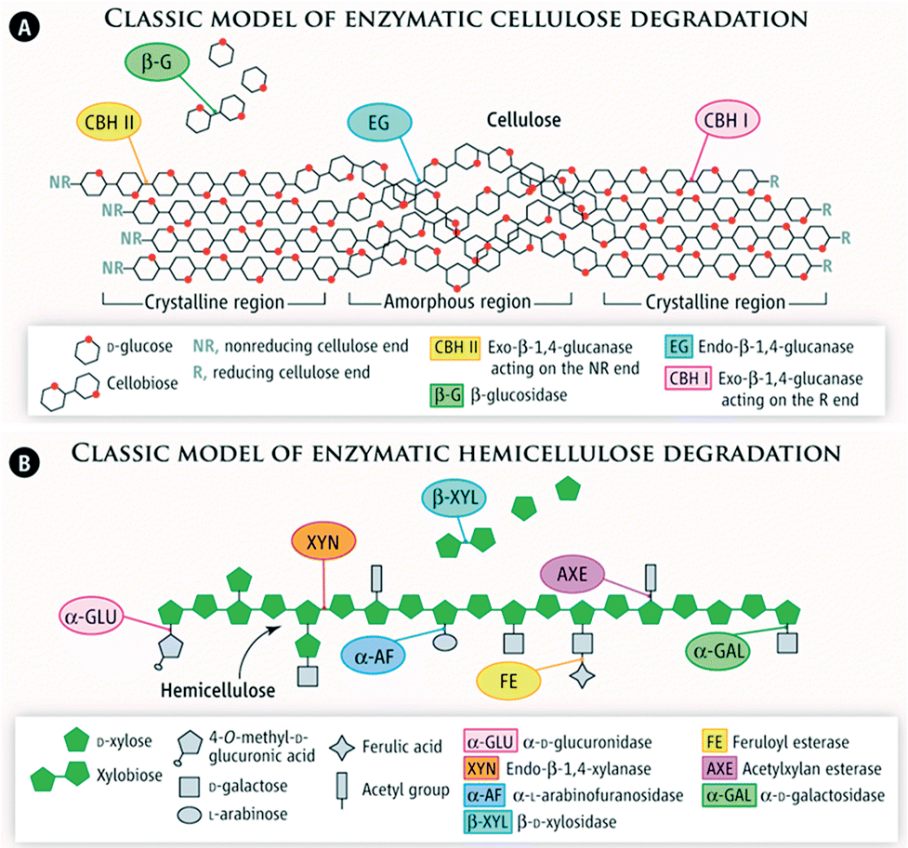


Figure 2 (A) The enzymatic degradation of cellulose includes the joint action of exoglucanases or cellobiohydrolases (CBHs), endoglucanases (EGs), and β-glucosidases. CBHs cleave cellulose in a processive manner, releasing primarily cellobiose from the ends of the cellulose chain. Unlike CBHs, EGs can act on both cellulose and hemicelluloses (Chen et al. 2012). The degradation of cellulose by the joint action of CBHs and EGs is the backbone of today's commercial cellulase products. (B) Hemicellulose degradation also involves the synergistic action of diverse enzyme activities (Goldman 2011). The concerted action of a myriad of hemicellulases efficiently hydrolyzes hemicellulose. The removal of the hemicellulose barrier enhances cellulase activity on biomass cellulose by increasing the accessible cellulose surface area (adapted from Berlin 2013).

al. 2010), current research focusses more on the second-generation feedstocks such as lignocellulose, which allows for more complete use of food crops. The downside to 2nd generation substrates is their recalcitrance to degradation that induces the need to pre-treat the substrate to make it amenable to enzymatic conversion. The second step of the biorefinery concept is the before-mentioned

pre-treatment. This pre-treatment is done using either a physical (*e.g.* milling), or chemical approach. The chemical treatment often consists of a dilute acid or alkali wash step at temperatures above 160°C and high pressures (Murciano Martínez et al. 2015). Especially the (dilute) acid pre-treatment is often used and has the advantage that a wide variety of substrates can be used and the recovery of hemicellulolytic sugars is often high. The main drawback is the possible further degradation of the sugar polymers that might convert to compounds inhibitory to fermentation, like furfural and hydroxymethyl furfural. The alkali treatment results in the removal of lignin making the polysaccharides more amenable to enzymatic hydrolysis. However, acetyl groups are often removed resulting in a lower accessibility of the hemicellulose and cellulose surface resulting in a less efficient enzymatic hydrolysis (Sindhu et al. 2016; Arantes & Saddler 2010). Next to the separate use of physical and chemical treatment, combinations of the two are also applied. A well-known example is steam explosion, in which high-pressure steam is injected to a vessel containing the biomass. By a sudden pressure drop the biomass decompresses at such a force that hemicellulose and lignin are degraded. This technique is sometimes used in combination with low amounts of acid or alkaline chemicals. The complete array of pre-treatment options has been extensively reviewed elsewhere (Xu et al. 2010; Ravindran & Jaiswal 2016; Sun & Cheng 2002). The result of pre-treatments depends strongly on the substrate of choice and the residence time, temperature, particle size and moisture content (Ravindran & Jaiswal 2016; Sun & Cheng 2002; Hendriks & Zeeman 2009). Since different fermentation processes require different sugar compositions, the choice for a certain pre-treatment depends heavily on the process conditions. It is widely accepted that pre-treatment of the substrate is crucial to utilize the full sugar potential of the biomass. Following pre-treatment in the biorefinery, the substrate will undergo enzymatic hydrolysis in which sugar polymers are degraded to oligomeric and subsequently monomeric sugars. The enzymatic hydrolysis is instrumental in high conversion yields from resource to product, but impacts the economics (Gaurav et al. 2017; Zabed et al. 2016). The expensive

enzymes, used in the conversion of lignocellulosic substrates to fermentable sugars, contribute thereby significantly (40-50%) to the production costs (Lee R. Lynd, Paul J. Weimer, Willem H. van Zyl 2002; Kubicek & Kubicek 2016). These enzymes are most often produced using fungal hosts like *Aspergillus niger* and *Trichoderma reesei* and are composed of enzymes which complement each other by different modes of action on the substrate, as described above. The enzymatic hydrolysis of lignocellulose yields fermentable sugars that consist of a (i) C6 fraction, composed mostly of glucose together with some mannose and galactose, (ii) a C_5 sugar fraction consisting of xylose and some arabinose, depending from which source the substrates originates, and (iii) lignin which optionally can be removed through pre-treatment. The hydrolysate obtained is used as substrate for microbial fermentation by mostly mesophilic organisms. These organisms, such as *Escherichia coli, Zymomonas mobilis* and *Saccharomyces cerevisiae*, often have a long history of research and a well-established genetic toolkit, and are selected based on their excellent production rates, yields and industrial robustness. However, despite these clear positive characteristics, they do come with some drawbacks to the biorefinery economics. Since the optimal temperature (50°C) for the hydrolytic enzymes during polymer hydrolysis does not match with the optimum growth and production temperature (30-37°C) of these organisms, a major conflict arises which adds up to the excessive costs related to the biorefinery concept (Ou et al. 2009). The hydrolysate needs to be cooled to the optimum temperature of the organism, requiring energy and time. In addition, the actual fermentation needs constant cooling to maintain this optimum temperature resulting in high energy demands for the process. The last step in the biorefinery comprises the downstream processing (DSP) where fermentation products are purified and waste streams are either reused or exploited as by-products. An example of such a waste stream would be the lignin fraction which might be separated and burned to generate heat. Despite exploiting possible waste streams, the large energy consumption in the fermentation step together with the high costs of hydrolytic enzymes should be reduced to create a

process that is competitive with the current petrochemical industry (Gaurav et al. 2017). Although many studies have been addressed to lower the enzyme dosage, to decrease the enzyme production costs and to increase the activity of enzymes in the hydrolysis of lignocellulosic substrates, the amounts and costs of enzymes needed are still substantial and a main bottleneck to the competitiveness of biobased products towards petrochemically derived products (Tomás-Pejó et al. 2008; Xiao et al. 2016; Jang et al. 2012).

Lactic acid

Lactic acid (LA) is one of the many interesting products that can be produced using a biorefinery approach. Lactic acid has as wide range of applications in the food and chemical industries (Elvers et al. 2016). Well-known is the role of lactic acid in the yoghurt and cheese production process, where LA is the product of fermentation by various lactic acid bacteria (LAB). Nowadays, lactic acid is increasingly exploited as a building block for polylactic acid (PLA), which can be used for the production of bioplastics (Becker et al. 2015). In contrast to the chemical synthesis route of lactic acid, in which usually a racemic mixture of D- and L-lactic acid is formed, the fermentative route can be used to produce optically pure lactic acid of both isomeric forms. The optical purity is of importance in the further processing of lactic acid since the strength of the polymers is highly influenced by the distribution of D- and L-lactic acid monomers. A racemic mixture results in an amorphous polymer and mixtures of various ratios are of interest in medicine research as drug delivery agents. However, PLA is mostly used as food packaging material in which heat stability and strength of the material are of great importance (Okano et al. 2009; Meng et al. 2012). Although the lactic acid and PLA markets grow annually with approximately 15% and 19% respectively, the production capacity of 714 kilo tons for lactic acid and 361 kilo tons for PLA is just a fraction of the 300 mega tons per year of total plastics produced, although not all plastics can be replaced by lactic acid derived plastics (PlasticsEurope 2016); http://www.grandviewresearch.com/press-release/global-lactic-acid-and-poly-lactic-acid-market). The discrepancy in production volume is

largely due to the bottlenecks discussed in the previous paragraphs. The high costs for raw substrates and their pre-treatment, the high costs for enzymatic hydrolysis and the high costs for separation in downstream processing (DSP) are the most important drawbacks of the fermentative production of lactic acid (Abdel-Rahman & Sonomoto 2016).

Biorefinery process integrations

Several variants on the biorefinery concept have been proposed to create a solution to the economic hurdle in the production of biobased products. Between these variants, a clear distinction can be made based on the extent of integration of subsequent steps (Figure 3). Where the least integrated variant is still mostly used, much research is now focusing on the more integrated variants like simultaneous saccharification and fermentation (SSF) and simultaneous saccharification and co-fermentation SSCF. These two variants, although different, are most often both named SSF. In SSF, the cellulose hydrolysis and fermentation occur in one step. This implies that the optima of the enzymes used and the fermentation organism need to match each other during the process to achieve maximum yield. Research has focused on the use of classical fermentation organisms to produce ethanol and chemicals in an SSF setup. However, in the last decade thermophilic organisms have gained much interest because of some key advantages they possess over mesophilic organisms. Thermophiles have temperature optima between 55-80°C, much more compatible with the optimal temperatures that are currently used for enzymatic hydrolysis during SSF. Next to the higher activity of the enzymes, which results in more efficient hydrolysis, less cooling is required for substrates and fermentation processes gaining a major energy benefit. Using *Bacillus coagulans*, which has an optimum growth temperature of 50°C, at a pH of 5.0 the process time and enzyme load was reduced three times compared to a similar process using mesophilic temperatures (Ou et al. 2009). Additionally, the risk of contamination is reduced and end-product inhibition of the enzymatic hydrolysis, due to increasing monomeric sugar concentrations, is circumvented

by the direct use of the sugars by the organism in an SSF-like setup. Furthermore, most chemical reactions are accelerated at higher temperatures and product solubility increases while viscosity is lowered. Succeeding SSF, the more integrated variant, Consolidated Bio-Processing, or CBP, is thought to have the highest economic viability of all biorefinery process setups. The costs of CBP are much lower than those of SSF because of an estimated eight time reduction in costs for biological conversion (Lee R. Lynd, Paul J. Weimer, Willem H. van Zyl 2002; Lynd et al. 2005). Overall, the cost savings would be more than fourfold lower than SSF assuming the lowest reported cellulase production price. Cost savings are mostly due to lower costs for capital expenditures, raw materials and other utilities related to cellulase production. However, the impact of the different cost saving factors is always dependent on the process it is compared to (Olson et al. 2012). In the last decade, much progress has been made in the field of CBP. Not only in applied technology, but certainly also in the field of fundamental research on enzymatic mode of actions and the interactions of enzymes with substrates. Although substantial progress has been made in several aspects of the process, an industrially relevant production organism for the CBP configuration has still to be developed.

Thermophilic CBP production organisms

The ideal organism for CBP processes should meet certain characteristics that have not yet been found altogether in a single organism. Next to the fact that it should be thermophilic, it should be able to completely degrade the (pre-treated) substrate and be able to convert the released sugars with a yield and productivity competitive with current industrial strains. Currently used industrial strains like *E. coli S. cerevisiae* and *Z. mobilis* are not able to ferment both C6 and C_5 sugars, which results in a major bottleneck for optimal usage of the substrate. Although these strains have been engineered to ferment both C6 and C_5 sugars, carbon catabolite repression (CCR) remains to be a big hurdle for efficient simultaneous fermentation. In the development of CBP organisms two main strategies have been proposed being the native strategy

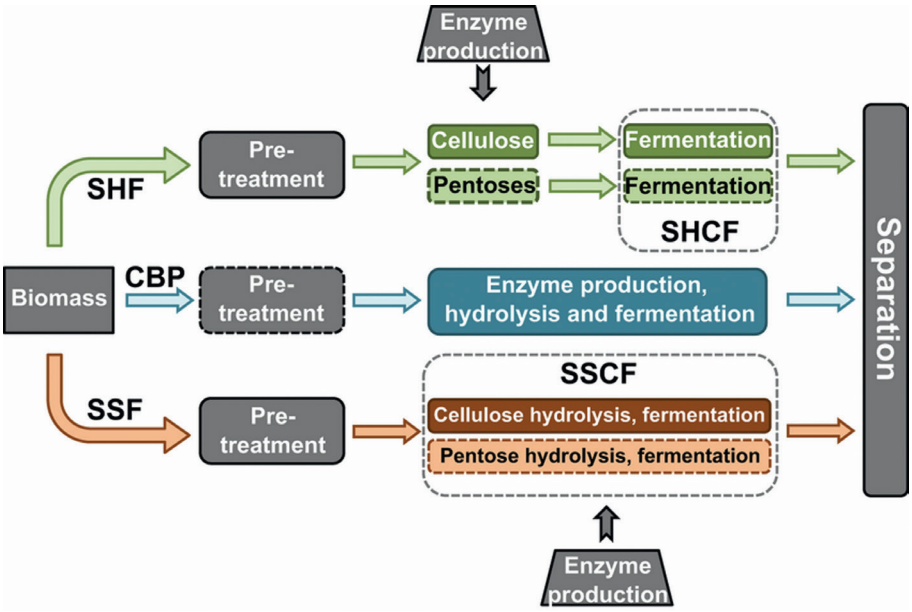


Figure 3 Strategies for the use of second-generation biomass as feedstock to produce biochemicals. Simultaneous hydrolysis and fermentation (SHF) (in green) is a two-stage process that requires a pre-treatment step to reduce recalcitrance after which the solids (cellulose and lignin) will be hydrolyzed separate from the liquid phase (mostly pentoses). Dashes around pentose hydrolysis and fermentation indicate that this step is not always included in the SHF strategy. Resulting hydrolysate will then be fermented by hexose- or pentose-fermenting organisms or co-fermented by a single organism (simultaneous hydrolysis and co-fermentation or SHcF). SSF is a one-stage process which reduces the volume required to produce comparable amounts of biofuel in comparison with SHF. Further efficiency can be gained using a microorganism capable of co-fermentation of hexoses and pentoses from the hydrolysate (SScF). In both SHF and SSF strategies, off- or on-site production of plant-biomass-deconstructing enzymes is required for hydrolysis, incurring an additional economic cost to biofuel production. A process designed to combat the economic costs of separate enzyme production is consolidated bioprocessing (CBP), which consolidates enzyme production, hydrolysis, and fermentation in a single-stage process. In this strategy, a single organism or community of microorganisms will produce plant-biomass-deconstructing enzymes that will release oligosaccharides that are then imported by the CBP microorganism(s) and co-fermented. (Blumer-Schuette et al. 2014)

and the recombinant strategy (Zheng et al. 2013; Mazzoli 2012; Lee R. Lynd, Paul J. Weimer, Willem H. van Zyl 2002). Whereas the native strategy aims at engineering naturally cellulolytic organisms into production strains, the recombinant strategy is more focused on the introduction of cellulolytic traits in industrially used production strains. Thermophiles on the other hand are a relatively unexplored niche of microbes and gained increased interest as CBP candidates

and as a potential source of robust hydrolytic enzymes. A substantial effort has been made to understand and improve the mode of action of several thermophilic cellulolytic species with *Clostridium thermocellum* as the most studied organism. The anaerobic, gram-positive *C. thermocellum* produces cellulosomes consisting of a large scaffolding protein, a non-catalytic subunit containing multiple cohesin modules. Various enzymatic subunits are able to bind to the cohesin modules by their dockerin module and in this way a variety of cellulolytic and hemicellulolytic enzymes can bind to the scaffoldin protein, together forming the cellulosome complex. Cellulosomes are typically found in anaerobic environments where organisms benefit from the proximity of the degradation products to the cell. By strictly regulating the expression of different modules on the cellulosome, these organisms are thought to create an energy efficient way of hydrolyzing and metabolizing sugars. The interest in *C. thermocellum* as a host for CBP processes is based on its excellent cellulose degrading capacities. Besides degrading cellulose and starch, it can readily ferment oligosaccharides like sucrose, maltose and raffinose and monosaccharides such as glucose, galactose, mannose, fructose and arabinose. The wide range of sugars it can ferment and the fast hydrolysis rate approaching 2.5g/L/h make this organism a very interesting candidate (Argyros et al. 2011).Although *C. thermocellum* is an efficient cellulose degrader, it cannot metabolize hemicellulose. To fully degrade the lignocellulosic sugars, pathways for C_5 sugar metabolism should be introduced or it should be combined in a co-culture with a hemicellulolytic organism to fully exploit the substrates potential. *C. thermocellum* produces ethanol as its main fermentation product, but concentrations higher than 30 g/L^{-1} have not been reported and concentrations above 80 g/L^{-1} are not tolerated by the organism (Blumer-Schuette et al. 2014).

Thermoanaerobacterium saccharolyticum is also being developed as a host for CBP (Joe Shaw et al. 2008; Tsakraklides et al. 2012). It grows between 45°C and 65°C and was demonstrated to degrade oat spelt xylan directly to ethanol, yielding 1.75 g/L ethanol from 10 g/L substrate (Table 1) (Lee et al. 1993). Comparable results on xylan were obtained with *T. aotearoense* (Yang et al. 2013). Although both organisms are

not cellulolytic, they are capable of fermenting cellobiose to ethanol (Table 1). Genetic modifications in *T. saccharolyticum* have largely been addressed to elevate ethanol productivities by knocking out organic acid pathways, thereby creating a mutant that produces almost exclusively ethanol (Desai et al. 2004; Shaw et al. 2009). *T. aotearoense* was engineered to produce mainly lactate by blocking the acetic acid pathway, resulting in a strain that produces up to 3 g/L lactic acid directly from xylan (Yang et al. 2013). However, despite numerous attempts, the heterologous expression of active cellulases has not yet been achieved in these *Thermoanaerobacterium* species (Currie et al. 2013).

Caldicellulosiruptor bescii is an extreme thermophile (T_{opt}= 75 °C) capable of degrading a wide variety of polysaccharides, including cellulose and hemicellulose. In contrast to the organisms described above, *C. bescii* is not capable of ethanol production as they lack the bifunctional alcohol and aldehyde dehydrogenase encoding gene *adhE*. Efforts providing *C. bescii* with a functional ethanol production pathway resulted in an ethanol producing strain, although production did not exceed 33% of the theoretical maximum (Chung et al. 2014). The main fermentation product of *C. bescii* is hydrogen, followed by near equal amounts of lactic acid and acetate (Cha et al. 2013). No attempts to direct metabolic fluxes towards lactic acid production are reported and genetic modification in *C. bescii* is challenging, as multiple restriction enzymes prevent plasmid replication and thereby the stable integration of DNA fragments into the genome (Zeldes et al. 2015).

In recent years the genus *Geobacillus* gained increased interest for the production of ethanol (Cripps et al. 2009; Taylor et al. 2009), as a source of thermostable enzymes (Ban et al. 2016; Dokuzparmak et al. 2016; Huang et al. 2016; Nisha & Satyanarayana 2016), and for its ability to degrade complex xylans (Bhalla et al. 2014; Canakci et al. 2012; Verma et al. 2013; Gerasimova & Kuisiene 2012). The genus *Geobacillus* contains Gram-positive, spore forming bacteria that optimally grow between 40-75°C (Nazina et al. 2001).

Table 1 Summary of the performance of several candidate CBP organisms on cellulose and hemicellulose

Product	Species and strain	Genotype	(hemi) cellulolytic	T °C	pH	Medium additions (g/L)	Substrate type	Substrate g/L	Titer g/L	Time (h)	Av. prod g/L/h	Reference
Lactate	*T. aotearoense*	Δ*pta /ack*	Xylanolytic	55	6.5	MS	Xylan	15.0	3.00	48	0.060	(Yang et al. 2013)
Acetate	*C. bescii*	Δ*ldh/adhE*$^{+}$	Cellulolytic and xylanolytic	72	6.7	0.5g YE	Avicel	20.0	0.65	120	0.005	(Chung et al. 2014)
Ethanol	*C. thermocellum*	ev. Δ*ldh*, Δ*hpt*, Δ*pta*	Cellulolytic	55	7.0	2g YE	Avicel	19.5	5.61	72	0.080	(Argyros et al. 2011)
Ethanol	*C. thermocellum and T. saccharolyticum coculture*	ev. Δ*ldh*, Δ*hpt*, Δ*pta*	Cellulolytic and xylanolytic	55	6.3	2g YE	Avicel	92.2	38.10	146	0.260	(Argyros et al. 2011)
Ethanol	*G. thermoglucosidasius*	Δ*ldhL, pdh up,Δpdf*	-	60	6.7	10g YE	Cellobiose	30.8	14.48	6.1	2.370	(Cripps et al. 2009)
Ethanol	*T. saccharolyticum*	wt	Xylanolytic	60	6.7	3g YE and 10g Tryptone	Xylan (autoclaved oat spelt)	10.0	1.29	11	0.110	(Lee et al. 1993)

Geobacilli have been isolated all over the world, from the heights of the Andes at 3653 m to the depths of ocean sediments at 10.897 m below sea level (Marchant et al. 2008; Takami et al. 2004) and isolates have been found on all seven continents (Zeigler 2014). *Geobacillus* isolates are often found in relatively warm environments like hot springs, geothermal soil, hydrothermal vents and self-heated compost (Wang et al. 2006; Pinzón-Martínez et al. 2010; DeFlaun et al. 2007). The abundancy of geobacilli in compost has led to research on the role of this genus in this complex microbial ecosystem (Ryckeboer et al. 2003; Takaku et al. 2006). Most work in this regard has been done by Shulami et al., who characterized the xylanolytic system of *G. stearothermophilus* T-6, isolated from soil (Shoham et al. 1992).

Geobacillus hemicellulose degradation

The xylanolytic system reveals a highly complex genomic island, containing multiple gene clusters encoding for a variety of hydrolytic activities and transport mechanisms (Shulami et al. 2014) (Figure 4). Characteristic for *Geobacillus* is the relative small amount of secreted enzymes that are able to cleave the backbone of several hemicellulosic structures to smaller oligosaccharides (Shulami et al. 2011; Shulami et al. 1999; Shulami et al. 2014). These oligosaccharides can then enter the cell through specific ABC transporters, after which these oligosaccharides are further processed intracellularly. The ability of *G. stearothermophilus* T-6 to take up oligosaccharides is a potential advantage over competing organisms that are not able to take up these products. By sensing residual amounts of free xylose though the histidine sensor kinase protein XynD, the response regulator XynC is phosphorylated and subsequently activates the expression of genes encoding the oligosaccharide ABC transporter XynEFG. Xylose not only activates the expression of oligosaccharide transport, it also interacts intracellularly with the master regulator XylR. XylR acts as a repressor on six transcriptional units, including the *xynX-axe2-xynB3* cluster encoding intracellular oligosaccharide degradation and the extracellular endoxylanase Xyn10A. The latter enzyme, encoded by the gene *xynA*, is the hallmark of the system, and without this enzyme the organism would not be able to degrade extracellular xylans.

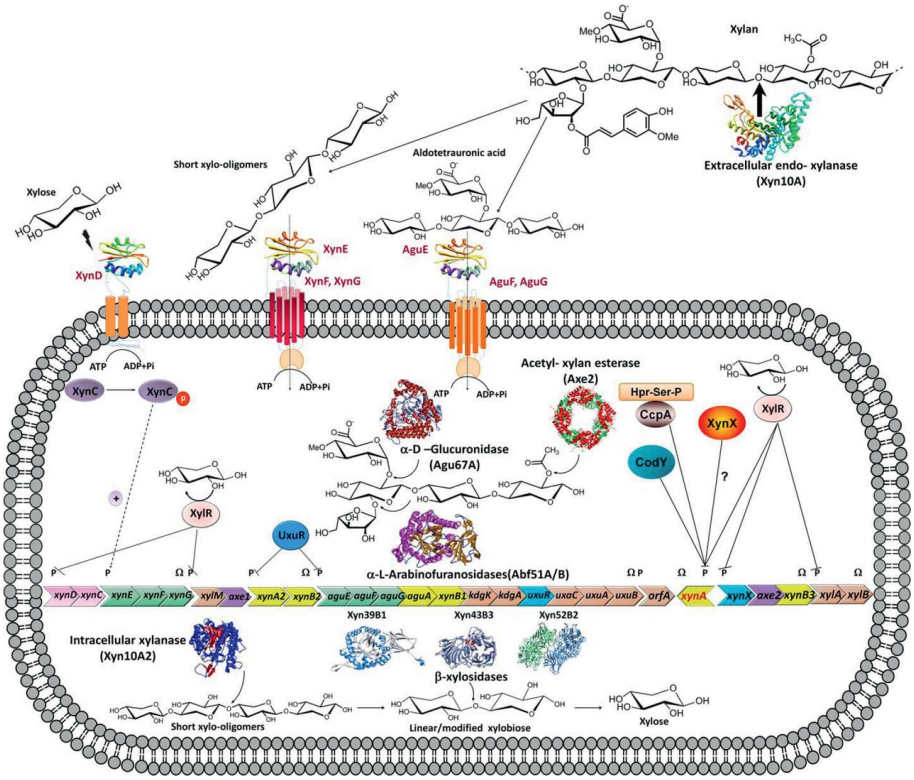


Figure 4 Schematic view of the xylanolytic system from *G. stearothermophilus* T-6**.** The utilization of xylan by *G. stearothermophilus* is initiated by residual levels of xylose or xylo-oligosaccharides present in the environment. These sugars interact with the extracellular domain of the class I histidine kinase sensor protein XynD, which triggers phosphorylation of the response regulator XynC (**21**). The phosphorylated XynC activates the expression of a dedicated ABC xylo-oligosaccharides transporter, XynEFG, which facilitated the entrance of xylosaccharides into the cell. Inside the cell, xylose serves as the molecular inducer and interacts with the master repressor XylR, which negatively regulates five transcriptional units. The expression of the extracellular xylanase provides the cell with ample amounts of decorated xylo-oligosaccharides that enter the cell via two dedicated ABC sugar transporters, XynEFG for xylooligosaccharides and AguEFG for aldotetrauronic acid. The decorated xylooligomers are hydrolyzed to their corresponding monomers by intracellular side chain-cleaving enzymes, including α-glucuronidase (Agu67A), two α-l-arabinofuranosidases (Abf51A and Abf51B), two xylan acetylesterases (CE4), an intracellular xylanase (Xyn10A), and three β-xylosidases (Xyn39B, Xyn52B2, and Xyn43B3) . The extracellular xylanase gene (*xynA*) is subjected to carbon catabolite repression mediated by the global negative repressor CcpA and most likely by CodY. Furthermore, *xynA* is also regulated by XynX and cell density by uncharacterized mechanisms. (Shulami et al. 2014).

In addition to the hemicellulose utilization (HUS) locus depicted in Figure 4, several other clusters are located on the 5' end of the locus including an arabinan degradation cluster, arabinose

transport and metabolism clusters and the cluster encoding for the α-L-arabinofuranosidase (Shulami et al. 2011; De Maayer et al. 2014). Recently, the HUS loci of 24 sequenced *Geobacillus* genomes were compared for their genetic variation and possible detection of novel enzyme activities (De Maayer et al. 2014) (Figure 5). It appeared that the HUS locus is a common feature among *Geobacillus* species, although great genetic variation was detected among their loci, like the absence of the α-glucuronidase and its associated transport system (*aguEFG*) and the lack of an arabinan degradation and transport cluster in all three *G. thermodenitrificans* strains. This variation might imply that geobacilli adapt to the substrate available, possibly through the action of transposases which are detected frequently in the HUS loci (black arrows, Figure 5). The *xynA* gene, encoding the extracellular endoxylanase, is (partially) absent in multiple strains, suggesting that these strains lack the ability to degrade extracellular xylans. Strains HTA426, B23 and CCB-US3F, which all lack the *xynA* gene, are also deficient in many other xylanolytic clusters and seem only able to transport and metabolize arabinose and xylose. Strains Y41MC1 and M10EXG seem to have lost also the xylose and arabinose clusters, and have thereby lost all functionality of the HUS locus except for an oligopeptide transport system. The wide diversity among the HUS loci of geobacilli gives ample opportunity to select for strains capable of degrading a specific substrate, and thereby to select for strains with desired characteristics.

Geobacillus cellulose degradation

The ability of *Geobacillus* to degrade cellulose has been reported in several studies (Rastogi et al. 2010; Assareh et al. 2012; Makky 2009; Ng et al. 2009). *Geobacillus* sp. T1 was found to produce cellulases on avicel, carboxymethylcellulose (CMC) and wheat straw, although protein identification was not performed (Assareh et al. 2012). Furthermore, strain T1 could not ferment glucose and/or xylose to organic acids and is therefore not a suitable candidate for CBP. Another strain, T4, was reported to ferment glucose and xylose to organic acids (Tai et al. 2004). This strain was isolated from sugar refinery wastewater and cellulase activity in the supernatant of liquid cultures was

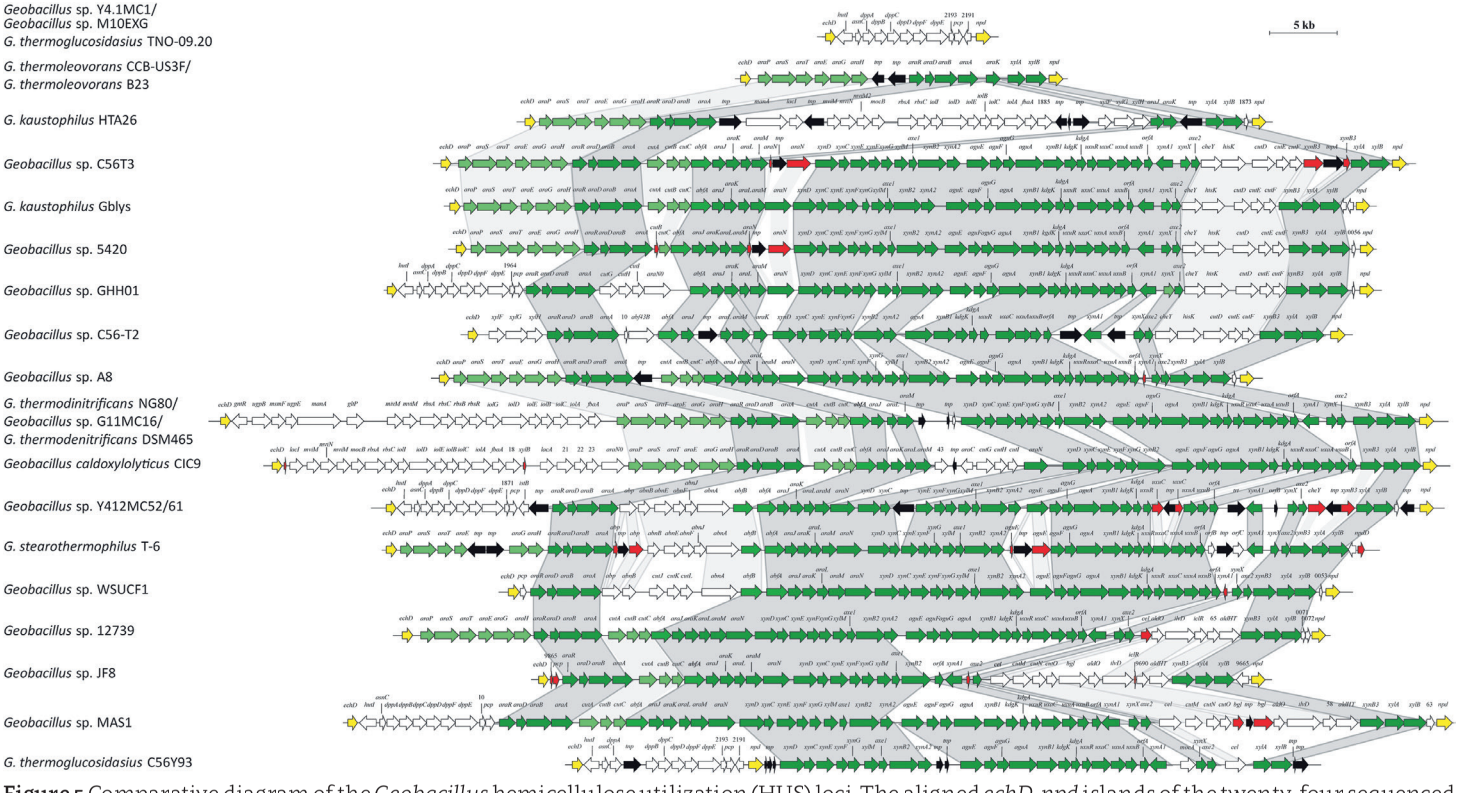


Figure 5 Comparative diagram of the *Geobacillus* hemicellulose utilization (HUS) loci. The aligned *echD-npd* islands of the twenty-four sequenced *HUS⁺ and HUS-Geobacillus* strains and that of G. *stearothermophilus* T-6 are shown. The flanking *echD and npd* genes are colored in yellow, genes coding for transposons in black, and genes in which the reading frames are transposon-disrupted are in red. Those genes conserved among >70% of the HUS⁺ strains are colored in green, those conserved among >50% and <70% in light green, while those conserved among <50% of the HUS⁺ strains are colored in white. (De Maayer et al. 2014)

demonstrated for CMC and avicel. Although its cellulolytic activity was already reported in 2004, there has been no report of the identified cellulase(s) to date. One of the strains whose genome has been sequenced, WSUCF1, was also reported to be cellulolytic (Rastogi et al. 2010). The compost isolate WSUCF1 showed the highest cellulase activity on crystalline cellulose, measured after six days of incubation. However, no fermentation products were detected indicating that its cellulolytic activity is not at par for efficient cellulose degradation. Again, no identification of its cellulolytic proteins was performed. Taken together the above, it seems like geobacilli occasionally do encode cellulases, although the reported activities are low and insufficient for organic acid production directly from cellulose. To enhance the cellulolytic activity, several attempts on the heterologous expression of cellulases have been taken (Suzuki et al. 2013; Bartosiak-Jentys et al. 2013). The heterologous expression of Cel5A, an endoglucanase from the thermophilic bacterium *Thermotoga maritima*, resulted in a *G. thermoglucosidans* strain expressing active endoglucanases, although at very low yields as no protein could be visualized by SDS-PAGE (Bartosiak-Jentys et al. 2013). The expression of the catalytic domain of a GH48 exoglucanase in the same strain yielded only marginal activity. Another attempt was performed with *G. kaustophilus* HTA426, in which the expression of a heterologous endoglucanase (WP_010885255.1) from the thermophilic archaeon *Pyrococcus horikoshii* rendered this strain capable of CMC degradation (Suzuki et al. 2013). Although both studies describe the successful introduction of endoglucanases, they do not report the production of fermentation products from cellulose. The successful genetic modification of most *Geobacillus* spp. is hindered by the lack of genetic tools, as we will describe in more detail below. As only a handful of plasmids and promoters are known, the options to optimize expression constructs are limited (Bartosiak-Jentys et al. 2012; Kananavičiūtė & Čitavičius 2015).

The genetic toolbox of *Geobacillus* spp.

Engineering of *Geobacillus* spp. was achieved for several strains using protoplast fusion (Liao & Kanikula 1990; Imanaka et al. 1982;

De Rossi et al. 1994; Wu & Welker 1989). However, Narumi et al. successfully applied electroporation to *G. thermodenitrificans* K1041 (previously known as *B. stearothermophilus*) and successfully introduced several plasmids to this strain (Narumi et al. 1992; Narumi et al. 1993; Nakayama et al. 1992). Electroporation is a method far less tedious than protoplast fusion, and is therefore the preferred method for transformation. The successful electroporation of *G. thermoglucosidans* TN, DL33 and derivatives meant a breakthrough in *Geobacillus* engineering (Thompson et al. 2008; Taylor et al. 2008). By creating an *ldh* gene mutant, Cripps et al. (2009) managed to shift the metabolic pathway from producing mainly lactic acid to mainly ethanol. Besides the achieved metabolic shift, the underlying techniques provided the first example of reliable transformation and engineering of *Geobacillus* species. Previously reported problems concerning tedious transformation protocols like protoplast transformation (Couñago & Shamoo 2005; Wu & Welker 1989) and low transformation efficiencies were circumvented and multiple cloning and integration vectors were developed. The previously used Gram-positive vector pUB110, containing a kanamycin resistance gene, supports growth up to 55°C, although it becomes more unstable at temperatures close to the upper limit. More stable vectors were developed with a replicon from pBST1, a plasmid isolated from *G. stearothermophilus* NRTL 1102 (Taylor et al. 2008). One such plasmid, pBST22, consist of a pUC19 backbone with kanamycin and chloramphenicol resistance markers stable up to 68°C, although its size of 7.6 kb limits transformation efficiencies (Kananavičiūtė & Čitavičius 2015). To increase transformation efficiencies, the 3.8 kb pUCG3.8 vector was created (Bartosiak-Jentys et al. 2013). This vector contains a kanamycin resistance marker, but lacks chloramphenicol and ampicillin resistance genes, thereby reducing its size. Vector pNW33n (3.9 kb), also based on the pUC19 backbone, contains the chloramphenicol resistance marker instead of the kanamycin marker and is stable at 60 °C (Zeigler 2014). Furthermore, a temperature sensitive suicide vector, pTMO31, was developed based on the pUB110 vector, specifically for genome integrations at temperatures above 55°C.

Various strains have been engineered to overexpress (heterologous) proteins (Lin et al. 2014; Suzuki et al. 2013; Bartosiak-Jentys et al. 2012). However, the variety of suitable promoters for *Geobacillus* spp. is limited, and characterization of new promoters is ongoing. The P_{sigA} promoter, found just upstream of two housekeeping genes, was successfully applied in the constitutive expression of α-amylase and β-galactosidase genes in *G. kaustophilus* HTA426 (Suzuki et al. 2012). Another constitutive promoter, $P_{RHIII,}$ based on the ribonuclease H III gene of *G. stearothermophilus* NUB3621, was used to express a superfolding green fluorescent protein (*sfGFP*) (Pédelacq et al. 2006). Bartosiak-Jentys et al. created P_{up2n38}, a constitutive promoter based on the upstream region of the uracil phosphoribosyltransferase gene of *G. thermoglucosidans*. Under the tested conditions, this promoter was demonstrated to be five times more active than the cellobiose induced promoter $P_{\beta glu}$. The latter promoter was obtained from the cellobiose-specific PTS operon of *G. thermoglucosidans* NCIMB 11955 and is inducible by cellobiose but was also found to be induced by glucose and xylose (Bartosiak-Jentys et al. 2013). The lactate dehydrogenase promoter P_{ldh}, isolated from *G. stearothermophilus* NCA1503 and *G. thermodenitrificans* DSM465 is considered the strongest promoter to date, although its activity seems to be highest under micro-aerobic conditions.

In the case of protein secretion, signal peptides are needed that are recognized by the host strain. The overexpression and secretion of α-amylase and β-galactosidase genes in *G. kaustophilus* HTA426 yielded active extracellular proteins by using heterologous signal peptides, although cellulase activity was also detected in the supernatant of a strain that expressed cellulases without signal peptide. Other studies have shown secretion of heterologous proteins in various strains of *Geobacillus*, mediated by a native signal peptide from the extracellular endoxylanase (Suzuki et al. 2013; Bartosiak-Jentys et al. 2013).

The ideal CBP organism

Research on thermophiles is still in its infancy, especially when compared to mesophiles like *E. coli* and *S. cerevisiae*. However, the

benefits of thermophiles for CBP are undisputed, and the search for the perfect host is ongoing. To date, no single organism isolated or engineered, is capable of efficiently hydrolyzing both hemicellulose and cellulose whilst converting the released sugars to desired end-product. As described in this chapter, thermophiles often are cellulolytic or hemicellulolytic, but are rarely able to hydrolyze both. Where the degradation of hemicellulose requires a large variety of different enzymatic actions, the degradation of cellulose is rather simple. Therefore, the aim of the project described in this thesis was to obtain a lactic acid producing isolate, that is at least hemicellulolytic, and preferably also cellulolytic.

Thesis outline

As described in **Chapter 1**, the conversion of lignocellulosic substrates to desired products is already possible, but not economically competitive with the petroleum-based industry. To overcome this economic hurdle, more integrated and efficient processes need to be developed like consolidated bioprocessing (CBP). A prerequisite for CBP is a microbial host capable of degrading both the hemicellulose and cellulose fraction of lignocellulose, and capable of subsequent fermentation of the released sugars to desired end-product. Out of several described thermophiles, *Geobacillus* poses the greatest opportunity to be developed as a suitable host for CBP and will therefore be the focus in this thesis.

Chapter 2 describes the isolation of *Geobacillus* spp. from compost based on growth on xylan and cellulose and the selection of the most suitable candidate for further research. We describe selection based on the fermentation profiles of the obtained isolates, as well as on hydrolytic activity against xylan and cellulose. We also describe the optimization of genetic accessibility for the selected isolate, *G. thermodenitrificans* T12, and demonstrate its potential for further development by heterologous gene expression. Strain T12 is further characterized for its ability to ferment glucose, xylose and xylan to lactic acid in different media.

Chapter 3 includes the further characterization of *G. thermodenitrificans* T12 based on genome analysis. Several interesting features are described that demonstrate the potential of strain T12 as a degrader of cellulose and hemicellulose. Of interest in this respect is the extended HUS locus of strain T12, including a novel pectate degrading pathway not previously described for *Geobacillus*. We also describe the host-defense mechanisms identified on the genome, with a CRISPR/Cas Type II system being the most notable feature. This Cas9 based CRISPR system is the first of its kind described for thermophiles, and might lead to a significant expansion of the *Geobacillus* genetic toolbox.

Chapter 4 is dedicated to the characterization of *Gt*XynA1, the extracellular endoxylanase produced by T12. Earlier studies describing orthologues of this enzyme focused on its temperature and pH optima but lacked an in-depth analysis of its degradation products. Therefore, we analyzed its mode of action on several xylans and provided a model for its mode of action. Substituents on the xylan backbone hindering the endoxylanase activity are demonstrated and give a unique insight in its role in xylan degradation by *Geobacillus*.

Chapter 5 describes three different approaches to complement the cellulolytic machinery of strain T12. A *Geobacillus* metagenome was screened for potential cellulases and two identified candidate genes were expressed and characterized. One of these was demonstrated to have activity against amorphous cellulose and was successfully expressed in strain T12. In addition, we identified more diversity between the *G. thermodenitrificans* HUS loci of strain T12 and strain T81, derived from the metagenome. Furthermore, earlier described *Geobacillus* endoglucanase encoding genes (*celA*, *M1* and *M2*) were introduced to T12. We discuss the potential misannotation of M1 and M2 and demonstrate the activity of CelA when produced by T12. Lastly, we describe the heterologous expression of a *C. thermocellum* derived endoglucanase and two exoglucanases in strain T12, and demonstrate the production of an active exoglucanase.

Chapter 6 contains the general discussion on this thesis, and puts the work described in a wider context and discusses its value to the field of *Geobacillus* research. Recent developments in genetic engineering and the expansion of the genetic toolbox for *Geobacillus* are described, along with a view on future directions. Furthermore, the genetic variety of the HUS locus is discussed, and some of our insights from the metagenome study are taken along to show the surprising diversity among the HUS loci of *G. thermodenitrificans*.

B. subtilis DSM 402
0.01
G. thermodenitrificans T22
G. thermodenitrificans T39
G. thermodenitrificans T76
G. thermodenitrificans T68
G. thermodenitrificans T3
G. thermodenitrificans T19
G. thermodenitrificans T11
G. thermodenitrificans T28
G. thermodenitrificans T54
G. thermodenitrificans T17
G. thermodenitrificans T73
G. thermodenitrificans T12
G. thermodenitrificans T29
G. caldoxylosilyticus T93
G. caldoxylosilyticus T37
G. caldoxylosilyticus T31
G. caldoxylosilyticus T32
G. caldoxylosilyticus T86
G. caldoxylosilyticus T92
G. caldoxylosilyticus T66
G. caldoxylosilyticus T90
G. caldoxylosilyticus DSM 12041
G. thermoglucosidans T89
G. thermoglucosidans T64
G. thermoglucosidans T63
G. thermoglucosidans T61
G. thermoglucosidans T91
G. thermoglucosidans T52
G. thermoglucosidans T67
G. thermoglucosidans T33
G. thermoglucosidans T57
G. thermoglucosidans T60
G. thermoglucosidans T58
G. thermoglucosidans DSM 2542
G. thermoglucosidans T88

Chapter 2

Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost

Martinus J.A. Daas[1] , Antonius H.P. van de Weijer[1] , Willem M. de Vos[1], John van der Oost[1] and Richard van Kranenburg [1,2]

Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, the Netherlands[1]; Corbion, Arkselsedijk 46, 4206 AC, Gorinchem, the Netherlands[2].

This chapter was adapted from:
Daas, M. J. A., Van de Weijer, A. H. P., De Vos, W. M., Van der Oost, J. & Van Kranenburg, R. Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. *Biotechnol. Biofuels* 9, (2016).

Abstract

Background: Due to the finite nature of global oil resources we are now faced with the challenge of finding renewable resources to produce fuels and chemicals in the future. Lactic acid has enormous potential as a precursor to produce bioplastics alternatives to conventional plastics. Efficient lactic acid fermentation from non-food lignocellulosic substrates requires pretreatment and saccharification to generate fermentable sugars. A fermentation process that requires little to no enzyme additions, *i.e.* consolidated bio-processing, would be preferred and requires lactic acid-producing organisms that have cellulolytic and/or hemicellulolytic activity.

Results: To obtain candidate production strains we have enriched and isolated facultative anaerobic (hemi)cellulolytic bacterial strains from compost samples. By selecting for growth on both cellulose and xylan, 94 *Geobacillus* strains were isolated. Subsequent screening for lactic acid production was carried out from C6 and C5 sugar fermentations and a selection of the best lactic acid producers was made. The denitrifying *Geobacillus thermodenitrificans* T12 was selected for further research and was rendered genetically accessible. In fermentations on a mixture of glucose and xylose, a total of 20.3 g of lactic acid was produced with a yield of 0.94 g product/g sugar consumed. In addition, strain T12 is capable of direct conversion of beechwood xylan to mainly lactic acid in minimal media.

Conclusions: We have demonstrated that *Geobacillus thermodenitrificans* T12 is genetically accessible and produces lactic acid as its main fermentation product on glucose, xylose and a mixture thereof. Strain T12 was additionally used for the direct conversion of xylan to lactic acid. The genetic accessibility of the T12 strain provides a solid basis for the development of this strain into a host for consolidated bioprocessing of biomass to lactic acid.

Background

The increasing consciousness regarding the sustainability of our current life standards has led to the accelerating development of alternative production strategies for fuels, energy and chemicals (Beauprez et al. 2010). Lactic acid is an organic acid that can be used as building block for poly lactic acid (PLA) (Abdel-Rahman et al. 2011). Petrochemically-produced lactic acid always yields a racemic mixture of both D- and L-lactic acid which results in a lower thermostability of the PLA polymer in comparison to its optically pure counterpart derived from microbial production. Microbial production of lactic acid currently dominates the chemical synthesis alternative; however, the efficiency largely depends on the substrate, microbe and mode of production. In recent years, the use of lignocellulosic or non-edible biomass as a resource has gained much interest as it does not compete with the food and feed industry. However, use of this type of substrate requires costly saccharolytic enzymes for successful conversion to fermentable sugars (Davis et al. 2013). Consolidated bioprocessing (CBP) is a method designed to exclude such enzymes and thus optimize lignocellulosic conversion in the most economically feasible way. It is believed that this mode of integration can result in a process four times more cost-efficient even when compared to simultaneous saccharification and fermentation (Lynd et al. 2005).

In the last decades, substantial efforts have been made to isolate and engineer organisms for the conversion of lignocellulosic substrates. The search for suitable saccharolytic organisms even from the most extreme and remote spots on earth has led to the discovery of many novel isolates (Blumer-Schuette et al. 2014; Schuster & Chinn 2013). Among these are isolates from the genus *Geobacillus*, a versatile group of thermophilic facultative anaerobic bacteria (Nazina et al. 2001). Geobacilli are Gram-positive, rod-shaped bacteria with the ability to sporulate. In recent years, this genus has gained much interest not only because of their thermostable enzymes, but also because of their ability to ferment C6 and C5 sugars simultaneously. Furthermore, several *Geobacillus* strains are capable of degrading

xylan or cellulose (Rastogi et al. 2010; Bhalla et al. 2014). Xylanolytic activity is common in this genus and the responsible hemicellulose utilization (HUS) locus, containing the majority of genes involved in the hemicellulose metabolism, has been described in detail elsewhere (De Maayer et al. 2014; Shulami et al. 2014). Cellulolytic activity has been demonstrated only in a few studies (Tai et al. 2004; Rastogi et al. 2009; Ng et al. 2009). In addition, *Geobacillus* strains have been proven to be useful for heterologous expression of a variety of proteins (Cripps et al. 2009; Bartosiak-Jentys et al. 2013). Nevertheless, only limited information on this genus is available, with only 18 complete genome sequences in the NCBI database and only a few strains that have been proven to be readily transformable (Studholme 2015). Moreover, genetic accessibility of (geo)bacilli seems to be strain-specific (Kananavičiute & Čitavičius 2015; Bosma et al. 2015). Because the lack of knowledge of the *Geobacillus* species is one of the main drawbacks to fully exploit members of this genus as platform organisms, more genome sequences, physiological data and improved genetic tools are required.

In this paper, we describe the isolation of novel *Geobacillus* strains of various species with the ability to degrade cellulose and/or xylan. The optimization of transformation and the heterologous expression of a reporter gene were shown for one selected strain. In addition, the production of lactic acid by this strain was demonstrated on glucose, xylose and on beechwood xylan.

Methods

Media and cultivation methods

Cellulolytic Thermophile Vitamin Medium (CTVM; based on Cripps et al. 2009; Bosma et al. 2015; Fong et al. 2006; Sizova et al. 2011) contained per liter: 8.37 g MOPS and salts mix consisting of 1 g NH_4Cl; 3 g NaCl; 1.50 g Na_2SO_4; 0.08 g $NaHCO_3$; 1 g KCl; 1.8 g $MgCl_2 \times 6H_2O$; 0.30 g $CaCl_2 \times 2H_2O$. pH was set to 6.6 at room temperature and the medium was autoclaved for 20 min at 121°C, after which 1 mL K_2HPO_4 (250 g/L; pH 6.6 10 mL filter sterile 100x metal mix and 1 mL filter sterile 1000x vitamin solution were added. 100x metal mix contained per liter: 1.60 g $MnCl_2 \times 6H_2O$; 0.1 g $ZnSO_4$; 0.2 g H_3BO_3; 0.01 g $CuSO_4 \times 5H_2O$; 0.01 g $Na_2MoO_4 \times 2H_20$; 0.1 g $CoCl_2 \times 6H_2O$; 0.7 g $FeSO_4 \times 7H_2O$; 5 g $CaCl_2 \times 2H_2O$; 20 g $MgCl_2 \times 6H_2O$. 1000x vitamin mix contained per liter: 0.1 g thiamine; 0.1 g riboflavin; 0.5 g nicotinic acid; 0.1 g pantothenic acid; 0.5 g pyridoxamine, HCl; 0.5 g pyridoxal, HCl; 0.1 g D-biotin; 0.1 g folic acid; 0.1 g *p*-aminobenzoic acid; 0.1 g cobalamin.

LB2 contains per liter: 10 g tryptone (Oxoid), 5 g yeast extract (Roth), 10 g sodium chloride and salts mix consisting of 1 g NH_4Cl; 3 g NaCl; 1.50 g Na_2SO_4; 0.08 g $NaHCO_3$; 1 g KCl; 1.8 g $MgCl_2 \times 6H_2O$; 0.30 g $CaCl_2 \times 2H_2O$. pH was set to 6.6 at room temperature and the medium was autoclaved for 20 min at 121°C, after which 10 mL K_2HPO_4 (250 g/L) was added.

Minimal Media (MM) contained per liter: 0.52 g K_2HPO_4; 0.23 g KH_2PO_4; 0.5 g NH_4NO_3 (MMy) or 0.3 g NH_4Cl (MMy+). After autoclaving, 1 mL of the following 1,000x concentrated sterile stocks were added: Nitrilotriacetic acid (200 g/L); $MgSO_4 \times 7H_2O$ (145.44 g/L); $CaCl_2 \times 2H_2O$ (133.78 g/L); $FeSO_4 \times 7H_2O$ (11.12 g/L).

For CTVMy/MMy medium, 0.5 g/L yeast extract (Roth) was added to the medium and CTVMy+/MMy+ contains 5 g/L yeast extract (Roth).

Glycerol stocks of cultures were made by adding 500 µl sterilized 60% glycerol to 1.5 mL culture, in a 2 mL cryogenic vial (Corning). Stocks were stored at -80°C.

In all plate and tube cultures, carbon substrates were used in a concentration of 10 g/L unless stated otherwise. Carbon sources were autoclaved separately with xylose being filter sterilized.

For plate cultures, 5 g/L gelrite (Roth) was added. Anaerobic cultivation of plates was done in an anaerobic jar (HP0011, Oxoid) containing an AnaeroGen sachet (AN0035, Oxoid).

All wild-type strains were isolated and cultured at 65°C. Strains harbouring the pNW33n plasmid are always cultures at 55°C to maintain the plasmids replication. Cultures in liquid media are shaken at 150RPM unless stated otherwise.

Sampling

Samples were collected from both a mature and an active compost heap at ReCom Ede (NL). The temperature of the mature compost was around 35°C at the sampling site on top of the compost heap and around 65°C for the active compost heap which was sampled at a depth of 30 cm. Both heaps were semi-anaerobic due to mixing once every week. The sampling site of the 1st isolation was aerobic since the sampling was done on top of the compost. Samples were taken by scraping compost into a plastic jar and were used to inoculate immediately after transport to the lab at room temperature. The sample of the 2nd isolation was kept under anaerobic conditions during transport by using an Oxoid AnaeroGen sachet in a sealed anaerobic box.

Isolation procedure

For the first isolation, 25 g mature compost was added to 250 mL CTVM-CMC in a 500 mL flask and shaken for 3 h at 150 rpm at 65°C. After 3 h, compost was sieved with a 3 mm pore size filter and dilution series were plated on CTVMy-CMC and CTVM-CMC and grown aerobically for 72 h. Subsequently, single colonies were picked and transferred to fresh plates containing either CTVMy-CMC or CTVM-CMC (according to their isolation plate). Identical plates lacking CMC were used as negative control. Grown isolates were transferred twice to assay consistent growth and eliminate false-positives. These false positives present colonies that possible grow on the degradation products produced by surrounding colonies. By transferring colonies to fresh plates, we could eliminate those from the selection. Remaining isolates were then grown anaerobically

for 96 h to assay their potential of anaerobic growth. The remaining isolates were used to obtain pure cultures and afterwards stored in 15% glycerol at -80°C.

In the second isolation, 10 g active compost was inoculated into 100 mL CTVM in a 500 mL beaker and stirred for 5 minutes at room temperature. Due to the active state of the compost, oxygen levels are believed to be limited at the sampling site. Therefore, the isolation of microorganisms from this sample was initiated under anaerobic conditions to mimic the site of sampling. Dilution series were plated on CTVMy-CMC and CTVM-CMC and grown anaerobically for 72 h. Subsequently, single colonies were picked and each colony was streaked to plates containing CTVMy or CTVM (according to their isolation plate) with or without CMC and grown anaerobically at 65°C for 168 h (7 days). Isolates lacking growth on the negative controls were transferred twice more to CTVM(y)-CMC plates to assay consistent growth and eliminate false-positives. Pure cultures were obtained on CTVMy-glc plates and subsequently stored in 15% glycerol stocks at -80°C.

16S rRNA-encoding gene identification and phylogenetic analysis

Single colonies were picked from plates and inoculated to pre-warmed CTVMy medium with 1% w/v glucose. When cultures reached exponential phase (OD_{600} between 0.5 and 1) 2 µl was transferred to a PCR tube. PCR mix was added containing 1.25 units of DreamTaq DNA polymerase (Fermentas), 10x DreamTaq buffer (Fermentas), 100 µM of dNTPs (Fermentas), 0.2 µM of both primers GM3 (AGAGTTTGATCATGGC) and GM4 (TACCTTGTTACGACTT) and milliQ water in a total volume of 50 µl (Muyzer et al. 1995). PCR products were checked on 1% agarose gels and products were purified by using a GeneJet PCR purification kit (Fermentas). CloneManager software was used to assemble and manually curate the GM3 and GM4 sequences into one contig after which BLASTn was used for identification against the non-redundant nucleotide collection of the NCBI database. Contigs were aligned using Mega6 (Tamura et al. 2013) software and MUSCLE v3.8.31 (Edgar 2004) with 5 iterations and, subsequently, were trimmed to equal length using Jalview version

2.0 (Waterhouse et al. 2009). The 94 sequences have been submitted to NCBI under GenBank accession numbers KX113522 - KX113615. A phylogenetic tree was created using the neighbor-joining method (Saitou & Nei 1987) and a bootstrap analysis (Felsenstein 1985) was done using 1,000 replicates. Type strains used in the phylogenetic tree were derived from the following GenBank accession numbers: NR_043021.1 (*G. thermodenitrificans* DSM 465), NR_043022.1 (*G. thermoglucosidasius* DSM 2542), NR_028708.1 (*G. caldoxylosilyticus* DSM 12041), NR_115285.1 (*G. kaustophilus* DSM 7263), FN428684.1 (*G. thermoleovorans* (DSM 5366), NZ_CP010052.1 (*B. subtilis* DSM 402).

Selection

All pure strains were inoculated from glycerol stocks to plates containing CTVMy-CMC and CTVMy-xylan and subsequently incubated anaerobically for 48 h. Plates were then washed using milliQ water to remove all cells and subsequently stained for 5 minutes using a 0.1% (w/w) Congo red dye. After staining, plates were de-stained for 15 minutes using 1M NaCl under constant swirling. Clearing zones were identified visually and graded using a 4-step scale ranging from 0 to 4 with zero being no visible clearing zone and four representing the biggest clearing zone.

Screening on fermentation products was carried out in duplicate in 15 mL Greiner tubes with a total volume of 10 mL containing CTVMy medium at an initial pH of 6.5 with cellobiose or xylose as carbon substrate. Tubes were inoculated from plate colonies and incubated at 65°C for 48 h with agitation at 50 rpm. After 48 h samples were taken for HPLC analysis on fermentation product profiles and quantities. Strains were then ranked based on highest total amount of products and highest total lactic acid formed. The average score of both cellobiose and xylose duplicates were combined and best scoring strains were then compared to the selected isolates from the Congo red plate assay (data not shown). Isolates with high enzymatic activity for either CMC and/or xylan combined with an above average ranking in the fermentation assay were selected to examine their genetic accessibility.

Genetic accessibility and heterologous expression of the *pheB* reporter gene

Three isolates closely related to *G. thermodenitrificans* (99%), *G. caldoxylosilyticus* (100%) and *G. thermoleovorans* (99%) were selected to evaluate genetic accessibility by electroporation. Strains were subjected to transformation in triplicate according to the protocols for *Geobacillus* described previously (Taylor et al. 2009; Cripps et al. 2009) using the pNW33n plasmid isolated from *E. coli* dH5α. After 2 h recovery at 60°C, electroporated cells were spread on LB2 plates containing 7 µg/mL chloramphenicol and incubated for 24 h at 60°C. Colonies that appeared within 48 h were subjected to PCR as described above to verify the presence of plasmid pNW33n inside the cells.

Optimization of the transformation protocol was performed on strain T12. Three concentrations (0, 0.25 and 2.5 g/L) of K_2HPO_4 in the growth medium were used and washing of the cells was done 2 times with 50mL milliQ water followed by 2 washing steps, of 25 and 10mL respectively, using 10% (v/v) glycerol (Narumi et al. 1992).

As a demonstration of genetic accessibility, we cloned the reporter gene *pheB,* derived from *G. stearothermophilus* DSM6285 (GenBank accession no. DQ146476.2) into strain T12. The *pheB* gene was synthesized by GeneArt (Thermo Scientific) as an expression construct using the constitutive promoter PuppT12. Promoter PuppT12 is derived from the upstream region (100nt) of the uracil phosphoribosyltransferase gene from T12 (Additional file FigS1). The promoter together with the *pheB* reporter gene were cloned to pNW33n between the *Acc*65I and *Pst*I restriction sites. Activity of the reporter gene was tested by spraying colonies with 100 mM catechol followed by incubation at 55°C for 5 minutes.

Fermentations

Fermentations were performed in an Eppendorf DASGIP parallel bioreactor system or an Applikon fermentor system. In the Eppendorf DASGIP parallel bioreactor system glass reactors of 1.4 L and a working volume of 0.5 L was used with Dasgip control 4.0 to control the process. Glass reactors of 2 L and a working volume of 1

L were used in the Applikon fermentor system with BioXpert V2 for control. The conditions were as follows: Temperature was controlled at 65°C, pH at 7.0 by addition of 3M KOH and the stirring speed was 150 rpm. Antifoam 204 (Sigma-Aldrich) was added as required.

Pre-cultures were grown overnight from glycerol stock in 10 mL medium in a 50 mL tube at 65°C and 150 rpm. The next morning 3 mL was transferred to 50 mL of medium in a 250 mL baffled Erlenmeyer and grown for 2 h before 2% (v/v) was inoculated to the reactor. Samples of 2 mL were taken for off-line OD-measurements by determining the turbidity at 600 nm. Concentrations of sugar and fermentation products were determined by HPLC.

HPLC

Sugars and fermentation products were measured using an ICS5000 HPLC system from Thermo Scientific equipped with a Dionex DP pump, Dionex AS-AP autosampler, Dionex VWD UV detector operated at 210 nm and Shodex RI detector at 35°C. An Aminex HPX-87H cation-exchange column was used with a mobile phase of 8 mM H_2SO_4 and was operated at 0.8 mL/min and 60°C. All samples were diluted with 10 mM dimethyl sulfoxide in 4 mM H_2SO_4 in a ratio of 5:1 sample/internal standard.

Results

To isolate facultative anaerobic thermophiles capable of degrading cellulose and/or xylan, mature and active compost of plant materials were sampled with temperatures of 35°C and 65°C respectively, at time of isolation.

Isolation

The mature compost was suspended in CTVM medium at 65°C and dilution series were plated on CTVM-CMC with or without 0.5 g/L yeast extract at 65°C to select for thermophiles. Addition of yeast extract resulted in approximately 1.5 times more colonies. From both media, 130 isolates were picked and grown aerobically for almost

4 days. Clear differences in growth speed/lag time and recovery rate were observed in favour of plates containing yeast extract (Additional file 2; Table S2). The control plates without carbon source (CTVM and CTVMy) showed no or very little growth of the isolates, indicating that CMC was required to obtain the growth monitored. Growth of some isolates became visible later than that of others, which may indicate either slower growth rates or an elongated lag phase, due to growth on degradation products of other isolates. When isolates were streaked again to fresh medium, lag time was shortened and the number of recovered isolates decreased, possibly due to the elimination of false positives (Additional file 2; Table S2).

The active compost sample used in the second isolation was resuspended in CTVM medium, plated on CTVMy-CMC medium and incubated anaerobically at 65°C for 72 h. After incubation, 130 colonies from the CTVMy-CMC plates and 130 colonies from CTVM-CMC plates were isolated and subsequently transferred to fresh plates. None of the isolates from the CTVM-CMC plates were recovered and only 17 could be recovered from the CTVMy-CMC plates. Further transfers demonstrated that maintaining the isolates on CMC medium was often unsuccessful and revival from glycerol stocks was even more challenging (Table 1).

Table 1. Isolate numbers recovered after multiple transfers on both CTVM-CMC and CTVMy-CMC media. From the initial 520 isolates 94 have been recovered.

Isolation round	Media	Isolates	Recovery 1st transfer	Recovery 2nd transfer	Recovery 3th transfer	Recovery -80 stock
1st (aerobic)	CTVM-CMC	130	63	31	31	**9**
	CTVMy-CMC	130	99	79	79	**79**
2nd (anaerobic)	CTVM-CMC	130	0	0	0	**0**
	CTVMy-CMC	130	17	16	13	**6**
Total		**520**	**179**	**126**	**123**	**94**

Selection

Pure cultures obtained from both isolation rounds were subjected to 16S rRNA-encoding gene sequencing for identification. All 94 isolates belonged to the *Geobacillus* genus (Figure 1). The majority, 73 of 94 isolates, were *G. thermodenitrificans*. The remainder was classified as *G. thermoglucosidans* (11), *G. caldoxylosilyticus* (8) and *G. kaustophilus* (2).

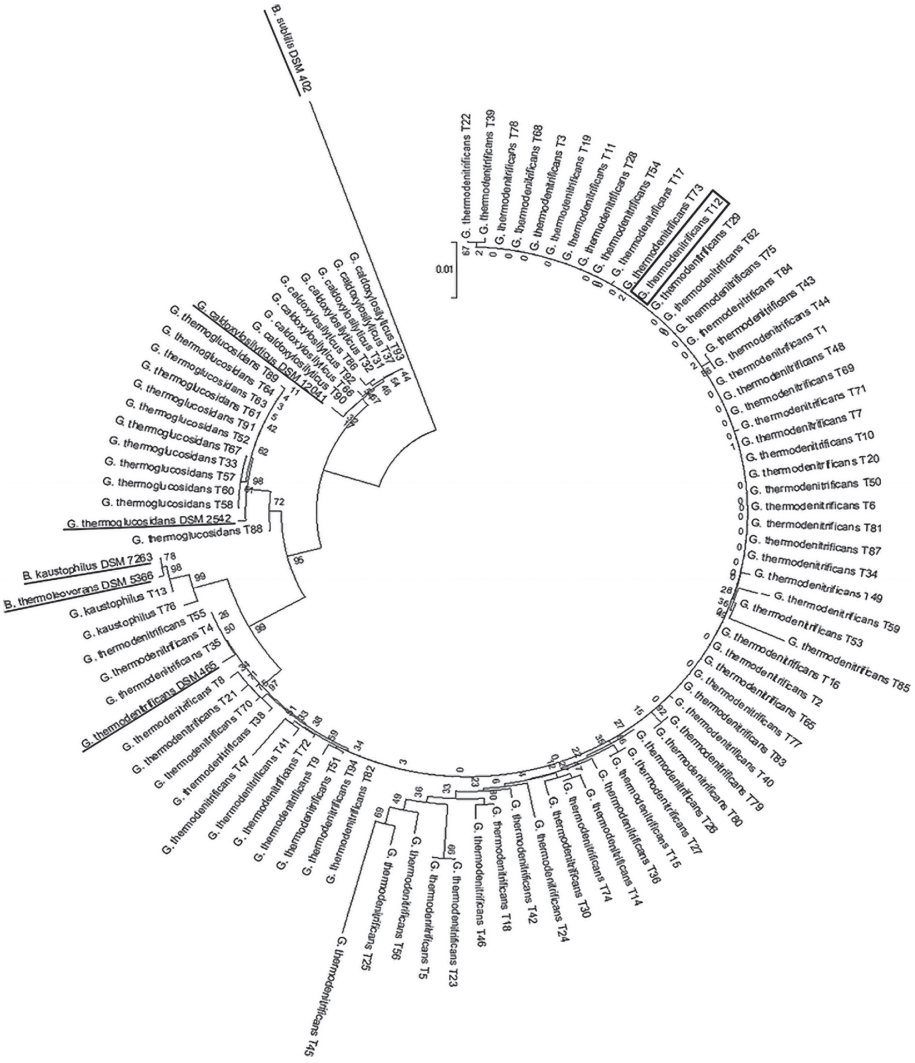


Figure 1. Phylogenetic tree of 16S rRNA gene sequences of *Geobacillus* isolates and their type strains (underlined). Phylogenetic tree was constructed by the Neighbor-joining method and tested with the bootstrap method using 1,000 replicates. Selected isolate *G. thermodenitrificans* T12 is indicated by a black box.

Further selection was based on fermentation product profiles. Isolates were pre-grown on CTVMy-glc and subsequently transferred to fresh CTVMy media containing either glucose or xylose as the carbon source. After 48 h fermentation products were measured by HPLC (Additional files 3 and 4; Tables S3 and S4). All strains produced lactic acid as their main fermentation product. Besides

lactic acid, acetic acid was present in all cultures. In some cases, independent of the amount of lactic acid, ethanol and formate were also produced. Succinate production was seen in some cultures, with *G. thermodenitrificans* T78 showing the highest average concentration of 4.3 mM succinic acid accounting for 19.3% (mol/mol) of the total products (data not shown). All strains belonging to *G. thermoglucosidans* showed similar production profiles implying low variability within this genus in comparison to for instance *G. caldoxylosilyticus*. *G. thermoglucosidans* is highly represented in the 25 strains with the highest organic acid production titers. This is demonstrated by the presence of 10 out of 11 strains in the cellobiose fermentations ranking and 8 of 11 strains in the xylose fermentations ranking (Additional files 3 and 4; Tables S3 and S4). Contrary to this observation, *G. thermoglucosidans* is underrepresented in the ranking of the 25 best lactic acid producers with only 3 (C6) and 4 (C_5) strains listed. Under such conditions, *G. thermodenitrificans* has a more homolactic phenotype while *G. thermoglucosidans* has a more mixed acid fermentation phenotype (data not shown). The mixed acid fermentation profiles of *G. thermoglucosidans* are in line with a previous study that demonstrated a mixed acid fermentation profile under micro-aerobic conditions (Tang et al. 2009).

The list of strains producing the highest amounts of lactic acid from cellobiose fermentation was overlaid with the list of isolates producing most lactic acid on xylose. No isolates belonging to *G. caldoxylosilyticus* or *G. kaustophilus* appeared in both lists. In addition to the fermentation assays described above, all 94 isolates were subjected to Congo red stain after growth on CMC and beechwood xylan. About 24 strains degraded CMC, as indicated by halo formation after staining with the Congo red dye. However, after repeated assays using Congo red it became evident that the formation of halos was highly irreproducible, although growth was similar in most cases. When grown on xylan, halos were more pronounced and a clear degradation was observed for about half of the isolates. Strain T12 showed strong and reproducible breakdown of xylan. Based on the concentrations of the products in the cellobiose and xylose fermentations, and the reproducibility of the plate assays, strains

T12, T62, and T85 (all originating from the 1st isolation round) were considered the best isolates for further study.

Electroporation and optimization of transformation

The genetic accessibility of these three selected isolates were then evaluated by electroporation performed based on protocols described previously (Cripps et al. 2009; Taylor et al. 2008), with some modifications. Cells were transformed by electroporation using the *E.coli-Bacillus* shuttle vector pNW33n, a vector that has been widely used in thermophilic bacilli (Bosma et al. 2015; Blanchard et al. 2014; Lin et al. 2014). Colonies found to be *chloramphenicol*-resistant were streaked to fresh LB2 plates supplemented with chloramphenicol to eliminate false-positives and successful transformation was confirmed by colony PCR. Following this, plasmids were isolated from the positive clones and subjected to restriction analysis to confirm integrity, before transfer back to *E. coli*. While T62 and T85 did not yield any colonies, transformation of strain T12 reproducibly did. In all cases the colonies obtained for T12 were positive for plasmid uptake. Although the efficiency was low, the high reproducibility provided the opportunity for optimization. Different culture media and wash protocols were evaluated (Table 2), and a combination of reduced concentrations of K_2HPO_4 with a wash protocol derived from (Narumi et al. 1992), proved to be the best protocol with an optimum CFU/µg DNA of 1.7×10^4. The LB2 medium initially used resulted in a short pulse time constant of 3.7ms. However, changing this growth medium to a variant with a lower salt concentration resulted in an increased pulse time constant and a CFU/µg DNA of 1240, which was in line with the results seen in previous research (Bosma et al. 2015). Consequently, decreasing the salt concentrations in the growth medium LB2 resulted in the cell pellets becoming less dense, leading to loss of many cells during washing. To prevent the loss of cells, 0.25 g/L K_2HPO_4 was added and wash buffers were changed to milliQ water and 10% glycerol, to increase the osmotic shock that results in weakening of the bacterial cell wall. Several variables need to be balanced to obtain a successful electroporation protocol. For example, with a more severe wash protocol, the cells will be weaker during

electroporation, which will probably require a lower voltage to reduce further damage to the cells. Although no statistical underpinning was performed, our protocol is demonstrated to be reproducible.

To verify the potential of the T12 genetic accessibility, we introduced *pheB* as a reporter gene. The reporter gene *pheB*, derived from *G. stearothermophilus* DSM6285 (GenBank accession no. DQ146476.2), encodes a heterologous catechol 2,3-dioxygenase (C23O). The *pheB* gene is demonstrated to function as a reporter gene in *G. thermoglucosidans*, where it has been used for the quantification of promoter strength (Bartosiak-Jentys et al. 2012). The protein C23O oxidizes catechol to form 2-hydroxymuconic semialdehyde (2-HMSA) which has a bright yellow color. Colonies harboring the *pheB* expression construct can therefore be visualized by spraying them with 100 mM catechol and incubating them for 5 minutes at 55°C. A bright yellow color was observed on all colonies harboring the *pheB* expression construct, while no color change was observed in colonies harboring empty pNW33n plasmids (Additional file 5; Figure. S5).

Table 2. Overview of transformation optimizations for *G. thermodenitrificans* T12.

Parameter changed	Medium	Growth[1] (h)	Final OD[2]	Voltage	Ω	µF	Cuvette width (mm)	Ptd[3] (ms)	DNA[4] (µg)	CFU/ µg DNA
N.A.	LB2 (2.50 g/L K_2HPO_4)	2	1.01	2,000	200	25	2	3.7	3,031	4
Medium	LB2 (0.25 g/L K_2HPO_4)	1.67	0.96	2,000	200	25	2	5.1	1,613	1,240
Medium	LB2 (0.00 g/L K_2HPO_4)	1.66	0.96	2,000	200	25	2	5	1,613	22
Wash buffer[a]	LB2 (2.50 g/L K_2HPO_4)	1.67	0.96	2,000	200	25	2	5.2	1,613	19
Medium/ wash buffer[a]	LB2 (0.00 g/L K_2HPO_4)	1.66	0.96	2,000	200	25	2	5.4	1,613	2,988
Medium/ wash buffer[a]	LB2 (0.25 g/L K_2HPO_4)	1.83	0.96	2,000	200	25	2	5.5	1,564	17,071

[1] Time of growth after dilution of the overnight pre-culture
[2] Final OD_{600} of the culture after growth prior to the washing step
[3] Pulse time duration in milliseconds
[4] Amount of plasmid (pNW33n) DNA added for transformation
[a] Wash protocol was changed to 2 times washing with milliQ water (50mL) followed by washing with 25mL and 10mL milliQ water+10% glycerol.

Growth and fermentation of T12

Fermentation characteristics of *G. thermodenitrificans* T12 were evaluated under micro-aerobic conditions. To determine an optimal pH, multiple pH-controlled fermentations were carried out in MMy media containing 30 g/L glucose, with a pH range of 5.5 to 8.5. Both growth and product formation were monitored during the fermentation (data not shown) to determine the optimal conditions. Growth was observed between pH 5.5 and 7.5, with an optimum between pH 6.5 and 7.5. This optimal growth pH of 7.5 coincided with the highest amount of lactic acid production followed closely by the culture at pH 6.5.

To assay the potential of strain T12 in converting biomass derived sugars, its fermentation products and productivity thereof, were quantified in fermentations with glucose, xylose and mixtures of these sugars as carbon source. For all carbon substrates, both MMy media (containing 0.5 g/L yeast extract and NH_4NO_3) as well as MMy+ media (containing 5 g/L yeast extract and NH_4Cl) were evaluated (Figure 2; Table 3). Ammonium nitrate was added to MMy medium to stimulate biomass formation (Mishima et al. 2009).

The average lactic acid productivity on 30 g/L glucose or 30 g/L xylose, either on minimal or rich medium, did not differ significantly. However, productivities differ when comparing a single carbon source on minimal and rich medium (Table 3). Co-fermentation of both glucose and xylose was performed in both media (Figure 2, Table 3). When MMy medium with a 2:1 ratio of glucose: xylose was used, a total of 17.9 g lactic acid was produced in 55 hours with an overall yield of 0.7 g lactic acid per gram sugar consumed. From the start of the fermentation, both glucose and xylose were consumed simultaneously though the consumption rates for xylose were lower in comparison to glucose (Table 3). Fermentations in MMy+ medium showed improved consumption rates, productivity and yield of lactic acid produced per amount of consumed substrate.

Table 3. Overview of fermentations of *G. thermodenitrificans* T12 on glucose, xylose and glucose/xylose mixtures. Fermentations were carried out at pH 7.0 at 150 rpm. Fermentations started with a 0.5 L air headspace without additional sparging. MMy media contained 0.5 g/L yeast extract whereas MMy+ media contained 5 g/L yeast extract.

Media	Substrate	Time (h)	Cons. (g)	Cons. rate (g/L/h)	Lactic acid production (g)	Lactic acid productivity (g/L/h)	Lactic acid yield (g product/ g substrate consumed)	Lactic acid yield (% of theoretical yield)
MMy	30 g/L Glucose	25.2	9.43	0.37	3.10	0.12	0.33	10.3
MMy	30 g/L Xylose	25	6.93	0.28	3.83	0.15	0.55	12.8
MMy	20 g/L Glucose + 10 g/L Xylose	54.95	19.87 glucose 6.55 xylose	0.33 glucose 0.11 xylose	17.93	0.30	0.69	59.8
MMy+	30 g/L Glucose	27	11.35	0.41	8.48	0.30	0.74	28.3
MMy+	30 g/L Xylose	27	15.23	0.54	8.73	0.31	0.57	29.1
MMy+	20 g/L Glucose + 10 g/L Xylose	25	14.51 glucose 7.05 xylose	0.50 glucose 0.24 xylose	20.27	0.71	0.94	67.6

In all mixed-sugar fermentations the xylose consumption rate was 2 to 3 times slower when compared to the consumption rate of glucose. To exclude the possibility that the sugar concentration is instrumental in this observation, a glucose/xylose co-fermentation was completed with equal concentrations of the two sugars (10 g/L) in MMy medium. Observed consumption rates of 0.2 g/L/h for glucose and 0.1 g/L/h for xylose confirm that glucose is in fact favored over xylose in co-fermentation.

The Congo-Red assay showed that strain T12 was capable of degrading beechwood xylan on solid medium. To quantify its fermentation products liquid cultures with 10 g/L beechwood xylan as carbon source were used. Fermentations were carried out in 50mL Greiner tubes with different volumes of head space and 250 mL shake flasks to demonstrate the impact of oxygen (Figure 3). Under micro-aerobic conditions in low nutrient (MMy) medium (Figure 3; A-B), lactic acid (11 mM) was the main fermentation product followed by acetic acid. In 250 ml baffled Erlenmeyer shake flasks using 50 mL culture (1:5 with more aeration), lactic acid concentration was reduced to 0.55 (± 0.04) mM and acetic acid increased to 19 (± 0.04) mM (data not shown). Use of the rich (MMy+) medium under any condition resulted in minimal amounts of lactic acid production (Figure 3; C) and increased acetate production, due to the conversion of the yeast extract (Figure 3; D). In addition to the increased lactic acid production we noticed an increase in cell densities in minimal media with xylan in comparison to cultures on minimal media without xylan (Figure 3; A). In rich media, no significant difference in cell densities were observed between cultures with and without xylan, likely due to high growth rates on yeast extract (Figure 3; C).

Discussion

This study was aimed at the discovery of novel industrially relevant strains that harbor saccharolytic enzymes to degrade cellulose and xylan, have lactic acid as their main fermentation product and are genetically accessible. We describe the isolation of 94 facultative

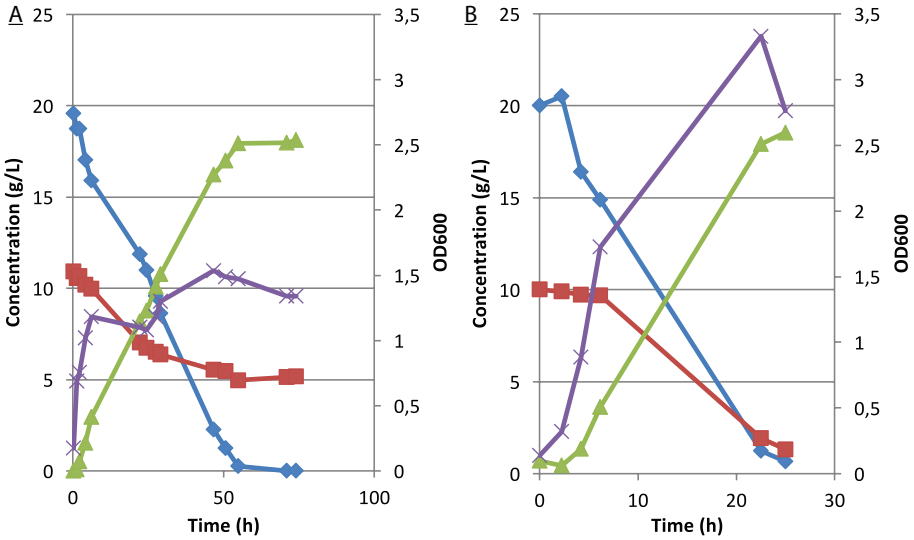


Figure 2 Fermentations of *G. thermodenitrificans* T12 on a 2:1 glucose/xylose mixture. Fermentations were carried out at pH 7.0 at 150 rpm. Fermentations started with a 0.5 L air headspace without additional sparging and media consisted of 500 mL MMy (left) or MMy+ (right). Diamonds: glucose; closed squares: xylose; triangles: lactic acid; asterisks: OD_{600}.

anaerobic Geobacillus strains that were capable of fermentation of both C6 and C5 sugars under micro-aerobic and anaerobic conditions. For isolate *G. thermodenitrificans* T12 direct conversion of beechwood xylan to lactic acid was demonstrated. The selection methods yielded several isolates that were capable of degrading xylan and cellulose.

All strains described in this study were isolated from compost at both moderate and elevated temperature by repetitive growth on both CMC and xylan. It has been previously shown that *Geobacillus* was the dominant genus in "mature" compost (Takaku et al. 2006; Tian et al. 2013; Ryckeboer et al. 2003). However, it cannot be concluded if the ratio of *Geobacillus* species found in this study represents the composition of the *Geobacillus* community in the compost heap sampled, or results from selection by the sampling and isolation methods. In a prior study, compost heaps on the same site were sampled and screened for thermophilic organisms (Bosma et al. 2015). The TMM media used in the study of Bosma *et al.* contained 0.9 g $NaNO_3$ which led, at 65°C, to the isolation of mainly

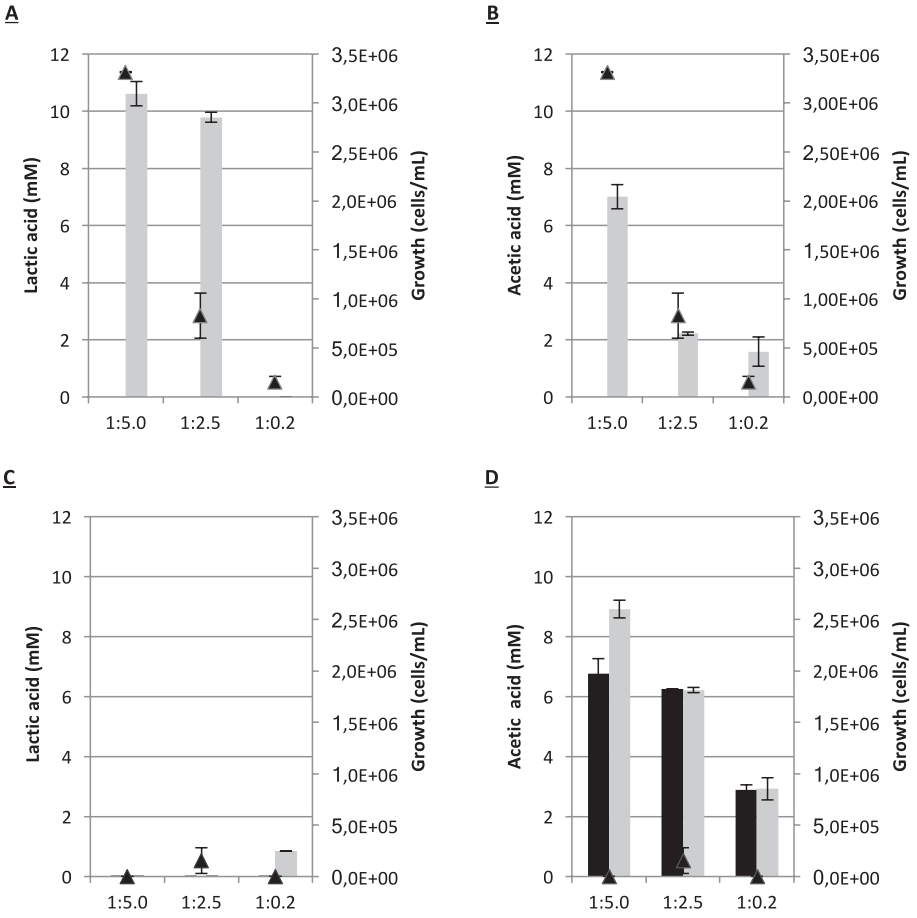


Figure 3. Xylan fermentations by G. *thermodenitrificans* T12. Lactic acid production (A, C) and acetic acid production (B, D) of strain T12 on MMy (A, B) and MMy+ (C, D) media with xylan (light-grey bars) and without xylan (dark-grey bars). The cultures were grown in 50mL Greiner tubes at pH 6.5 and 150 rpm. Ratios on the X-axis represent the ratio of medium and headspace during fermentation. Growth (▲) is expressed as increase of cells on xylan compared to cultures without xylan.

G. thermodenitrificans strains. When both vitamins and metals were added, a larger diversity in *Geobacillus* species was found. In isolations done at 60°C without nitrate in the medium approximately half the isolates belonged to *Geobacillus* but no *G. thermodenitrificans* isolates were found (E.F. Bosma, unpublished data). With the selection criteria used in this study, it was anticipated that geobacilli would be isolated. The overrepresentation of *G. thermodenitrificans* (79% of isolates) in this isolation procedure, however, was surprising

since nitrate was excluded from the isolation media to prevent this species benefiting over other species.

All strains were isolated after repeated growth on CMC and beechwood xylan. Degradation of xylan was always detectable and most strains showed clearing zones when assayed on Congo-red plates. In contrast, the results on CMC in Congo-red assays appeared difficult to reproduce. Although never described for *Geobacillus* or closely related species, the irreproducibility is most likely caused by regulation of the expression of saccharolytic enzymes, as has been described for xylanolytic enzymes (Shulami et al. 2014). The latter study describes that the hemicellulose utilization (HUS) locus is most likely induced by small amounts of free xylose. It is known that the coverage of the HUS locus varies substantially among *Geobacillus* species and even differs per strain (De Maayer et al. 2014). From our study, it has become clear that within geobacilli there is major variation in the regulation of the xylanolytic cassette. In apparently identical situations, the breakdown of xylan differs greatly between *Geobacillus* species and even between strains of the same species. Regulation of cellulase gene expression has not yet been described for *Geobacilli* and details of its control remain elusive. Growth is no direct evidence for the presence of extracellular cellulases although free glucose is limited in this substrate which implies at least some cellulolytic activity to be present.

When grown under micro-aerobic conditions, all isolates generated lactic acid as their main fermentation product from glucose and xylose. Acetate and succinate were always present as minor by-products, and occasionally, formate and ethanol were produced. Isolates belonging to the *G. thermoglucosidans* clade showed low differentiation in terms of fermentation products and amounts, whereas *G. caldoxylosilyticus* and *G. thermodenitrificans* produce quite a diverse spectrum.

Genetic accessibility was tested for three selected isolates. Strain T12 was found to be reproducibly transformable, albeit with low efficiencies of 3-5 transformants per µg of plasmid DNA. In previous research, only 2 of 25 strains of *G. thermodenitrificans* could be transformed (Bosma et al. 2015). The transformation efficiencies of

G. thermodenitrificans isolates in the latter study are in line with the efficiency of transformation described here. Efficient transformation of a *G. thermodenitrificans* strain by electroporation has been described before (Kananavičiute & Čitavičius 2015) yielding approximately 2.8 × 10^6 colonies per µg DNA. However, this protocol resulted in only 3-5 T12 transformants per µg of plasmid DNA. The efficiency reported by Kananavičiute and Čitavičius was obtained by using plasmid DNA isolated from the same strain, circumventing possible degradation by a native restriction modification system. Conversely, there is no literature available describing a systematic approach in optimizing electroporation protocols for *Geobacillus* species. We have optimized the transformation for *G. thermodenitrificans* T12 by adjusting the concentration of K_2HPO_4 in its growth media used in preparation of making the cells competent. The change of wash buffers further increased the transformation efficiency to CFU/µg DNA of 1.7×10^4. With this transformation efficiency, we were able to introduce the heterologous reporter gene *pheB* to T12. The expression of the *pheB* gene, controlled by a constitutive promoter (PuppT12), shows the potential of strain T12 for further genetic engineering. Strain T12 was selected based on its ability to grow on CMC plates, its reproducible xylan degradation, an above average product titer on C6 and C5 fermentations, and its genetic accessibility, which provides opportunities to develop this strain into a host for CBP processes. To reach this potential, the organism should be engineered to produce cellulolytic enzymes.

Strain T12 was found to have an optimum pH between 6.5 and 7.5, and temperature of 65°C, identical to that of isolation. When T12 is grown on pure glucose or xylose the sugar consumption rate was almost equal, whereas during co-fermentation the glucose consumption rate was approximately two to three times as high compared to the xylose consumption rate. This indicates that the presence of glucose partially inhibits xylose uptake and/or metabolism. Reduced xylose consumption by *Geobacillus* in a glucose/xylose mixture has been reported before (Tang et al. 2009). On average, hemicellulose makes up 25-35% of lignocellulosic biomass, which makes simultaneous sugar uptake advantageous for the CBP

economy (Murciano Martínez et al. 2015). There are many microbes capable of xylose consumption; however, most of these organisms suppress the uptake of xylose when glucose is present. This negative feedback results in sequential consumption of available sugars, leading to elongated fermentation times or accumulation of xylose in the reactor during fed-batch.

We have demonstrated the direct conversion of beechwood xylan into lactic acid by *G. thermodenitrificans* T12. The ratio of medium to headspace was important for this process, where increased oxygen transfer seems to positively affect xylan degradation, but negatively influences the ratio of lactic acid to acetic acid produced. Higher acetic acid production was seen with higher oxygen concentrations, while lactic acid was mostly produced in micro-aerobic conditions. The reduced degradation of xylan under high nutrient conditions (MMy+) was most likely caused by the repression of *xynA* by the global regulator CodY. It is known from *B. subtilis* that the CodY regulator is responsive to branched-chain amino acids (nutrient-rich conditions) (Sonenshein 2005). A CodY orthologue has been found in the genome sequence of *G. thermodenitrificans* T12 (to be published), and two potential binding regions for CodY, based on the binding region consensus sequences from (Shulami et al. 2014; Sonenshein 2005), were identified in the promoter region of the T12 *xynA* gene (Figure 4). For *G. stearothermophilus* T6, which has a xylanolytic operon similar to that of T12, CodY binding to the *xynA* promoter was demonstrated and its function as repressor of *xynA* expression was postulated (Shulami et al. 2014). Our results are in line with this hypothesis (Figure. 3 C, D).

Conclusions

In conclusion, a collection of 94 *Geobacillus* isolates were obtained from an initial selection of 520 thermophilic organisms, by screening for lactic acid production on both C6 and C5 sugars, together with the ability to degrade xylan and/or cellulose. Genetic accessibility was confirmed and optimized for one isolate, which was designated *G.*

thermodenitrificans T12. This strain can grow on CMC plates, although Congo red assays are inconsistent. Strain T12 is capable of fermenting both xylose and glucose simultaneously with lactic acid as its main fermentation product. In addition, beechwood xylan was directly converted to lactic acid. The capacity of strain T12 to ferment xylan together with its genetic accessibility makes *G. thermodenitrificans* T12 a potential candidate for consolidated bioprocessing.

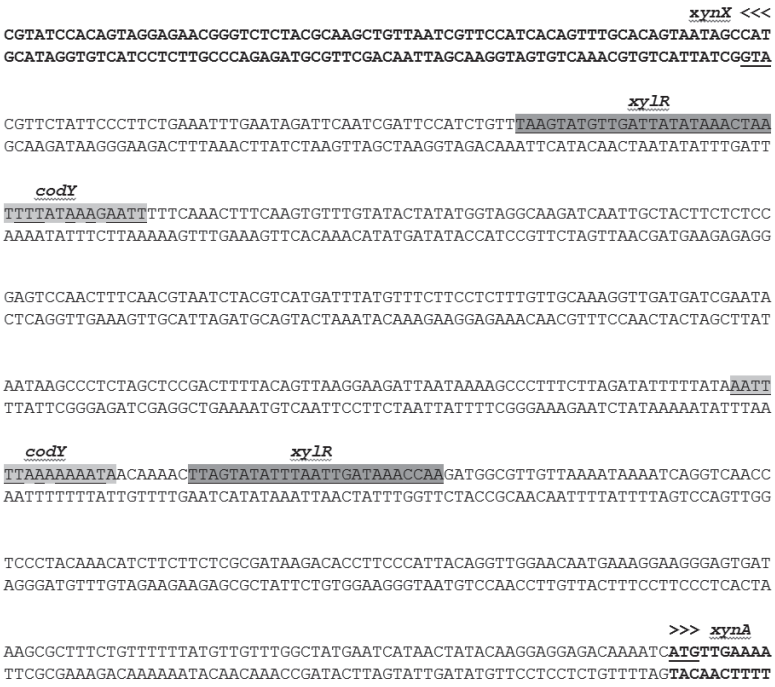


Figure 4 Intergenic region of the divergently oriented *xynX* and *xynA* genes. The predicted start codons of *xynA* and *xynX* genes are shown boldface and underlined. Predicted CodY-binding regions and predicted xylose regulator XylR-binding sites are shaded grey. Nucleotides matching the *B. subtilis* consensus sequences have been underlined. The CcpA-binding sites found in *G. stearothermophilus* T6 (Shulami et al. 2014) were not found.

Acknowledgements

We thank Willem van Harten from Recom Ede for allowing us to sample their compost and Sofia Thomaidou and Ioanna Papakosta for their excellent technical assistance.

Supplementary information

5' - TAAGTGTGCCTTTTCCTTTGCTTCAACGGTTGAACGGGCGCCCGTTTTCCAGTAGAATGT

ATAGAAGTGTACTGCATACATACGGAAGAGGAGATGACCT**ATG** - 3'

File S1. Sequence of the PuppT12 promoter including the first ATG codon of the uracil phosphoribosyltransferase gene.

Table S2. Overview of the number of isolates recovered from the first isolation. Overview of the number of isolates recovered from the first isolation on CTVMy-CMC and CTVM-CMC media after 20h, 48h and 88h. The pH was set at 6.5 and plates were incubated at 65°C. Dilution series were plated from compost that had been shaken at 150 rpm for 3h in the same medium and temperature.

	1st transfer			2nd transfer		
Media	20h	48h	88h	20h	48h	88h
CTVMy-CMC	28	94	99	62	79	79
CTVM-CMC	1	13	63	3	7	31

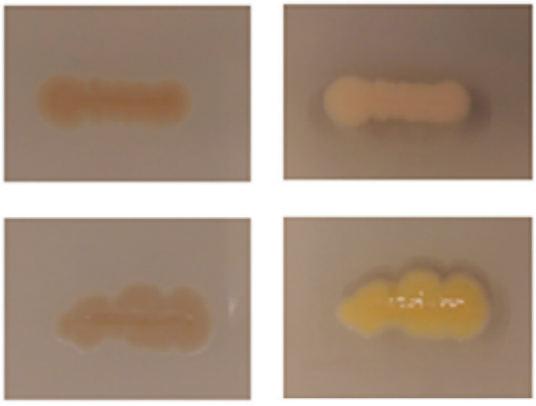


Figure S5. *G. thermodenitrificans* T12 transformed with pNW33n (A1) and pNW33n+Pheb (B1). Colonies were sprayed with 100mM catechol and incubated for 5 minutes at 55°C. The colony containing pNW33n did not show any color formation (A2) while the colony with the *pheB* gene under control of the constitutive uracil phosphoribosyltransferase promoter (PuppT12) shows a yellow color (B2) indicating the conversion of catechol to 2-hydroxymuconic semialdehyde.

Table S3. HPLC data from isolates ranked on total organic acid production and total lactic acid production on cellobiose (C6)

Strain	Identification	Lactate (mM)	SD.	**Total (mM)**	SD.	Strain	Identification	**Lactate (mM)**	SD.	Total	SD.
T62	*G. thermodenitrificans*	17.187	0.41	**32.429**	0.397	T85	*G. thermodenitrificans*	**19.727**	0.017	27.906	0.216
T87	*G. thermodenitrificans*	17.336	0.675	**31.606**	0.656	T13	*G. kaustophilus*	**17.772**	0.334	22.334	0.342
T83	*G. thermodenitrificans*	16.43	0.414	**31.431**	0.503	T12	*G. thermodenitrificans*	**17.669**	0.191	28.885	0.309
T39	*G. thermodenitrificans*	16.481	1.543	**29.055**	0.852	T35	*G. thermodenitrificans*	**17.414**	0.666	21.168	0.646
T12	*G. thermodenitrificans*	17.669	0.191	**28.885**	0.309	T87	*G. thermodenitrificans*	**17.336**	0.675	31.606	0.656
T85	*G. thermodenitrificans*	19.727	0.017	**27.906**	0.216	T50	*G. thermodenitrificans*	**17.277**	1.156	20.471	1.463
T64	*G. thermoglucosidasius*	14.652	0.182	**27.212**	0.587	T40	*G. thermodenitrificans*	**17.199**	0.415	20.768	1.139
T63	*G. thermoglucosidasius*	15.128	0.214	**27.15**	0.083	T62	*G. thermodenitrificans*	**17.187**	0.41	32.429	0.397
T57	*G. thermoglucosidasius*	15.053	0.528	**27.102**	0.358	T48	*G. thermodenitrificans*	**16.94**	0.49	20.448	0.349
T60	*G. thermoglucosidasius*	15.364	0.608	**26.878**	0.088	T55	*G. thermodenitrificans*	**16.87**	0.991	23.051	0.713
T61	*G. thermoglucosidasius*	15.065	0.173	**26.805**	0.136	T45	*G. thermodenitrificans*	**16.859**	0.897	20.395	0.783
T58	*G. thermoglucosidasius*	15.579	0.558	**26.726**	0.79	T52	*G. thermoglucosidasius*	**16.818**	0.043	22.245	0.341
T34	*G. thermodenitrificans*	14.815	0.432	**26.687**	0.495	T44	*G. thermodenitrificans*	**16.713**	1.24	20.779	0.853
T59	*G. thermodenitrificans*	15.514	0.132	**26.551**	1.476	T49	*G. thermodenitrificans*	**16.677**	1.913	19.93	2.463
T33	*G. thermoglucosidasius*	15.396	0.43	**26.504**	0.175	T39	*G. thermodenitrificans*	**16.481**	1.543	29.055	0.852
T93	*G. caldoxylosilyticus*	15.979	1.352	**26.278**	0.032	T83	*G. thermodenitrificans*	**16.43**	0.414	31.431	0.503
T88	*G. thermoglucosidasius*	15.178	0.542	**23.068**	0.336	T26	*G. thermodenitrificans*	**16.063**	1.083	19.792	1.047
T55	*G. thermodenitrificans*	16.87	0.991	**23.051**	0.713	T93	*G. caldoxylosilyticus*	**15.979**	1.352	26.278	0.032
T13	*G. kaustophilus*	17.772	0.334	**22.334**	0.342	T51	*G. thermodenitrificans*	**15.767**	1.008	21.194	1.484
T52	*G. thermoglucosidasius*	16.818	0.043	**22.245**	0.341	T30	*G. thermodenitrificans*	**15.725**	0.577	19.35	0.385
T31	*G. caldoxylosilyticus*	13.625	0.698	**21.616**	0.262	T91	*G. thermoglucosidasius*	**15.604**	0.023	20.773	0.015
T73	*G. thermodenitrificans*	15.435	0.286	**21.415**	0.45	T58	*G. themoglucosidasius*	**15.579**	0.558	26.726	0.79
T51	*G. thermodenitrificans*	15.767	1.008	**21.194**	1.484	T42	*G. thermodenitrificans*	**15.538**	3.213	20.06	1.884
T35	*G. thermodenitrificans*	17.414	0.666	**21.168**	0.646	T59	*G. thermodenitrificans*	**15.514**	0.132	26.551	1.476
T54	*G. thermodenitrificans*	15.371	0.284	**21.063**	0.74	T73	*G. thermodenitrificans*	**15.435**	0.286	21.415	0.45

Table S4. HPLC data from isolates ranked on total organic acid production and total lactic acid production on xylose (C_5)

Strain	Identification	Lactate (mM)	SD.	**Total (mM)**	SD.	Strain	Identification	**Lactate (mM)**	SD.	Total (mM)	SD.
T58	*G. thermoglucosidasius*	13.225	1.447	**27.555**	0.272	T90	*G. caldoxylosilyticus*	**18.553**	0.264	23.964	1.191
T60	*G. thermoglucosidasius*	15.638	0.054	**27.513**	0.187	T40	*G. thermodenitrificans*	**18.021**	0.198	22.553	0.074
T61	*G. thermoglucosidasius*	14.207	0.456	**27.446**	0.121	T1	*G. thermodenitrificans*	**17.681**	0.097	22.225	0
T57	*G. thermoglucosidasius*	14.331	0.746	**27.414**	0.689	T6	*G. thermodenitrificans*	**17.549**	0.355	23.335	1.925
T63	*G. thermoglucosidasius*	15.583	0.669	**27.322**	0.295	T70	*G. thermodenitrificans*	**17.501**	0.002	21.148	0.182
T59	*G. thermodenitrificans*	11.428	0.692	**26.955**	0.999	T88	*G. thermoglucosidasius*	**17.422**	0.057	21.555	0.198
T85	*G. thermodenitrificans*	14.14	0.122	**25.709**	1.628	T62	*G. thermodenitrificans*	**17.366**	0.209	24.157	2.753
T89	*G. thermoglucosidasius*	16.112	0.823	**25.705**	0.074	T80	*G. thermodenitrificans*	**17.36**	0.177	21.029	0.137
T39	*G. thermodenitrificans*	11.184	0.104	**25.596**	1.925	T7	*G. thermodenitrificans*	**17.311**	0.663	23.451	0.772
T51	*G. thermodenitrificans*	16.242	1.527	**25.353**	1.386	T44	*G. thermodenitrificans*	**17.165**	0.314	21.232	0.75
T91	*G. thermoglucosidasius*	16.489	0.167	**25.194**	0.639	T92	*G. caldoxylosilyticus*	**17.164**	0.158	25.131	0.903
T92	*G. caldoxylosilyticus*	17.164	0.158	**25.131**	0.903	T52	*G. thermoglucosidasius*	**16.916**	0.47	24.922	0.556
T52	*G. thermoglucosidasius*	16.916	0.47	**24.922**	0.556	T71	*G. thermodenitrificans*	**16.671**	0.154	20.371	0.047
T86	*G. caldoxylosilyticus*	15.527	0.619	**24.433**	0.141	T4	*G. thermodenitrificans*	**16.556**	0.45	19.97	0.139
T62	*G. thermodenitrificans*	17.366	0.209	**24.157**	2.753	T91	*G. thermoglucosidasius*	**16.489**	0.167	25.194	0.639
T24	*G. thermodenitrificans*	15.131	0.26	**24.107**	0.004	T25	*G. thermodenitrificans*	**16.415**	2.247	20.405	2.194
T90	*G. caldoxylosilyticus*	18.553	0.264	**23.964**	1.191	T51	*G. thermodenitrificans*	**16.242**	1.527	25.353	1.386
T7	*G. thermodenitrificans*	17.311	0.663	**23.451**	0.772	T2	*G. thermodenitrificans*	**16.151**	0.094	20.854	0.291
T6	*G. thermodenitrificans*	17.549	0.355	**23.335**	1.925	T89	*G. thermoglucosidasius*	**16.112**	0.823	25.705	0.074
T40	*G. thermodenitrificans*	18.021	0.198	**22.553**	0.074	T60	*G. thermoglucosidasius*	**15.638**	0.054	27.513	0.187
T10	*G. thermodenitrificans*	13.88	0.759	**22.482**	0.404	T63	*G. thermoglucosidasius*	**15.583**	0.669	27.322	0.295
T1	*G. thermodenitrificans*	17.681	0.097	**22.225**	0	T26	*G. thermodenitrificans*	**15.549**	0.958	18.806	1.41
T78	*G. thermodenitrificans*	14.062	0.105	**22.078**	0.124	T86	*G. caldoxylosilyticus*	**15.527**	0.619	24.433	0.141
T11	*G. thermodenitrificans*	11.909	0.244	**21.907**	5.901	T31	*G. caldoxylosilyticus*	**15.514**	1.405	21.728	0.426
T31	*G. caldoxylosilyticus*	15.514	1.405	**21.728**	0.426	T15	*G. thermodenitrificans*	**15.317**	0.474	19.554	0.741

Chapter 3

Complete genome sequence of *Geobacillus thermodenitrificans* T12, a potential host for biotechnological applications

Martinus J.A. Daas[1], Bastienne Vriesendorp[2], Antonius H.P. van de Weijer[1], John van der Oost[1] and Richard van Kranenburg [1,2]

Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, the Netherlands[1];
Corbion, Arkselsedijk 46, 4206 AC, Gorinchem, the Netherlands[2].

This chapter was adapted from:
Daas, M. J. A., Vriesendorp, B., Van de Weijer, A. H. P., Van der Oost, J. & Van Kranenburg, R. Complete genome sequence of *Geobacillus thermodenitrificans* T12, a potential host for biotechnological applications. *Current Microbiology*, (Submitted, 2017).

Abstract

In attempt to obtain a thermophilic host for the conversion of lignocellulose derived substrates into lactic acid, *Geobacillus thermodenitrificans* T12 was isolated from a compost heap. It was selected from over 500 isolates as a genetically tractable hemicellulolytic lactic acid producer, requiring little nutrients. The strain can ferment glucose and xylose simultaneously and can produce lactic acid from xylan, making it a potential host for biotechnological applications. The genome of strain T12 consists of a 3.64 Mb chromosome and two plasmids of 59 kb and 56 kb. It has a total of 3.676 genes with an average genomic GC content of 48.7%. The T12 genome encodes a denitrification pathway, allowing for anaerobic respiration. The identity and localization of the responsible genes is like those of the denitrification pathways found in strain NG80-2. The hemicellulose utilisation (HUS) locus was identified based on sequence homology against *G. stearothermophilus* T-6. It appeared that T12 has all the genes that are present in strain T-6 except for the arabinan degradation cluster. Instead, the HUS locus of strain T12 contains genes for both an inositol and a pectate degradation pathway. Strain T12 has complete pathways for the synthesis of purine and pyrimidine, all 20 amino acids and several vitamins except D-biotin. The host-defence systems present are comprised of a Type II and a Type III restriction-modification system, as well as a CRISPR-Cas Type II system. It is concluded that *G. thermodenitrificans* T12 is a potentially interesting candidate for industrial applications.

Introduction

Bioprospecting to discover novel organisms for the use in industrial applications is widely used and this strategy has led to an increased interest *Geobacillus spp.* (Adiguzel et al. 2009; Banat et al. 2004; Chamkha et al. 2008; DeFlaun et al. 2007; Lentini et al. 2007; Poli et al. 2011; Rastogi et al. 2009; Studholme 2015). The *Geobacillus* genus, which was reclassified in 2001 from *Bacillus,* consists of Gram positive, rod shaped, spore-forming bacteria that thrive under thermophilic conditions and are capable of aerobic and anaerobic respiration (Nazina et al. 2001). Recently, Aliyu *et al.* have suggested that the genus *Geobacillus* consists of two distinct genera, *Geobacillus* and *Parageobacillus.* The latter genus includes *P. caldoxylosilyticus, P. thermoglucosidans, P. antarticus, P. toebii and Parageobacillus* genomospecies 1 (NUB3621) (Aliyu et al. 2016). The ability of *Geobacillus* spp. to metabolize C6 and C5 sugars, starchy substrates and xylans is of particular interest for their application in biomass conversions (Studholme 2015; Hussein et al. 2015).

A range of recently reported applications of geobacilli, including the production of several (heterologous) proteins, fuels and chemicals, underpin the potential of this genus (Lentini et al. 2007; Bartosiak-Jentys et al. 2013; Bartosiak-Jentys et al. 2012; Cripps et al. 2009; Thompson et al. 2008; Xiao et al. 2012). Still, a greater understanding of their genomes, metabolism and the development of a robust genetic toolbox would facilitate the development of *Geobacillus* spp. into valuable whole-cell catalysts. Several strains of (*Para*) *Geobacillus* have proven to be genetically accessible, but the genomes of *Parageobacillus stearothermophilus* NUB3621, *Parageobacillus thermoglucosidans* DSM2542 and *Geobacillus kaustophilus* HTA426 remain the only genetically accessible strains that are publicly available to date (Studholme 2015; Kananavičiute & Čitavičius 2015). Although an impressive number of genome sequences have been completed in the past few years, the genomic information on geobacilli is still limited.

Here, we describe the complete genome sequence of G. thermodenitrificans T12, which is accessible in the GenBank

database (CP020030-CP020032). Strain T12 was isolated from compost and was characterized for its ability to co-ferment C5 and C6 sugars, as well as its ability to produce lactic acid and acetic acid from beechwood xylan. Furthermore, a relatively efficient transformation protocol has been developed for strain T12 (Daas et al. 2016). In this paper, several genomic highlights are described, demonstrating the potential of strain T12 not only as an interesting whole-cell catalyst, but also as a valuable resource of useful genetic elements.

Methods

Media, Cultivation methods and DNA isolation

Geobacillus thermodenitrificans T12 was isolated from a compost heap at Recom Ede B.V. in the Netherlands (52.043941 latitude° and 5.616682° longitude). It was demonstrated to co-ferment C6 and C5 sugars, to utilize xylan, and to be genetically accessible (Daas et al. 2016). The isolated organism is a facultative anaerobic thermophile capable of denitrification. It was grown in LB2 or MMy media at 65°C in a rotor shaker with an agitation speed of 150 RPM. Genomic DNA of strain T12 was isolated from 10 ml of a logarithmic growing culture of an OD_{600nm} of 0.87 by using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, Wisconsin, USA) according to manufacturer's protocol.

LB2 contains per liter: 10 g tryptone (Oxoid), 5 g yeast extract (Oxoid), 10 g NaCl and salts mix consisting of 0.25 g K_2HPO_4 (after autoclaving); 1 g NH_4Cl; 3 g NaCl; 1.50 g Na_2SO_4; 0.08 g $NaHCO_3$; 1 g KCl; 1.8 g $MgCl_2 \times 6H_2O$ and 0.30 g $CaCl_2 \times 2H_2O$; MOPS was added as a buffering agent and pH was set to 7.04 at room temperature. The medium was autoclaved for 20 min at 121 °C after which 1 ml filter sterile metal mix and 1 ml filter sterile vitamin solution were added. Metal mix contains per liter: 1.60 g $MnCl_2 \times 6H_2O$; 0.1 g $ZnSO_4$; 0.2 g H_3BO_3; 0.01 g $CuSO_4 \times 5H_2O$; 0.01 g $Na_2MoO_4 \times 2H_20$; 0.1 g $CoCl_2 \times 6H_2O$; 0.7 g $FeSO_4 \times 7H_2O$; 5 g $CaCl_2 \times 2H_2O$ and 20 g $MgCl_2 \times 6H_2O$. Vitamin mix contains per liter: 0.1 g thiamine; 0.1 g riboflavin; 0.5 g nicotinic acid; 0.1 g pantothenic

acid; 0.5 g pyridoxamine, HCl; 0.5 g pyridoxal, HCl; 0.1 g D-biotin; 0.1 g folic acid; 0.1 g *p*-aminobenzoic acid; 0.1 g cobalamin.

MMy medium contains per liter: 0.5 g yeast extract (Oxoid), 0.52 g K_2HPO_4; 0.23 g KH_2PO_4; 0.34 g NH_4Cl; 8.37 g MOPS was used as a buffering agent and pH was set to 7.04 at room temperature. After autoclaving of the medium 0.14 g $CaCl \times 2H_2O$; 1 ml vitamin mix and 1 ml metal mix were added.

Glucose (10 g/l) was used as carbon source (unless stated otherwise) and separately autoclaved. Pectic substrates were used at a concentration of 5 g/L and were sterilized by dry incubation at 120 °C for 30 minutes. For plate cultures, 5 g/l gelrite (Carl Roth, Karlsruhe, Germany) was added. API 50 CHB/E test was inoculated from a 10mL overnight LB2 culture of T12. Cells from the LB2 culture were centrifuged at 4000xg for 10 min and then transferred to the API 50 CHB/E Medium to obtain a cell density of 0.24 AUs at 600nm. After inoculation, the API 50 CHB/E strips were inoculated for 48h at 60°C.

Genome sequencing, assembly and annotation

The *G. thermodenitrificans* T12 genome was sequenced, assembled and annotated by Baseclear B.V. (Leiden, The Netherlands). Pair-end sequence reads were generated by using the Illumina HiSeq2500 system followed by PacBio sequencing using the PacBio RS system. The quality of the Illumina FASTQ sequences was enhanced by trimming off low-quality bases with the "Trim sequences" option of the CLC Genomics Workbench version 7.0.4. Subsequently, the quality-filtered sequence reads were puzzled into several contig sequences. The data collected from the PacBio RS instrument were processed and filtered with the SMRT Analysis software suite. Filtering of the Continuous Long Read (CLR) data was done by Read-length (>50), Subread-length (>50) and Read quality (>0.75). The average read length of the paired-end reads was 254.15 bp and those of PacBio were 2795 bp. Analysis of methylation patterns was included in the SMRT portal (v2.3), using the RS_Modification_and_Motif_Analysis.1 workflow. A summary of the discovered motifs is given in Table 1.

The assembly has been performed with the "De novo assembly" option of the CLC Genomics Workbench version 7.0.4. and the optimal k-mer size was automatically determined with KmerGenie (Chikhi & Medvedev 2014). The assembled contigs were linked and placed into super-scaffolds based on the alignment of the PacBio CLR reads. This alignment was performed with BLASR (Chaisson & Tesler 2012). From the alignment the orientation, order and distance between the contigs were estimated by using the SSPACE-LongRead scaffolder version 1.0 (Boetzer & Pirovano 2014). The gapped regions within the super-scaffolds were closed in an automated manner by GapFiller version 1.10 (Boetzer & Pirovano 2012). Closing the gaps resulted in 7 scaffolds: one that covers the T12 genome, two that cover the plasmids and 3 contaminant scaffolds that were assembled and then removed from the dataset.

Genome annotation was performed on the assembled scaffolds by using the BaseClear annotation pipeline, which is based on the Prokka Prokaryotic Genome Annotation System (**http://vicbioinformatics.com/**). Genes were predicted with Prodigal V2 and translated CDSs were used to search the Uniprot database to identify EC number, CAZY number and function annotation. A domain analysis was performed with *hmmer-3* and signal peptides, with their corresponding cellular localisation, were predicted by using *signalP v4.0*. To predict rRNAs and tRNAs we used barrnap v0.2 and Aragorn v1.2.36, respectively.

The genome of *Geobacillus thermodenitrificans* T12 (Bioproject acc. Nr. PRJNA377291) encompasses a 3.64 Mb chromosome and two plasmids of 59 kb and 56 kb that together contain 3,676 genes and an average GC content of 48.71% (Table 2). Most genes predicted were assigned with a known function (2088 genes) and a total of 47 genes with a CAZy number were annotated (Table 3).

Table 1 Methylation analysis summary

MotifString[a]	CenterPos[b]	R - M Type[c]	ModificationType	Fraction[e]
5'-GATC-3'	1	III	m6A[d]	0.992
5'-TAAYNNNNNNRTTA-3'	2	I	m6A	0.983
5'-GCCAT-3'	3	II	m6A	0.957

a = Detected motif sequence for this site such as "GATC"
b = Position in motif of modification (0-based)
c = Type of restriction modification system (Wilson & Murray 1991)
d = N6-Methyladenosine (m6A) refers to a type of methyltransferase that prefers to methylate the adenosine base at the nitrogen-6 position
e = The percent of time this motif is detected as modified in the genome.

Table 2 Genome statistics

Attribute	Value
Chromosome size (bp)	3,640,708
Plasmid 1 size (bp)	58,808
Plasmid 2 size (bp)	56,976
GC percentage chromosome	49.10
GC percentage plasmid 1	38.91
GC percentage plasmid 2	39.40
Number of genes	3,676
Number of tRNA genes	94
Number of rRNA genes	30
Gram	+
Organism	*Geobacillus thermodenitrificans*

Table 3 Statistics of Coding Sequences (CDS)

Attribute	Value
Number of genes	3,676
with a known function	2,088
with a GO annotation	2,379
with a CaZy number	47
with a signal peptide	146
Gene density [gene/Mbp]	977
Total gene size	3,233,578
Max gene size	7,155
Min gene size	71
Average gene size	879

Insights to the genome

Geobacillus thermodenitrificans T12 is a rod-shaped, thermophilic, facultative anaerobic, Gram-positive bacterium that was isolated from compost (Figure 1). The phylogeny of strain T12 was determined using its 16S RNA encoding gene. This sequence was used to create a phylogenetic tree that demonstrates its relationship to several *Geobacillus* type strains (Daas et al. 2016). *G. thermodenitrificans* T12 was selected for whole genome sequencing based on its fermentation product profile, co-utilisation of glucose and xylose, its ability to degrade xylan and its amenability to transformation.

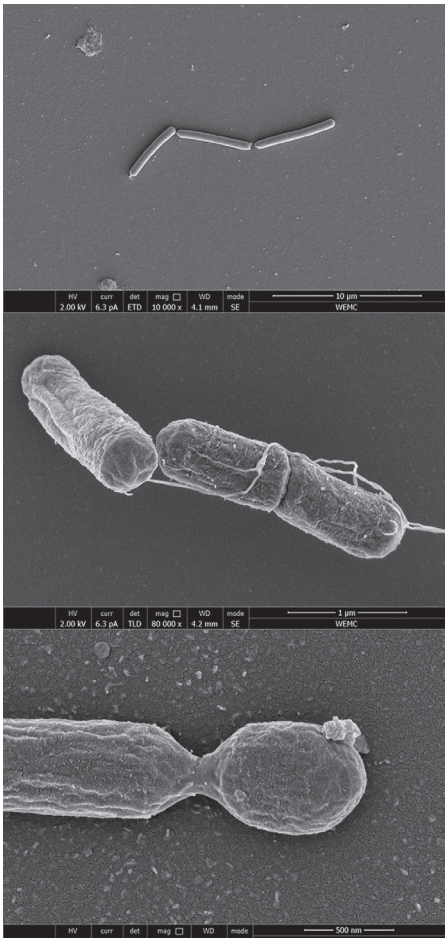


Figure 1 Scanning electron micrographs of *G. thermodenitrificans* T12

General metabolism

T12 can utilize a wide variety of carbohydrates (Table S1) as was proven by fermentation test with an API 50 CHB/E test kit (BioMérieux, Marcy l'Etoile, France). We also determined the presence of complete pathways for the synthesis of purine and pyrimidine, all 20 amino acids and several vitamins except D-biotin. Pathways for the synthesis of these compounds are identical to those identified for *G. thermodenitrificans* NG80-2 (KEGG entry: T00496). The D-biotin pathway in strain T12 lacks the genes required for the conversion of Pimeloyl-(acp) to Dethiobiotin of the biotin pathway, however, a putative biotin transporter encoding gene (GTHT12_02457) was found.

Denitrification

Geobacillus thermodenitrificans T12 is capable of anaerobic respiration, reducing nitrate to molecular nitrogen (Daas et al. 2016). The metabolic pathways of strain T12 involved in this reduction have been identified based on the annotated genome of *Geobacillus sp.* NG80-2, which was shown to be capable of denitrification. The genes encoding all enzymes of the denitrification pathway have been identified in its genome (Feng et al. 2007). Two copies of the nitrate reductase operon *narGHJI*, responsible for the reduction of nitrate to N_2O, are present in the genomes of both strain T12 and strain NG80-2. The homology of genes involved in both copies of the *narGHJI* operon is 99% to 100% sequence identity between the two strains. Not only does the denitrification cluster of T12 show high identity to their homologs of strain NG80-2, also their genomic localisation is identical to the genomic localisation in strain NG80-2.

The *nos* gene cluster in strain NG80-2 is required for the reduction of N_2O to N_2, but this cluster was not described for gram positive bacteria until its discovery in strain NG80-2 (Feng et al. 2007). We here reported the identification of genes encoding all enzymes for denitrification in strain T12 with 97% to 100% sequence identity to their orthologs of strain NG80-2.

HUS locus

The hemicellulose utilisation (HUS) locus of strain T12 contains 95 genes with a total size of 121 kb and is thereby the largest hemicellulolytic cluster reported for *Geobacillus* spp. (De Maayer et al. 2014). The HUS locus shows extensive variability among geobacilli, even among strains of the same species. This variability might be induced by the action of mobile genetic elements and the selection for desired traits based on the environment from which *Geobacillus* spp. are isolated.

When the HUS locus of strain T12 is compared to the HUS locus of *Geobacillus stearothermophilus* T-6, which is the most extensively studied HUS locus among *Geobacillus* spp., all T-6 genes are present in the T12 cluster, except for the arabinan degradation cluster. The arabinose metabolic genes of strain T12 show highest similarity to *Geobacillus* sp. 1MC16 with both strains lacking the gene cluster containing the extracellular and intracellular arabinase, although both do contain the enzymes needed for arabinose and arabinooligosaccharide degradation (Brumm et al. 2015).

Compared to T-6, the HUS locus of strain T12 has an additional inositol pathway, an oligopeptide-transport gene cluster, and a pectate degradation pathway. The inositol gene cluster (*iolG/DEBCA*) of strain T12 has a 93-100% amino acid identity to the inositol clusters of *G. thermodenitrificans* strains NG80-2, DSM465 and G1MC16 (Brumm et al. 2015). Directly upstream of the inositol cluster in strain T12, an oligopeptide-transport gene cluster (*dppABCDFE*) is found which is not present in any other sequenced *G. thermodenitrificans* strain. This peptide-transport gene cluster is identical to its orthologs found in multiple other *Geobacillus* spp. (Brumm et al. 2015). *G. thermodenitrificans* strains T12, NG80-2, DSM465 and G1MC16 also encode an α-mannosidase at the 5'-end of their HUS-locus. Surprisingly, a 23.3Kb insert was found in between the mannosidase-encoding gene and the polypeptide transport gene cluster in strain T12. The insert contains a pectate degradation gene cluster containing both a pectate lyase (PL1) and a rhamnogalacturonyl hydrolase (GH88). Furthermore, genes GTHT12_1416 to GTHT12_1424 show high sequence identity to the rhamnogalacturonan degradation pathway

of *Bacillus* spp. (Table S2). When grown on various pectic substrates, strain T12 is capable of growth and organic acid production on both rhamnogalacturonan I and galactan. In contrast, strain *G. thermodenitrificans* DSM465 cannot grow by converting these substrates and lacks the pectate degradation gene cluster in its HUS locus (Figure 2). We conclude that the ability of strain T12 to ferment both rhamnogalacturonan and galactan most likely is the result of the pectate degradation cluster in its HUS locus.

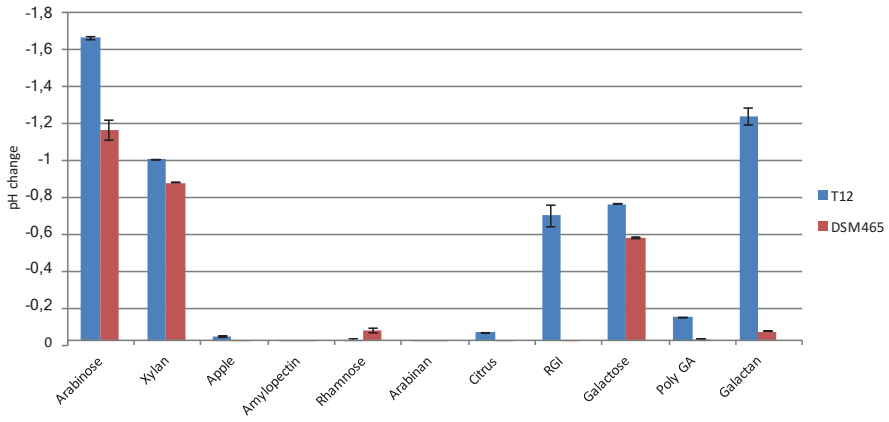


Figure 2 Acidification of MMy medium due to fermentation of various pectic substrates by *G. thermodenitrificans* strains T12 and DSM465. Cultures were incubated at 65 °C for 24 h in a rotary shaker at 150 RPM. Xylan: beechwood xylan, Apple: apple pectin; Citrus: citrus pectin, RGI: rhamnogalacturonan I, Poly GA: poly galactic acid.

Additional CAZymes

Next to the carbohydrate active enzymes (CAZymes) present on the HUS locus, several additional genes have been localized that encode for CAZymes. Most notable are a starch degrading cluster (GTHT_00158 through GTHT_00163) encoding for both an intracellular α-amylase and an extracellular α-amylase, interspaced by a three component ABC transporter. In addition, two oligo-1,6-glucosidases were identified, located far apart on the genome. These oligo-1,6-glucosidases are predicted to degrade the oligosaccharides obtained from the degradation of starch by alpha-amylases, thereby complementing the starch degradation pathway. Besides the starch degradation pathway, we identified a 6-phospho-β-

glucosidase (GTHT_01331) just upstream of a putative cellobiose specific phosphotransferase uptake mechanism (GTHT_01332 through GTHT_01334). Additionally, three more β-glucosidases were identified scattered over the genome (GTHT_01847, GTHT_02694 and GTHT_2696). Two proteins annotated as peptidases (GTHT_02208 and GTHT_02311) showed high identity to proteins (EPR27003.1 and EPR26354.1, respectively) that were characterized as endoglucanases (Rastogi, 2010). However, although strain T12 demonstrated growth on carboxymethylcellulose, we were unable to detect degradation of this substrate (data not shown). Furthermore, on plasmid pGeo12b, a levanase encoding gene has been identified (GTHT_3754) located in between a three component ABC transporter and a fructokinase, suggesting the ability of *G. thermodenitrificans* T12 to degrade fructans.

Host defense systems

On the genome of T12, multiple methylation motifs were discovered and the corresponding restriction-modification (R/M) systems were identified using the REBASE database (Table 1) (Roberts et al. 2015). A Type II R/M system was identified on plasmid pGeo12b, encoded by ORFs GTHT12_03786 through GTHT12_03788. Methylation pattern analysis on the T12 genome revealed 5'-GATC-3' as potential Type II R/M system recognition sequence (Table 1). ORFs GTHT12_00809 and GTHT12_00810 encode a Type III R/M system that may recognize the 5'-GCCAT-3' sequence retrieved from the methylation analysis. This Type III system is most closely related (99% amino acid identity) to a system in *Geobacillus* sp. PA-3. PacBio data also showed methylation at motif 5'-TAAYNNNNNNRTTA-3', which is a typical Type I R/M system recognition sequence. Although we did find a putative Type I methylase (GTHT_3783), we were unable to discover the corresponding R and S subunits. The mentioned R/M systems might influence the genetic accessibility and most likely play a crucial role in phage resistance in the microbiome of the compost from which the T12 strain was isolated.

In addition to the R/M systems, we discovered a Type II-C CRISPR-Cas system (called GtCas9 hereafter) on the genome of strain T12

(GTHT12_03309 through GTHT12_03401) (van der Oost et al. 2016). The CRISPR-Cas locus architecture is a typical CRISPR-Cas Type II-C system (Chylinski et al. 2014), but contains a *cas6* gene located after the CRISPR array (Figure 3). This *cas6* gene is likely to be a remnant of a Type I or Type III CRISPR-Cas system and is believed not to be part of the Cas9 locus in strain T12. The CRISPR-finder tool (http://crispr.i2bc.paris-saclay.fr/Server/) was used to identify the CRISPR array that contains 11 repeat sequences (36 bp) interspaced by 10 spacer sequences (29-31 bp). The tracrRNA sequence was predicted by searching for sequences with strong complementarity to the repeat sequences in a 1-kb window flanking the CRISPR locus. A 34 bp tracr sequence (TCATAGTAACCCTGAGATCATTGCTGTGGTATAA) was found 164 bp upstream of the *cas9* gene. The tracrRNA, which can pair with the repeat sequence of the crRNA, is essential to crRNA maturation in this system (Jinek et al. 2012).

Spacers were blasted against all available databases of CRISPRTarget (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html) and were found to have hits against *Anoxybacillus flavithermus* WK1 phage DNA and several spacers matched against *Geobacillus* Virus E2 DNA. Virus E2 is present as prophage in the genomes of *G. thermodenitrificans* NG80-2 and *P. thermoglucosidans* C56-YS93 but was not found in the genome of *G. thermodenitrificans* T12.

Figure 3 CRISPR-Cas Type II-C system architecture of *G. thermodenitrificans* T12. Rectangles: repeats; diamonds: spacers; dashed arrow: predicted promoter

To date there is no experimental evidence for active Cas9 proteins in thermophiles. Based on a comparative genome screening on the presence of Cas9 in bacteria, it was found that the Type II-C CRISPR-Cas system is only present in approximately 3.3% of all bacterial genomes (Chylinski et al. 2014). The classification and evolution of

Type-II CRISPR-Cas systems has previously been described and, in particular, three thermo-tolerant species have been identified which exhibit these Type-II systems, however, these species grow optimally between 50-60°C and do not exhibit strictly thermophilic growth; by definition thermophiles have an optimum growth temperature above 60 °C (Chylinski et al. 2014). Comparative analysis to the non-redundant protein database of the NCBI revealed that Cas9 proteins are widespread among *Geobacillus spp*. A phylogenetic tree reveals the close relatedness of Cas9 proteins among thermophiles and their distinct mesophilic orthologs (Figure 4). Although mesophilic Cas9 proteins show little sequence identity to GtCas9, protein sequence alignment against the well characterized Cas9 proteins of *A. naeslundii* (Type II-C), *S. pyogenes* and *S. thermophilus* (Type II-A) reveals the conservation of important active site residues in GtCas9 (Figure S1). CRISPR-Cas is often used for genome editing and the engineering toolbox expands rapidly (Mougiakos et al. 2016; Singh et al. 2017).

Conclusions

The isolated *Geobacillus thermodenitrificans* T12 possesses several features that make this strain unique and certainly interesting for further research. Strain T12 has low nutritional requirements and based on its genome analysis we concluded that it contains complete pathways for the synthesis of purine and pyrimidine nucleotides, all 20 amino acids and several vitamins except D-biotin. The xylanolytic cluster of strain T12 is similar to the cluster described for *G. stearothermophilus* T6. In addition, the cluster from strain T12 extends that of T6 by a complete inositol and ribose metabolism pathway but lacks genes encoding arabinan degrading enzymes. *G. thermodenitrificans* T12 is capable of denitrification and the involved genetic pathways share high identity with the genes described for strain NG80-2. Both organisms can reduce nitrate to dinitrogen, allowing for growth under oxygen-limited/anaerobic conditions. In addition to the metabolic features described, strain T12 contains R-M systems of Type II and III. In addition, we showed the presence of a CRISPR-Cas Type II-C system. The ability of the organism to grow

and ferment a wide range of substrates at temperatures between 50-70°C and the potential of the CRISPR-Cas system for genome editing make this organism into a potential interesting candidate for industrial fermentations.

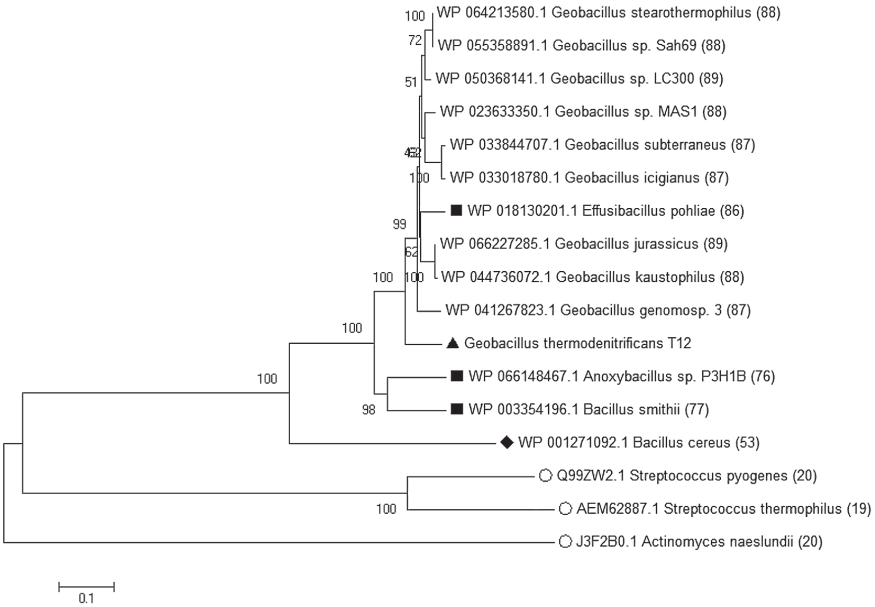


Figure 4 Neighbour-Joining tree of Cas9 protein sequences. The evolutionary history was inferred using the Neighbor-Joining method (Saitou & Nei 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). All sequences found in *Geobacillus* spp. were included, as well as currently well-characterized sequences (Open circles: *S. pyogenes, S. thermophiles* and *A. naeslundii*), as well as the closest non-thermophilic species *Bacillus cereus* (closed diamond). Non-*Geobacillus* strains capable of thermophilic growth have been included (closed squares). For all sequences, the percentage of amino acid sequence identity to T12 is indicated after the strain name between brackets.

Acknowledgements

We thank Fons Janssen for technical assistance and Jasper Koehorst for his assistance in bioinformatics analysis. This work was financially supported by BE-Basic and Corbion as part of the BE-Basic C3-acids project.

Supplementary information

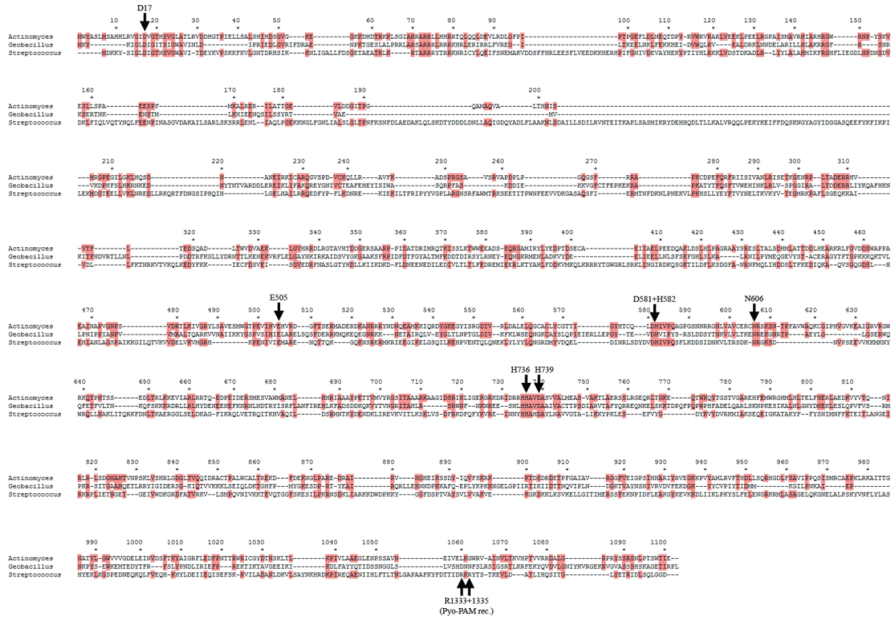


Figure S1 Protein sequence alignment for T12-Cas9 (Type II-C) with well-characterized Type II-C (A. naeslundii/'ana') and Type II-A (S. pyogenes/'pyo' and S. thermophilus) Cas9 sequences. Important active site residues are well conserved and indicated with black arrows. Protein domains as described for Ana-Cas9 and Pyo-Cas9 (Jinek et al. 2014) are indicated with shaded boxes and similarly coloured letters. The PAM recognition domain has been determined for the *S. pyogenes* Type II-A system, but not for any Type II-C system and is therefore only indicated in the *S. pyogenes* sequence

Table S2 Fermentation test results after 48h using API 50 CHB/E. Red: no acidification, orange: slight acidification, green: acidification

0	Control		25	Esculin ferric citrate	
1	Glycerol		26	Salicin	
2	Eryhritol		27	D-Cellobiose	
3	D-Arabinose		28	D-Maltose	
4	L-Arabinose		29	D-Lactose (bovine origin)	
5	D-Ribose		30	D-Melibiose	
6	D-Xylose		31	D-Saccharose (sucrose)	
7	L-Xylose		32	D-Trehalose	
8	D-Adonitol		33	Inulin	
9	Methyl-β-D-Xylopyranoside		34	D-Melezitose	
10	D-Galactose		35	D-Raffinose	
11	D-Glucose		36	Amidon (starch)	
12	D-Fructose		37	Glycogen	
13	D-Mannose		38	Xylitol	
14	L-Sorbose		39	Gentiobiose	
15	L-Rhamnose		40	D-Turanose	
16	Dulcitol		41	D-Lyxose	
17	Inositol		42	D-Tagatose	
18	D-Mannitol		43	D-Fucose	
19	D-Sorbitol		44	L-Fucose	
20	Methyl-α-D-Mannopyranoside		45	D-Arabitol	
21	Methyl-α-D-Glucopyranoside		46	L-Arabitol	
22	N-Acetylglucosamine		47	Potassium gluconate	
23	Amygdalin		48	Potassium 2-ketogluconate	
24	Arutin		49	Potassium 5-ketogluconate	

Table S3 The pectate degradation cluster of *Geobacillus thermodenitrificans* T12

Locus tag	Gene name	Product size (aa)	Subcellular Localization	Predicted protein function	Closest non-*Geobacillus* ortholog	Strain	AA identity (%)	Bitscore	E-value
GTHT12_01416	*kduI1*	277	-	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase 1	*Bacillus* sp.	LF1	79	472	5.00E-167
GTHT12_01417	*kduD*	256	-	2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase	*Bacillus* sp.	JCM 19034	74	395	2.00E-137
GTHT12_01418	*GlpR*	254	-	Putative transcriptional repressor	*Bacillus* sp.	SA1-12	62	308	5.00E-103
GTHT12_01419	-	404	-	putative response regulatory protein	*Bacillus novalis*	NBRC 102450	41	286	2.00E-90
GTHT12_01420	*ypdA_1*	599	-	Sensor histidine kinase	*Bacillus novalis*	NBRC 102450	54	651	0.00E+00
GTHT12_01421	*pel*	442	Extracellular	Pectate lyase	*Bacillus* sp.	TS-47	100	916	0.00E+00
GTHT12_01422	*yesR*	377	-	Unsaturated rhamnogalacturonyl hydrolase YesR	*Bacillus novalis*	NBRC 102450	65	527	0.00E+00
GTHT12_01423	-	449	Cytoplasmic Membrane	ABC-type transport systems, binding protein	*Bacillus novalis*	NBRC 102450	67	654	0.00E+00
GTHT12_01424	*ugpE*	279	-	ABC transporter protein	*Bacillus* sp.	UNC41MFS5	77	459	5.00E-162
GTHT12_01425	-	315	-	ABC-type transport systems, permease protein	*Pelosinus fermentans*	JBW45	69	468	4.00E-164
GTHT12_1502	*yesY*	273		putative rhamnogalacturonan acetylesterase YesY	*Anoxybacillus geothermalis*	GSSed3	83	435	6.00E-153

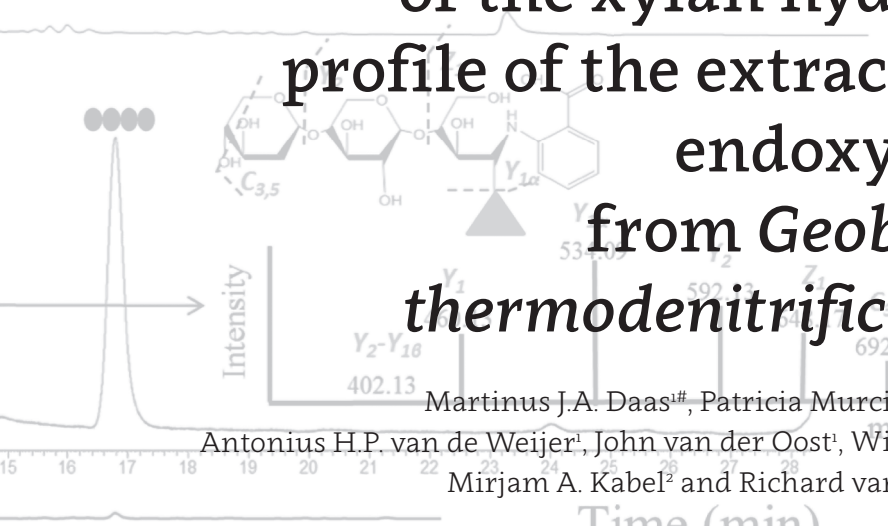
Table S4 Additional CAZymes identified on the genome of *G. thermodenitrificans* T12

Locus tag	Gene name	Product size (aa)	Subcellular Localization	Predicted protein function	Closest non-*Geobacillus* ortholog	Strain	AA identity (%)	Bitscore	E-value
GTHT12_00075	*malL_1*	556	-	oligo 1,6-glucosidase	*Anoxybacillus geothermalis*	-	92	2796	0.00E+00
GTHT12_00158	-	589	Extracellular	Alpha- amylase	*Thermus* sp.	YBJ-1	87	2842	0.00E+00
GTHT12_00162	-	512	Cytoplasm	Alpha- amylase	*Anoxybacillus amylolyticus*	MR3C	92	2479	0.00E+00
GTHT12_01209	*malL_2*	564	-	oligo 1,6-glucosidase	*Anoxybacillus geothermalis*	-	93	2819	0.00E+00
GTHT12_01331	*bglA*	471	-	Beta-glucosidase	*Anoxybacillus geothermalis*	-	95	2401	0.00E+00
GTHT12_01664	*aga*	730	-	alpha-galactosidase	*Parageobacillus* genomosp.	1	81	3267	0.00E+00
GTHT12_01847	*arbB*	456	-	6-phospho-beta-glucosidase	*Anoxybacillus* sp.	SK3-4	95	2262	0.00E+00
GTHT12_02317	*pulA*	727	-	Pullulanase	Mixed culture bacterium	AmyA1	99	3873	0.00E+00
GTHT12_02694	*chbF*	448	-	6-phospho-beta-glucosidase	*Anoxybacillus geothermalis*	-	95	2262	0.00E+00
GTHT12_02696	*bglH*	479	-	6-phospho-beta-glucosidase	*Parageobacillus* genomosp.	1	88	2299	0.00E+00
GTHT12_02766	-	566	-	Sucrose phosphorylase	*Parageobacillus* genomosp.	1	82	2535	0.00E+00
GTHT12_02767	*mngB*	888	-	Alpha-mannosidase	*Parageobacillus* genomosp.	1	77	3795	0.00E+00
GTHT12_03754	-	494		Levanase	*Parageobacillus toebii*	-	99	2610	0.00E+00

A
B
B'
C
Response (nC)
UV-Signal
0
13
14

Chapter 4

Biochemical characterization of the xylan hydrolysis profile of the extracellular endoxylanase from *Geobacillus thermodenitrificans* T12



Martinus J.A. Daas[1#], Patricia Murciano Martínez[2#], Antonius H.P. van de Weijer[1], John van der Oost[1], Willem M. de Vos[1], Mirjam A. Kabel[2] and Richard van Kranenburg [1,3]

Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, the Netherlands[1];
Laboratory of Food Chemistry, Wageningen University, Bornse Weilanden 9, 6708 WG Wageningen, the Netherlands[2];
Corbion, Arkelsedijk 46, 4206 AC Gorinchem, the Netherlands[3].

# These authors contributed equally

This chapter was adapted from:

Daas, M. J. A., Murciano Martinez, P., Van de Weijer, A. H. P., Van der Oost, J., De Vos, W. M., Kabel, M. A. & Van Kranenburg, R. Biochemical characterization of the xylan hydrolysis profile of the extracellular endoxylanase from *Geobacillus thermodenitrificans* T12. *BMC Biotechnol. 17:44, (2017)*.

Abstract

Background: Endoxylanases are essential in degrading hemicellulose of various lignocellulosic substrates. Hemicellulose degradation by *Geobacillus* spp. is facilitated by the hemicellulose utilization (HUS) locus that is present in most strains belonging to this genus. As part of the HUS locus, the *xynA* gene encoding an extracellular endoxylanase is one of the few secreted enzymes and considered to be the key enzyme to initiate hemicellulose degradation. Several *Geobacillus* endoxylanases have been characterized for their optimum temperature, optimum pH and generation of degradation products. However, these analyses provide limited details on the mode of action of the enzymes towards various substrates resulting in a lack of understanding about their hydrolytic potential.

Results: A HUS-locus associated gene (*Gt*XynA1) from the thermophile *Geobacillus thermodenitrificans* T12 encodes an extracellular endoxylanase that belongs to the family 10 glycoside hydrolases (GH10). The *GtxynA1* gene was cloned and expressed in *Escherichia coli*. The resulting endoxylanase (termed *Gt*XynA1) was purified to homogeneity and showed activity between 40°C and 80°C, with an optimum activity at 60°C, while being active between pH 3.0 to 9.0 with an optimum at pH 6.0. Its thermal stability was high and *Gt*XynA1 showed 85% residual activity after 1 h of incubation at 60°C. Highest activity was towards wheat arabinoxylan (WAX), beechwood xylan (BeWX) and birchwood xylan (BiWX). *Gt*XynA1 is able to degrade WAX and BeWX producing mainly xylobiose and xylotriose. To determine its mode of action, we compared the hydrolysis products generated by *Gt*XynA1 with those from the well-characterized GH10 endoxylanase produced from *Aspergillus awamori* (*Aa*XynA). The main difference in the mode of action between *Gt*XynA1 and *Aa*XynA on WAX is that *Gt*XynA1 is less hindered by arabinosyl substituents and can therefore release shorter oligosaccharides.

Conclusions: The *G. thermodenitrificans* T12 endoxylanase, *Gt*XynA1, shows temperature tolerance up to 80°C and high activity to a variety of xylans. The mode of action of *Gt*XynA1 reveals that arabinose substituents do not hamper substrate degradation by *Gt*XynA1. The extensive hydrolysis of branched xylans makes this enzyme particularly suited for the conversion of a broad range of lignocellulosic substrates.

Introduction

Many *Geobacillus* species contain a hemicellulose utilization (HUS) locus in their genome that encodes for multiple hydrolytic enzymes and sugar transporters responsible for the degradation of hemicellulose (Shulami et al. 2007; De Maayer et al. 2014). The most well characterized HUS locus belongs to *G. stearothermophilus* T-6 and its proposed degradation mechanism is based on a limited number of secreted hydrolytic enzymes and further processing of the generated degradation products intracellularly (Shulami et al. 2014). One of these secreted hydrolytic enzymes is an endoxylanase that cleaves the xylan backbone of hemicellulose into shorter oligosaccharides (Shulami et al. 2014; Khasin et al. 1993).

The xylan backbone is composed of xylosyl residues linked by ß-(1→4)-glycosidic bonds, and is typically decorated with a varying degree of substitutions containing (4-*O*-methyl)-α-D-glucuronic acids, α-L-arabinofuranosyl residues and acetyl residues (Ebringerová et al. 2005). The degree of substitution and polymerization is dependent on the origin of the substrate. Substrates from grasses and annual plants are rich in arabinoxylans, whereas hardwoods mostly consist of (4-*O*-methyl)-α-D-glucuronoxylan. This variety of substitutions requires multiple enzymes acting in synergy to fully degrade the xylans. Endo(1→4)-ß-D-xylanases (E.C. 3.2.1.8) hydrolyse the ß-(1→4)-xylosidic bonds to release xylose and (substituted) xylooligosaccharides of various lengths that can be degraded further by other hemicellulases.

Endoxylanases belonging to glycoside hydrolase family 10 and 11 are the best characterized and described (Shallom & Shoham 2003). Endoxylanases belonging to family 10 (GH10) can tolerate the presence of substituents, such as arabinosyl residues, better than those belonging to glycoside hydrolase family 11 (GH11) (Beaugrand et al. 2004). An example of the mode of action of two endoxylanases, belonging to families GH10 and GH11, from *Aspergillus awamori* (*Aa*) towards an arabinoxylan model substrate has been described (Kormelink et al. 1993). The *Aa*GH10 was found to be able to cleave the xylan backbone next to a xylosyl residue substituted with one

or two arabinosyl residues. In contrast, xylanases from family GH11 are hindered by arabinosyl substituents and cleave the xylan backbone only next to an unsubstituted xylosyl residue, producing longer oligosaccharides in comparison to those obtained from GH10 (Kormelink et al. 1993).

Endoxylanases from *Geobacillus* species belong to the GH10 family and have been described in multiple studies (Bhalla et al. 2014; Canakci et al. 2012; Gerasimova & Kuisiene 2012; Irfan et al. 2016; Gao et al. 2009; Marcolongo et al. 2015; Mitra et al. 2015; Verma et al. 2013). These studies mainly focused on the characterization of activity ranges at higher temperatures and thermo-stability but lack detailed analyses of the mode of action of the enzyme on its substrate. Although some studies have described xylo-oligosaccharides formation upon hydrolysis of xylan by the endoxylanase, the position of hydrolysis and the possible hindrance of substituents present on the xylan backbone have not been characterized (Bhalla et al. 2014; Irfan et al. 2016; Marcolongo et al. 2015; Verma et al. 2013).

In this paper, we describe the cloning and overproduction in *E.coli* of *Gt*XynA1, a GH10 family endoxylanase isolated from *Geobacillus thermodenitrificans*. We also determined the enzyme's mode of action by identifying the generated hydrolysis products and propose a model for the degradation of substituted xylans.

Methods

Media, strains and cultivation methods

Wheat arabinoxylan (WAX) was of medium viscosity and supplied by Megazyme (Wicklow, Ireland). The chemical composition of this type of WAX has previously been described (Van Gool et al. 2011). Beechwood xylan (BeWX) and all chemicals were purchased from Sigma-Aldrich (St Louis, MO, USA) unless otherwise specified. *Aa*XynA was kindly provided by the laboratory of Food Chemistry (Wageningen University) as described elsewhere (Kormelink et al. 1993).

Luria-Bertani (LB) medium contained per L: 10 g NaCl, 10 g tryptone (Oxoid), 5 g yeast extract (Roth). LB2 contains per liter: 10 g tryptone (Oxoid), 5 g yeast extract (Roth), 10 g sodium chloride and salts mix consisting of 1 g NH_4Cl; 3 g NaCl; 1.50 g Na_2SO_4; 0.08 g $NaHCO_3$; 1 g KCl; 1.8 g $MgCl_2 \times 6H_2O$; 0.30 g $CaCl_2 \times 2H_2O$. pH was set to 6.6 at room temperature and the medium was autoclaved for 20 minutes at 121°C, after which 1 mL K_2HPO_4 (250 g/L) was added.

For plasmid propagation and protein expression *E. coli* strains DH5α and BL21(DE3) were used respectively. *E. coli* strains were grown in LB medium at 37°C, pH 7.0 with an agitation speed of 150 RPM.

Geobacillus thermodenitrificans T12 was isolated from compost (Daas et al. 2016). Strain T12 was grown in LB2 medium at 65°C, 150 RPM and pH 7.0.

Cloning of *gtxynA1*

Genomic DNA of strain T12 was obtained using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre) according to manufacturer's protocol. The XynA coding sequence was cloned without the signal peptide, which was determined by SignalP4.1 to consist of the first 28 amino acids of the protein. Primers used for amplification of the *xynA* gene (GenBank accession number KX962565) were XynA-FW (5'-GCGCTCATGAAAACTGAACAATCATACGCTAAAAAG-3') and XynA-RV (5'-GCGCCTCGAGCTTATGATCGATAATAGCCCAATACG-3') that contain the *Bsp*HI and *Xho*I restriction sites, respectively. PCR amplification was performed on 50 ng of genomic DNA using PhusionHF DNA polymerase (Thermo Scientific) in the following conditions; 1 cycle 98°C 5 minutes., 30 cycles of 98°C 30 seconds, 50°C 30 seconds and 72°C 45 seconds followed by a final elongation step at 72°C for 5 minutes. For enzyme characterization studies *Gt*XynA1 was cloned into pET24d(+) using the restriction sites of *Bsp*HI and *Xho*I. The reverse primer was designed in such a way that the stop codon was removed to include the His-tag from the pET24d vector. Plasmids were purified from *E. coli* DH5α using the GeneJET™ Plasmid Miniprep Kit (Thermo Scientific) and subsequently transformed into *E. coli* BL21(DE3).

Overproduction and purification of *Gt*XynA1

Protein production was induced by adding 0.1 mM IPTG to a 250 mL culture at the start of inoculation. Culture was provided with 50 µg/mL kanamycin as antibiotics and incubated O/N at 37°C at an agitation speed of 150 RPM. Cells were then centrifuged (4000x*g*, 15 minutes, 4°C) and cell pellet was suspended in 25 mL phosphate buffer at a pH of 7.4. Cell suspension was disrupted using French Press and cell free extract (CFE) was obtained by centrifugation (30,000x*g*, 20 minutes, 4°C). Filtered CFE was applied to a HisPrep FF 16/10 Ni-NTA Sepharose column (GE Healthcare) equilibrated with 50 mM sodium phosphate buffer (pH 7.4) containing 300 mM NaCl. Bound *Gt*XynA1 was eluted using a linear gradient of 0-500 mM imidazole in the same buffer. Purified *Gt*XynA1 was then desalted by applying the protein solution to a HiPrep 26/10 desalting column (GE Healthcare) equilibrated with 50 mM sodium phosphate buffer (pH 7.4). Recombinant xylanase was eluted at a rate of 5 mL/minute and stored at -20°C for further research. Protein purity was assayed by SDS-PAGE analysis using a 10% (*w/v*) polyacrylamide gel according to the method of Laemmli (1970). A PageRuler™ protein ladder was used and bands were visualized using PageBlue™ Protein Staining Solution (Thermo Scientific).

Enzyme activity assays

Wheat arabinoxylan (WAX), Beechwood xylan (BeWX) and Birchwood xylan (BiWX) were incubated with the purified *Gt*XynA1 in 10 mM NaOAc buffer, pH 6.0 (1 mL, 10 mg substrate dry matter) at 60°C for 24 hours. The enzymes were dosed at 0.1% (w/w) protein based on substrate dry matter. Next, 2 µl of 4 M HCl was added to stop the enzyme activity and the sample was centrifuged (10,000x*g* , 10 minutes, 10°C) prior to analysis.

The pH and temperature optima were tested by incubating WAX with *Gt*XynA1 in a range of pH 3 to 9. For the pH range of 3 to 8 we used 50 mM sodium citrate buffer and for the pH range of 8 to 10 we used 200mM Tris-HCl buffer. Buffers were adjusted for correct pH at the temperature of incubation. Optimum temperature was determined in a range from 30 to 100°C. Incubation time, enzyme dosage and stopping the reaction were as described above.

All samples collected were submitted to the PAHBAH-assay, HPSEC, HPAEC and RP-UHPLC-UV-MS.

Analysis of hydrolysis products

The supernatants obtained were analyzed for the amount of reducing ends present by the PAHBAH reducing assay in duplicate (Lever 1972).

HPSEC

High performance size exclusion chromatography (HPSEC) of WAX and BiWX before and after incubation with *Gt*XynA1 was performed on an Ultimate 3000 HPLC system (Thermo Scientific, Sunnyvale, CA, USA) equipped with a set of three TSK-gel columns (6.0 mm x 15.0 cm per column) in series (SuperAW4000, SuperAW3000, SuperAW25000, Tosoh Bioscience, Stuttgart, Germany) in combination with a PWX-guard column (Tosoh Bioscience). HPSEC was controlled by the Chromeleon software (Thermo Scientific). Elution took place at 40°C with 0.2 M sodium nitrate at a flow rate of 0.6 mL/minute. The eluate was monitored using a refractive index (RI) detector (Shoko Scientific Co., Yokohama, Japan). Calibration was made by using pullulan series (Polymer Laboratories, Union, NY, USA) with a molecular weight in the range of 0.18-788 kDa.

HPAEC

Oligosaccharides and monosaccharides released after enzymatic incubation of WAX and BeWX with *Gt*XynA1 were analyzed by HPAEC as described elsewhere (Murciano Martínez et al. 2015).

RP-UHPLC-UV-MS

Xylo-oligosaccharides produced by the enzymes were 2-AA labelled and analyzed by RP-UHPLC-UV-MS as previously described (Murciano Martínez et al. 2015).

Results

In this study, we cloned and overexpressed the *Geobacillus thermodenitrificans* T12 *xynA1* gene in *E.coli* and determined the mode of action of the *Gt*XynA1 xylanase enzyme on wheat arabinoxylan (WAX), beechwood xylan (BeWX) and birchwood xylan (BiWX). The *xynA* gene is 1224 bp long and encodes a protein of 407 amino acid residues, which belongs to GH family 10. The amino acid sequence contains a signal peptide of 28 amino acids and two catalytic residues were predicted at positions H264 and D295. Phylogenetic analysis revealed high amino acid identity to several *Geobacillus* endoxylanases (Table 1).

Table 1 Sequence identity of several characterized endoxylanases to the *G. thermodenitrificans* T12 endoxylanase.

Origin	DNA identity (%)	AA identity (%)	Reference
G. thermodenitrificans T12	-	-	This study
G. thermodenitrificans JK1	99	100	(Gerasimova & Kuisiene 2012)
G. thermodenitrificans TSAA1	99	100	(Verma et al. 2013)
G. sp. TC-W7	99	99	(Liu et al. 2012)
G. sp. 71	91	91	(Canakci et al. 2012)
G. stearothermophilus T-6	84	84	(Khasin et al. 1993)

Cloning, expression and characterization of *Gt*XynA1

The gene coding for *Gt*XynA1 (1224 bp) was PCR-amplified, cloned into the pET24d vector and successfully expressed in *E.coli* BL21(DE3). The endoxylanase was purified to apparent homogeneity, as it appeared as a 50 kDa single band on SDS-PAGE after visualization with PageBlue stain (Figure S1; Figure S2; lane 6). The purified endoxylanase *Gt*XynA1 showed highest activity at pH 6.0 in a broad activity range from pH 4.0 to 9.0 (Figure 1A). *Gt*XynA1 showed activity between 40°C to 80°C with an optimum at 60°C (Figure 1B). The thermostability of *Gt*XynA1 was tested in the range of 30°C to 80°C. After 1 hour at 50 and 60°C the enzyme retained 90 and 45% residual activity, respectively, whereas the residual activity after 1 hour at 70°C was only 18% (Figure 2).

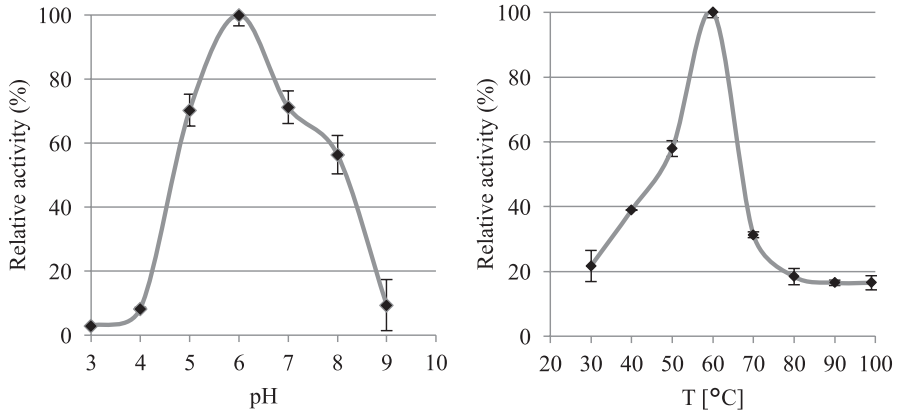


Figure 1 pH (A) and temperature (B) profiles of the incubated *Gt*XynA1 with wheat arabinoxylan (WAX).

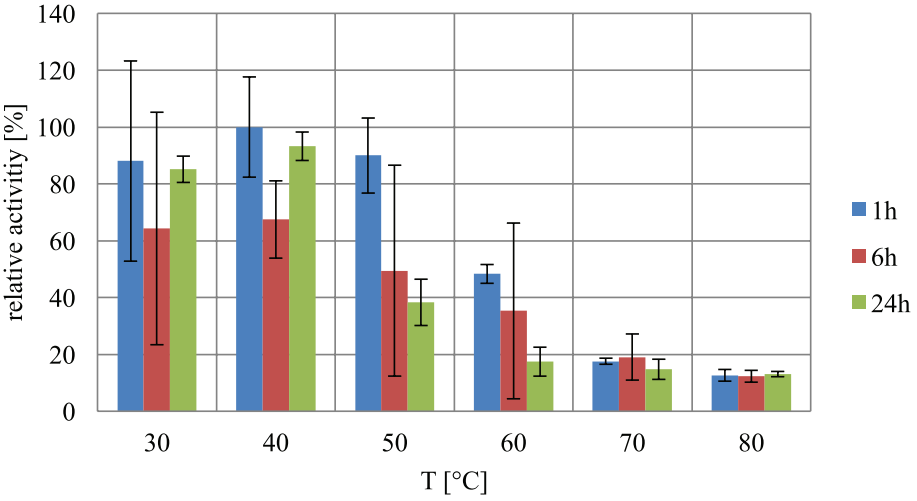


Figure 2 Temperature stability of *Gt*XynA1. The temperature stability was determined by incubating *Gt*XynA1 at temperatures between 30 and 80°C for 1 hour, 6 hours, and 24 hours after which the residual activity was measured against wheat arabinoxylan at 60°C and a pH of 6.0.

Mode of action

Activity of *Gt*XynA1 towards wheat arabinoxylan (WAX), birchwood xylan (BiWX), and beechwood xylan (BeWX) was determined as outlined in materials and methods. Degradation profiles of BiWX and BeWX were similar (data not shown), therefore only the BeWX

degradation patterns will be further explained. WAX contains β-(1→4) linked xylosyl residues substituted with α-(1→2) linked arabinosyl residues or with both α-(1→2) and α-(1→3) linked arabinosyl residues. BiWX and BeWX contain β-(1→4) linked xylosyl residues substituted with α-(1→2) linked 4-*O*-methylglucuronic acid (UA_{me}). Both WAX and BeWX were incubated with *Gt*XynA1 at its optimum pH and temperature for 24h (Figure 3). Both WAX and BeWX were degraded to oligomers and monomers, as deduced from the disappearance of soluble polymers (WAX around 55 kDa, BeWX around 12 kDa; Figure 3) and the increase of lower molecular weight products of around 1 kDa after 24 h incubation. The difference in molecular weight of the polymeric structures analyzed at 0 h incubation is because WAX is completely soluble and BeWX is only partly soluble (van Gool et al. 2012).

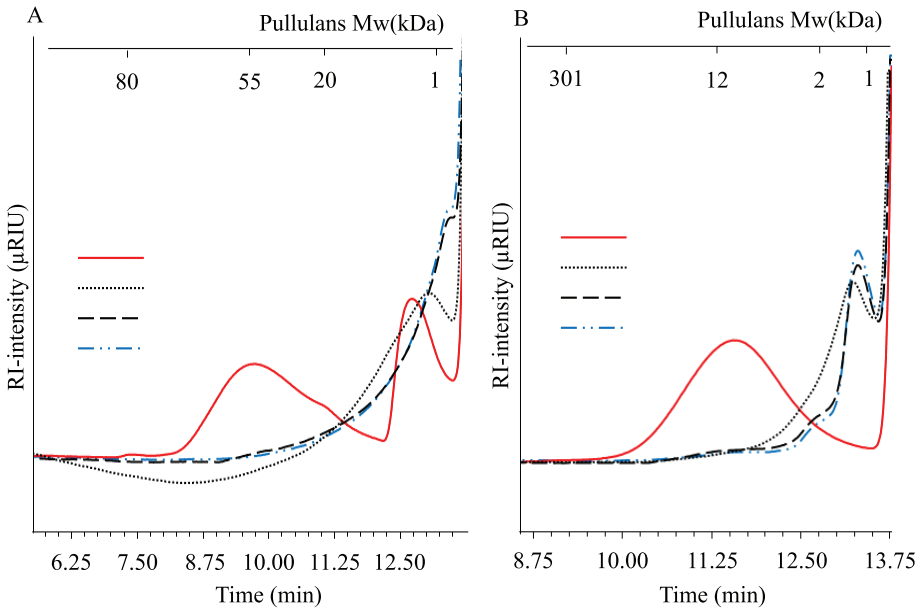


Figure 3 Molecular weight distributions of wheat arabinoxylan (WAX) (A) and beechwood xylan (BeWX) (B) incubated with *Gt*XynA1 for 0, 1, 6 and 24 hours.

*Gt*XynA1 incubated with WAX and BeWX released a series of linear and substituted xylo-oligosaccharides that were identified based on the known HPAEC elution of WAX digested with the

well-characterized GH10 endoxylanase from *Aspergillus awamori* (*Aa*XynA) (Kormelink et al. 1993). The oligomeric profile released by *Gt*XynA1 is comparable to the profile produced by *Aa*XynA, supporting that *Gt*XynA1 is a GH10 endoxylanase (Figure 4). Only one substituted oligosaccharide was produced from BeWX, eluting at 16 minutes (Figure 4), which was identified as a xylotriose substituted by 4-*O*-methylglucuronic acid located at the non-reducing end (Figure 4B') after 2-AA labelling and analysis by UPLC-UV-MS. The UV chromatogram at 340nm of 2-AA labelled degradation products from BeWX and the full MS spectra overlapped. The two predominant masses seen in the full MS spectra were 724 (m/z) and 666 (m/z) (data not shown), and they overlapped with UV chromatogram peaks eluting at 13 and 16.8 min, respectively. Figure 4 also shows that *Gt*XynA1 is able to degrade WAX and BeWX to smaller-sized linear xylo-oligosaccharides, mainly xylose and xylobiose. This observation is substantiated on the integrated peak area of linear xylooligosaccharides (Table 2). On the contrary, *Aa*XynA is able to degrade WAX and BeWX producing mainly xylobiose and xylotriose, but also xylotetraose, xylopentaose and xyloheaxaose (Table 2).

Table 2 HPAEC integrated peak area of linear xylo-oligosaccharides produced from WAX and BeWX after 24h incubation with *Gt*XynA1 and *Aa*XynA. ● = xylosyl residue; * = not detected

Enzyme tested	***Gt*XynA1**		***Aa*XynA**	
Substrate	WAX	BeWX	WAX	BeWX
Quantification	peak area			
Xylose (●)	87.8	165.6	4.4	17
Xylobiose (● ●)	192.7	457.9	17.5	157.1
Xylotriose (● ● ●)	*	15.5	10.6	131.7
Xylotetraose (● ● ● ●)	*	6.9	4.3	81.1
Xylopentaose (● ● ● ● ●)	*	*	1.2	25.3
Xylohexaose (● ● ● ● ● ●)	*	*	*	10.2

When comparing the diversity of substituted xylooligosaccharides produced (Figure 4), it is clear that *Gt*XynA1 produces shorter oligomers from both WAX and BeWX compared to *Aa*XynA. Substituted xylo-

oligosaccharides are identified in a range of DP from 2 to 6 by both HPAEC (Figure 4A) and MS (not shown), suggesting that *Aa*XynA is less tolerant towards 4-*O*-methylglucuronic acid substituted xylan.

Degradation products formed during the degradation of WAX by *Aa*XynA that where eluting at 15, 17 and 25 minutes are absent in the profile of *Gt*XynA1 incubated on WAX (Figure 4C, D). These xylo-oligosaccharides are characteristic because of the presence of two arabinosyl residues located at the *O*-2 and *O*-3 of the same xylosyl residue. Hence, *Gt*XynA1 cleaves next to an adjacent non-substituted residue, from the non-reducing end when a single substituted residue is present. In case the xylosyl residue is doubly substituted by arabinosyl residues, *Gt*XynA1 needs two adjacent unsubstituted residues at the non-reducing end. In case the cleavage takes place from the reducing end, the enzyme can cleave directly next to the substituted (single or doubly substituted) residue in most cases (Figure 4D, minutes 18, 22, 25.5). However, when the substituted residue (single or doubly substituted), is followed by a doubly substituted xylosyl residue, the enzyme cleaves next to an unsubstituted residue (Figure 4D, minutes 23, 27). On the contrary, the presence of more than one doubly or single substituted xylosyl residues obstructs the action of *Aa*XynA. Based on the information provided by the HPAEC chromatograms and RP-UV-UHPLC-MS profile on the linear and substituted xylo-oligosaccharides produced by the enzymes, an schematic mode of action of *Gt*XynA1 is proposed in Figure 4 B* and D*.

Discussion

The ability of *Geobacillus* species to degrade xylan relies on the activity of the hemicellulose utilization (HUS) locus. The HUS-locus contains a variety of xylan degrading enzymes, sugar transporters and metabolic pathways for the complete utilization of (substituted) xylans. Xylan degradation initiates by the action of an extracellular endoxylanase (XynA1) that cleaves the xylan backbone, thereby producing (substituted) xylo-oligosaccharides. The linear

oligosaccharides generated by the action of XynA1 can then enter the cell through an ABC-transporter that is encoded by the *xynEFG* operon whereas the substituted oligosaccharides enter the cell *via* other substrate specific ABC transporters (Shulami et al. 2007; Shulami et al. 2014; Shulami et al. 2011). In this research, we cloned and expressed *Gt*XynA1, the endoxylanase of *G. thermodenitrificans* T12 and analyzed its mode of action.

The purified endoxylanase *Gt*XynA1 showed highest activity at pH 6.0 in a broad activity range from pH 4.0 to 9.0 (Figure 1A), which is in line with previous reports on endoxylanases from geobacilli (Gerasimova & Kuisiene 2012; Gao et al. 2009; Verma et al. 2013). *Gt*XynA1 showed activity between 40°C to 80°C with an optimum at 60°C (Figure 1B) where previous reports on *Geobacillus* endoxylanases demonstrated optimum temperatures from 60°C up to 75°C (Gerasimova & Kuisiene 2012; Gao et al. 2009; Marcolongo et al. 2015; Verma et al. 2013). The thermostability of *Gt*XynA1 was tested in the range of 30°C to 80°C. After 1 hour at 50 and 60°C the enzyme retained 90 and 45% residual activity, respectively, whereas the residual activity after 1 hour at 70°C was only 18% (Figure 2). The endoxylanase characterized from *Geobacillus stearothermophilus* T-6 remains 90% active (1h time point) after incubation at 70°C (Khasin et al. 1993). This higher temperature stability of the T-6 endoxylanase may be caused by its production from the native strain instead of a recombinant expression system in *E.coli*. A lower thermostability of an endoxylanase produced in *E. coli* compared to the same protein produced in *Geobacillus* has been reported for *Geobacillus* sp. TC-W7 (Liu et al. 2012), which produces an endoxylanase highly similar to *Gt*XynA1 (Table 1). Alternatively, the difference in thermostability between *Gt*XynA1 and the T-6 endoxylanase may be (partly) caused by differences in their amino acid composition. The *Gt*XynA1 enzyme was more thermostable in comparison with other bacterial endoxylanases such as XynC from *Bacillus subtilis* 168, that shows around 20% residual activity at 50°C and no residual activity at 60°C (St. John et al. 2006). The thermostability of *Gt*XynA1 has great advantage for the degradation of xylan at higher temperatures (60 °C) over enzymes produced by most mesophilic organisms.

In previous studies, *G. stearothermophilus* T-6 was grown on 4-*O*-methyl –D-glucurono-D-xylan and the main end products found were xylotriose substituted with 4-O-methyl–D-glucuronic acid together with xylobiose and xylose (Shulami et al. 1999). Although we found the same end products in the present study, we also detected xylotriose and xylotetraose as minor end products (Figure 4). Only one substituted oligosaccharide was produced, identified as a xylotriose substituted by 4-*O*-methylglucuronic acid located at the non-reducing end (Figure 4B').

When comparing the mode of action on WAX, the main difference between *Gt*XynA1 and *Aa*XynA is that the former enzyme is less hindered by arabinosyl substituents. This is deduced from the difference in degradation products formed between the incubations of *Gt*XynA1 and *Aa*XynA on WAX. Similar end products were found in previous studies using TLC analysis but, no quantification of the linear end products and no identification of substituted end products was performed (Marcolongo et al. 2015; Anand et al. 2013). Formed xylo-oligosaccharides are taken up by the cell *via* specific sugar transporters (Shulami et al. 2014). These transporters are encoded by the operon *xynEFG*, located in the HUS locus, and have a preference for short xylo-oligosaccharides within a DP of 2 to 6. More specifically, highest affinity is towards trisaccharides and for other xylo-oligosaccharides the affinity is as follows: X2<X3>X4>X5>X6 (Shulami et al. 2007). The relative quantity of different xylo-oligosaccharides formed by the action of *Gt*XynA1 on BeWX shows that mostly xylobiose and xylose is formed (Table 2, Figure 5). The discrepancy between end products formed by *Gt*XynA1 and the affinity of the sugar transporter XynEFG is most likely caused by the 24 h incubation as *Gt*XynA1 is also active towards xylotriose. Previous studies demonstrated the formation of xylotriose within 30 minutes of incubation and prolonged incubation resulted in a decrease of xylotriose and an increase in xylobiose and xylose (Marcolongo et al. 2015; Anand et al. 2013).

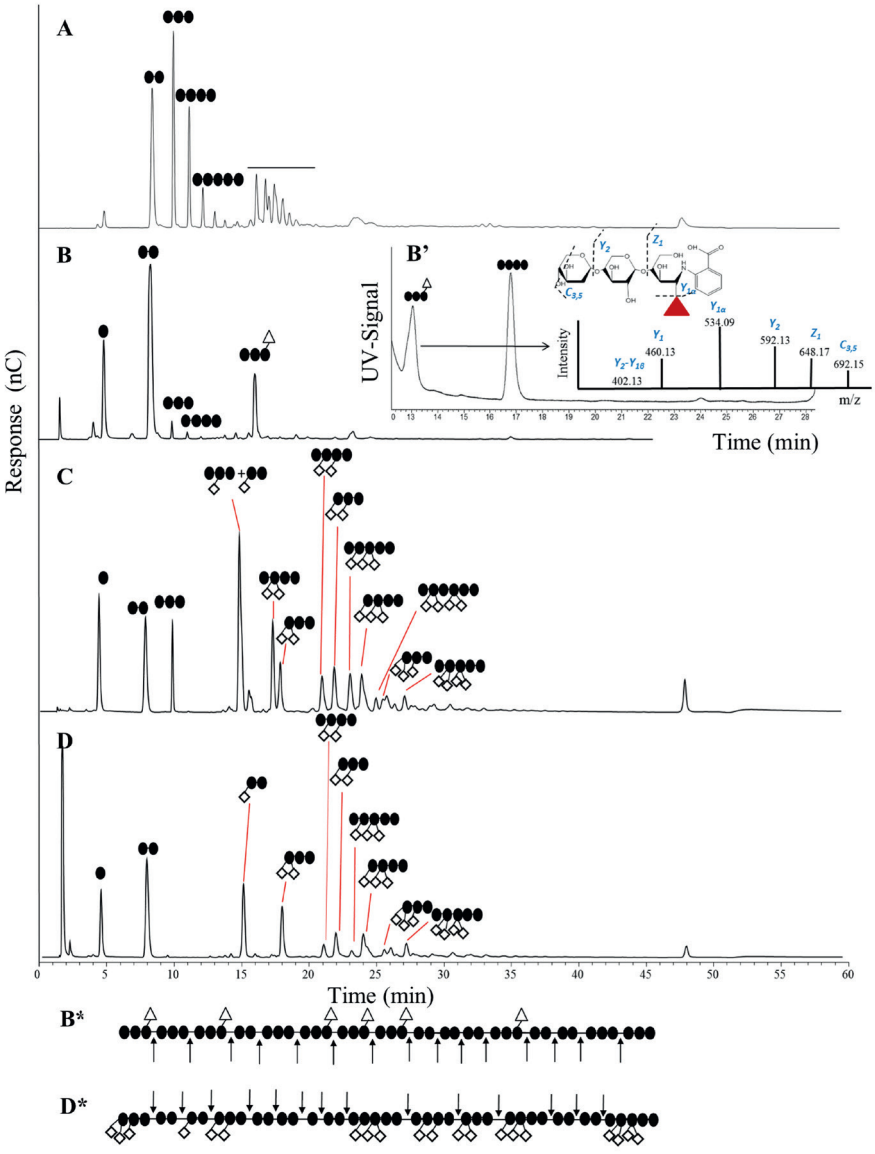


Figure 4 HPAEC elution pattern of with *Aa*XynA and *Gt*XynA1 incubated with beech wood xylan (BeWX) (A, B respectively) and wheat arabinoxylan (WAX) (C, D respectively) for 24h, and respective cleavage pattern of *Gt*XynA1 on BeWX (B*) and WAX (D*). •=xylosyl residue, ◊=arabinosyl residue, ▲= 4-O-methylglucuronic acid. B´ shows the identification of the main substituted xylo-oligosaccharide from chromatogram B. The structure was identified based on its UV signal and the MS2 fragmentation pattern.

The mode of action of *Gt*XynA1 creates linear and branched xylo-oligosaccharides of which most are in the DP range of 2 to 6. These small branched oligosaccharides are believed to enter the cell via specific ABC-transporters (Shulami et al. 2014). For most of these transporters their affinity towards linear oligosaccharides was determined (Shulami et al. 2007; Shulami et al. 2011; Shulami et al. 1999). However, for the arabinooligosaccharide transporter AbnEFJ it was shown that the binding constants of branched oligosaccharides where 2 orders of magnitude higher than those obtained for linear oligosaccharides (Shulami et al. 2011). It is likely that also other ABC transporters that are involved in the uptake of oligosaccharides have higher affinity for branched oligosaccharides in comparison to linear oligosaccharides. The possibility of transporting branched oligosaccharides enables the *Geobacillus* species to use natural substrates in a rapid and efficient way.

Conclusions

*Gt*XynA1 is a GH10 endoxylanase isolated from *G. thermodenitrificans* T12 and is active towards WAX and BeWX, showing extensive degradation of the soluble polymeric structures to linear xylo-oligosaccharides in a DP range from 1 to 4 and to relatively short substituted xylo-oligosaccharides. These degradation products are taken up by the cells *via* specific ABC transporters. We have demonstrated that *Gt*XynA1 is able to cleave the xylan backbone of WAX next to an arabinose substituent on the reducing end and the need for one to two free xylose units on the non-reducing end. In contrast, the hydrolysis of BeWX by *Gt*XynA1 occurs next to a 4-*O*-methylglucuronic acid located at the non-reducing end. The thermophilic origin and its extensive degradation of xylans from different origins demonstrate the potential of *Gt*XynA1 for biomass conversion processes at elevated temperatures.

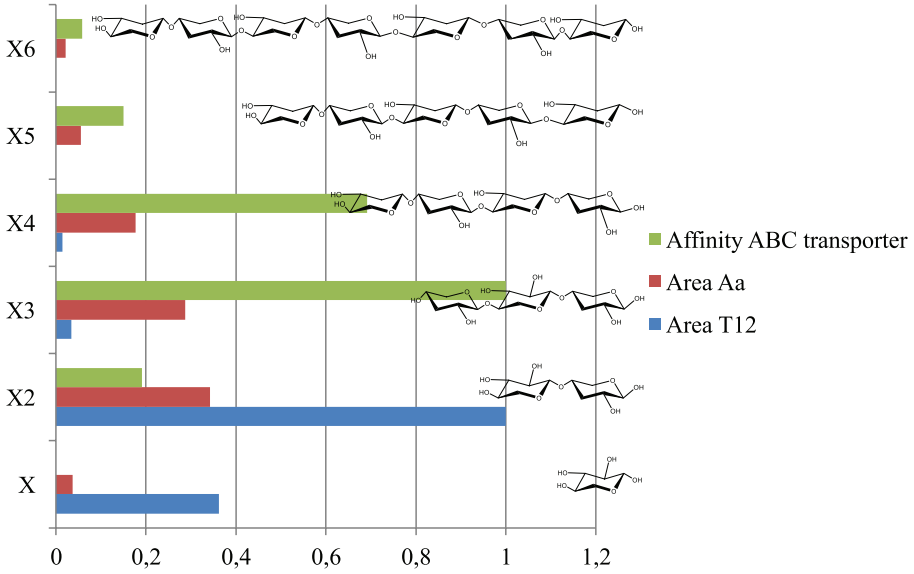


Figure 5 Peak area ratio of linear end products after 24h incubation on beech wood xylan in comparison to oligosaccharide transporter (*xynEFG*) affinity. Green: ratio of affinity of the XynEFG oligosaccharide transporter of *G. stearothermophilus* towards different oligosaccharides. Red: ratio of the end product area after a 24h incubation of *Aa*XynA on beechwood xylan. Blue: ratio of the end product area after a 24h incubation of *Gt*XynA1 on beechwood xylan.

Acknowledgements

We thank Leandra Vermeulen for her excellent technical assistance.

Supplementary information

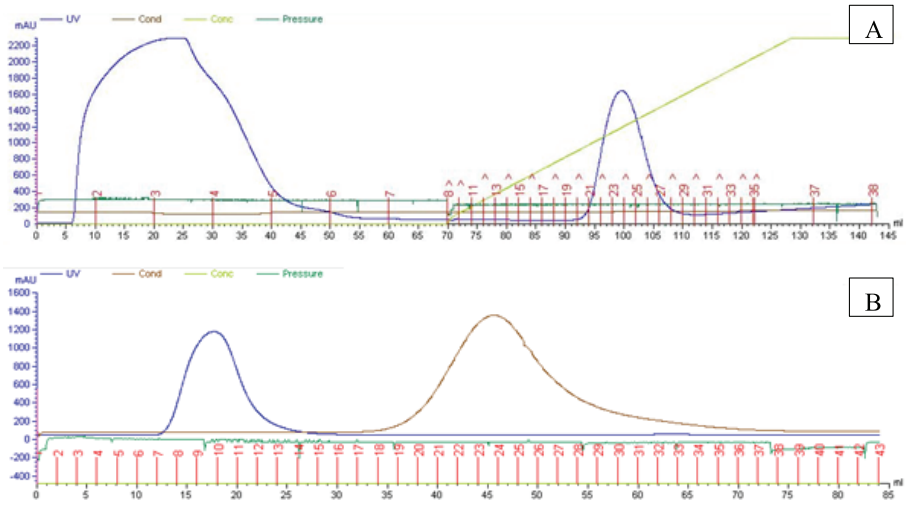


Figure S1 FPLC purification of *Gt*XynA1. Protein fraction not bound to the nickel column eluted with the first 65 mL eluent. Bound *Gt*XynA1 protein was removed from the nickel column using a imidazole gradient (yellow line) which increased from 0 mM to 500 mM over a time span of 20 minutes (A). Fractions 22-25 were pooled and used for desalting the purified *Gt*XynA1 (B). Fractions 8-10 of the desalting column were pooled and then used for further experiments.

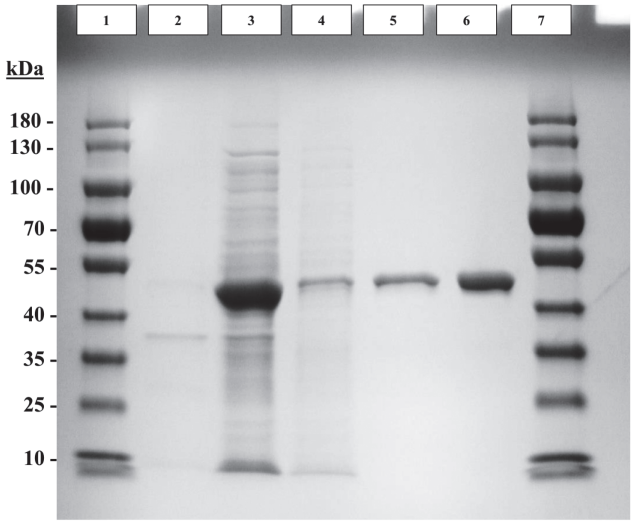


Figure S2 10% SDS-PAGE of purified endoxylanase from *G. thermodenitrificans* T12 followed by PageBlue staining. Lane 1: Protein marker; Lane 2: Pellet fraction; Lane 3: cell-free extract; Lane 4: non-binding protein fraction from FPLC; Lane 5: Purified recombinant *Gt*XynA1; Lane 6: Purified and desalted recombinant *Gt*XynA1; Lane 7: Protein marker

Chapter 5

Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes

Martinus J.A. Daas[1], Bart Nijsse[2], Antonius H.P. van de Weijer[1], Bart W.A.J. Groenendaal[1], Fons Janssen[1], John van der Oost[1], and Richard van Kranenburg [1,3]

Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, the Netherlands[1];
Laboratory of Systems and Synthetic Biology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, the Netherlands[1];
Corbion, Arkelsedijk 46, 4206 AC Gorinchem, the Netherlands[3].

This chapter was adapted from:
Daas, M. J. A., Nijsse, B., Van de Weijer, A. H. P., Groenendaal, B., Janssen, F., Van der Oost, J. & Van Kranenburg, R. Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes.
Microbial Cell Factories, 2017 (Submitted)

Abstract

Background Consolidated bioprocessing (CBP) is a promising approach for the conversion of lignocellulosic biomass to biofuels and biochemicals. In order to obtain a cost-effective conversion, the organism of choice is required to hydrolyze and metabolize both hemicellulose and cellulose. The enzymatic conversion of cellulose to glucose requires the synergistic action of three types of enzymes: exoglucanases, endoglucanases and β-glucosidases. The thermophilic, hemicellulolytic *Geobacillus thermodenitrificans* T12 was shown to be a potential candidate for CBP but lacks the desired endo and exoglucanases needed for the conversion of cellulose. Here we report the heterologous expression of both endoglucanase and exoglucanase genes by *G. thermodenitrificans* T12, in an attempt to complement the enzymatic machinery of this strain and its suitability for CBP.

Results A metagenome screen was performed on open reading-frames (ORFs) encoded in the metagenome of 73 *G. thermodenitrificans* strains using HMM profiles of all known CAZy families that contain endo and/or exoglucanases. Two putative endoglucanases, GE39 and GE40, belonging to glucoside hydrolase family 5 (GH5) were isolated and expressed in both *E. coli* and *G. thermodenitrificans* T12. Structure modeling of GE39 revealed a folding similar to a GH5 exo-1,3-β-glucanase from *S. cerevisiae*. However, we determined GE39 to be a β-xylosidase having pronounced activity towards *p*-nitrophenyl-β-d-xylopyranoside. Structure modelling of GE40 revealed its protein architecture to be similar to a GH5 endoglucanase from *B. halodurans*, and its endoglucanase activity was confirmed by enzymatic analysis against 2-hydroxyethylcellulose, carboxymethylcellulose and barley β-glucan. In addition, we introduced expression constructs to T12 containing *Geobacillus* sp. 70PC53 endoglucanase gene *celA* and both endoglucanase genes (*M1* and *M2*) from *Geobacillus* sp. WSUCF1. Finally, we introduced expression constructs to T12 containing the *C. thermocellum* exoglucanases *celK* and *celS* genes and the endoglucanase *celC* gene.

Conclusions We have identified a novel *G. thermodenitrificans* β-xylosidase (GE39) and a novel endoglucanase (GE40) using a metagenome screen based on multiple HMM profiles. We successfully expressed both genes in *E. coli* and functionally expressed the GE40 endoglucanase in *G. thermodenitrificans* T12. In addition, the heterologous production of active CelK, a *C. thermocellum* derived exoglucanase, and CelA, a *Geobacillus* derived endoglucanase, was demonstrated with strain T12. The native hemicellulolytic activity and the heterologous cellulolytic activity described in this research provide a good basis for the further development of *G. thermodenitrificans* T12 as a host for consolidated bioprocessing.

Background

Lignocellulosic biomass is considered a potential alternative to fossil resources as substrate for biofuels and biochemicals. Although lignocellulosic biomass itself is cheap, saccharification of this substrate is costly due to the variety and amount of enzymes needed for its conversion. In order to reduce costs of this conversion, Lynd *et al.* (2002) proposed consolidated bioprocessing (CBP) as the most promising solution. CBP requires an organism capable of saccharolytic enzyme production and fermentation of the released sugars.

Due to the production of a variety of cellulosic enzymes and its thermophilic nature, *Clostridium thermocellum* is considered an excellent candidate for CBP (Akinosho et al. 2014). The use of a thermophile offers several advantages such as reduced contamination risk, reduced substrate viscosity and a reduced amount of energy needed for cooling (Turner et al. 2007). The downside of *C. thermocellum* is that it is not able to ferment the C_5 sugars released from the lignocellulosic substrate and thus requires extensive engineering of metabolic pathways or the use of co-culturing with pentose utilizing microbes (Argyros et al. 2011; Blumer-Schuette et al. 2008; Zhang & Lynd 2005).

Alternatively, a CBP organism can be derived from the genus *Geobacillus*. Species from this genus are thermophilic, facultative anaerobes that are able to degrade and metabolize hemicellulose (Cripps et al. 2009; Hussein et al. 2015; Shulami et al. 2014). Unlike *C. thermocellum,* most *Geobacillus* strains are not able to efficiently degrade cellulose even when isolated from microcrystalline cellulose or composted plant biomass (Rastogi et al. 2010; Daas et al. 2016). In the majority of isolates β-glucosidases are present, but endo and exoglucanases are missing. Several isolates with endoglucanase activity have been reported, however, the endoglucanase CelA from *Geobacillus* sp. 70PC53 is to date the only characterized true cellulolytic enzyme native to *Geobacillus* spp. (Rastogi et al. 2010; Ng et al. 2009; Bhalla et al. 2014).

An approach to overcome the hurdle of cellulose conversion is to engineer the required cellulose encoding genes into a suitable

host of the genus *Geobacillus*. The expression of a heterologous endoglucanase (WP_010885255.1) from *Pyrococcus horikoshii* in *G. kaustophilus* HTA26 enabled this mutant strain to degrade carboxymethylcellulose and filter paper. However, for complete hydrolysis of cellulose both endo and exoglucanases are required and the HTA426 strain lacks genes required for hemicellulose conversion, making it less suited for CBP (De Maayer et al. 2014).

Heterologous production of *C. thermocellum* cellulases has been demonstrated in *Bacillus subtillis* (Lan Thanh Bien et al. 2014). Here, the production of CelK (reducing end exoglucanase; GH 9) and CelS (non-reducing end exoglucanase; GH48) was demonstrated by the clearing zones of mutant colonies on CMC plates. Both exoglucanases described are also highly expressed in *C. thermocellum* when grown on cellulosic and lignocellulosic substrates and are therefore expected to be of great importance for the cellulolytic activity of *C. thermocellum* (Gold & Martin 2007; Rydzak et al. 2012; Wei et al. 2014).

In this study, we screened the metagenome of 73 *Geobacillus thermodenitrificans* isolates for novel endo and exoglucanases. Subsequently, three *Geobacillus* derived endoglucanases were expressed in *E.coli* and *G. thermodenitrificans* T12. To complement the cellulolytic activity of strain T12 we also introduced an endoglucanase (*celC*) and two exoglucanases (*celK* and *celS*) from *C. thermocellum*.

This is the first study on the expression of a full set of cellulolytic enzymes in *Geobacillus* and thereby provides new insights in the applicability of members of this genus as potential hosts for consolidated bioprocessing.

Materials & Methods

Media, strains, primers and constructs

Cellulolytic Thermophile Vitamin Medium (CTVM; based on (Cripps et al. 2009; Fong et al. 2006; Sizova et al. 2011; Bosma et al. 2015)) contained per liter: 8.37 g MOPS; 1 g NH_4Cl; 3 g NaCl; 1.50 g Na_2SO_4; 0.08 g $NaHCO_3$; 1 g KCl; 1.8 g $MgCl_2 \times 6H_2O$; 0.30 g $CaCl_2 \times 2H_2O$. pH was set to 6.6 at room temperature and the medium was autoclaved for

20 min at 121°C, after which 1 mL K_2HPO_4 (250 g/L; pH 6.6), 10 mL filter sterile 100x metal mix and 1 mL filter sterile 1000x vitamin solution were added. 100x metal mix contained per liter: 1.60 g $MnCl_2 \times 6H_2O$; 0.1 g $ZnSO_4$; 0.2 g H_3BO_3; 0.01 g $CuSO_4 \times 5H_2O$; 0.01 g $Na_2MoO_4 \times 2H_20$; 0.1 g $CoCl_2 \times 6H_2O$; 0.7 g $FeSO_4 \times 7H_2O$; 5 g $CaCl_2 \times 2H_2O$; 20 g $MgCl_2 \times 6H_2O$. 1000x vitamin mix contained per liter: 0.1 g thiamine; 0.1 g riboflavin; 0.5 g nicotinic acid; 0.1 g pantothenic acid; 0.5 g pyridoxamine, HCl; 0.5 g pyridoxal, HCl; 0.1 g D-biotin; 0.1 g folic acid; 0.1 g *p*-aminobenzoic acid; 0.1 g cobalamin.

LB2 contained per liter: 10 g tryptone (Oxoid), 5 g yeast extract (Roth), 10 g sodium chloride and salts mix consisting of 1 g NH_4Cl; 3 g NaCl; 1.50 g Na_2SO_4; 0.08 g $NaHCO_3$; 1 g KCl; 1.8 g $MgCl_2 \times 6H_2O$; 0.30 g $CaCl_2 \times 2H_2O$. pH was set to 6.6 at room temperature and the medium was autoclaved for 20 min at 121°C, after which 10 mL K_2HPO_4 (250 g/L) was added.

Minimal Media (MM) contained per liter: 0.52 g K_2HPO_4; 0.23 g KH_2PO_4; 0.5 g NH_4NO_3 (MMy) or 0.3 g NH_4Cl (MMy+). After autoclaving, 1 mL of the following 1,000x concentrated sterile stocks were added: Nitrilotriacetic acid (200 g/L); $MgSO_4 \times 7H_2O$ (145.44 g/L); $CaCl_2 \times 2H_2O$ (133.78 g/L); $FeSO_4 \times 7H_2O$ (11.12 g/L).

For CTVMy/MMy medium, 0.5 g/L yeast extract (Roth) was added to the medium and CTVMy+/MMy+ contains 5 g/L yeast extract (Roth).

Glycerol stocks of cultures were made by adding 500 µl sterilized 60% glycerol to 1.5 mL culture, in a 2 mL cryogenic vial (Corning). Stocks were stored at -80°C.

In all plate and tube cultures, carbon substrates were used in a concentration of 10 g/L unless stated otherwise. For plate cultures, 5 g/L gelrite (Roth) was added.

Geobacillus thermodenitrificans T12 was isolated from compost (Daas et al. 2016). *E. coli* DH5α and *E. coli* TG-90 were used for DNA manipulation and *E. coli* BL21(DE3) and *G. thermodenitrificans* T12 were used for protein production. *E. coli* DH5α and BL21(DE3) were grown at 37°C in Luria-Bertani (LB) medium and *E. coli* TG-90 was grown at 30 °C. Wild-type *G. thermodenitrificans* T12 was grown at 65 °C and at 55 °C when harbouring plasmid DNA to maintain plasmid replication.

DNA isolation, sequencing and assembly

All strains were grown in LB2 media at 65 °C in a rotary shaker at 150 RPM. Genomic DNA was isolated from 10 mL of logarithmic growing cultures of an OD_{600} of approximately 1.00 AUs by using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre, Madison, Wisconsin, USA) according to manufacturer's protocol. Genomic DNA was then pooled and sent for sequencing by the company Baseclear B.V. (Leiden, The Netherlands). Paired-end sequence reads were generated using the Illumina HiSeq2500 system. FASTQ sequence files were generated using the Illumina Casava pipeline version 1.8.3. The initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control signal were removed using an in-house filtering protocol. Reads were aligned to the genome of *G. thermodenitrificans* T12, a strain known to contain no functional cellulases, with Bowtie2 v2.2.4 (parameters: –local --no-mixed –no-discordant) (Langmead & Salzberg 2012). The unaligned reads were assembled with Ray v2.3.1 with a kmer set to 81. The same unaligned reads were aligned to the assembled scaffolds with Bowtie2 v2.2.4 and converted to sorted BAM files with Samtools v1.1 (Li et al. 2009). The sorted BAM files were used as an input for Pilon v.1.10 (Walker et al. 2014) for automatic error correction resulting in 2616 scaffolds with a total of 28814104 bp and a n50 of 41189 bp with average coverage of 715.

Selection and sequence analysis of metagenome putative cellulases

Prodigal v2.6.1 (Hyatt et al. 2010) was used for gene prediction with the *G. thermodenitrificans* T12 genome as a training set. The predicted proteins were used as input for hmmsearch v3.1b1 (Eddy 2011) to identify possible cellulases with Hidden Markov models (HMMs) from PFAM (Finn et al. 2016) and dbCAN (Yin et al. 2012) of all known glycosyl hydrolase (GH) families that contain endoglucanases and/or exoglucanases: GH 1, 5, 6, 7, 8, 9, 10, 11, 12, 26, 44, 45, 48, 51, 74 and 124 (accession numbers resp. PF00150.16, PF01341.15, PF00840.18, PF01270.15, PF00759.17, PF00331.18, PF00457.15, PF01670.14, PF02156.13, PF12891.5, PF02015.14, PF02011.13). Models for GH 51, 74 and 124 were obtained from dbCAN. Proteins fitting the models were functionally annotated with BLASTp v2.6.0+ using the TrEMBL v2017_1 protein database. A schematic overview of the procedure is given in Figure 1.

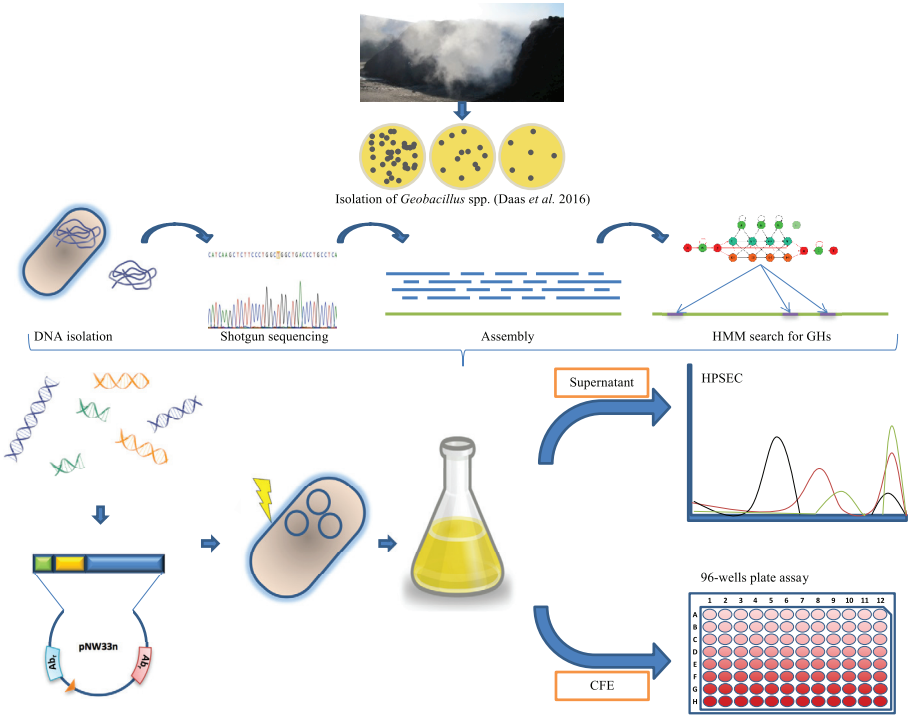


Figure 1 Overview of the workflow for detecting novel cellulases from a *G. thermodenitrificans* metagenome. From compost a total of 73 *G. thermodenitrificans* strains were isolated of which the genomic DNA was pooled and subjected to shotgun sequencing. By screening the assembled metagnome data on proteins matching against HMM profiles of all known glycoside hydrolase families that contain endoglucanase and/or exoglucanases we identified several novel putative cellulases. The genes encoding these putative cellulases were used to create expression constructs in both *E. coli* and *G. thermodenitrificans* T12. The supernatant and CFE of these cultures were then subject to activity assays against a variety of substrates.

Initial protein function was based on best hits in the TrEMBL v2017_1 protein database, and then manually filtered on their predicted function. The remaining selection (Table 1) was subject to further sequence analysis, which involved protein model prediction by Phyre2 (Kelley et al. 2015), determination of conserved active site residues and signal peptide prediction by SignalP4.1 (Petersen et al. 2011).

Table 1 Selection of metagenome derived putative endoglucanases and their closest orthologs from the TREMBL v080914 protein database

Query	E-value	Bitscore	Hit-Gene ID	AA identity (%)	Protein names	Organism
GE32	0	2673	A0A0Q1EJD2	100	GEPA3_1816	*Geobacillus* sp. PA-3
GE33	0	2640	A0A0Q1EJD2	98	GEPA3_1816	*Geobacillus* sp. PA-3
GE39	0	2559	A0A0Q1EJD2	96	GEPA3_1816	*Geobacillus* sp. PA-3
GE40	0	1691	E6TRT3	58	LPXTG-motif cell wall anchor domain protein	*Bacillus cellulosilyticus* (DSM 2522)

Expression of recombinant putative endoglucanases GE39 and GE40

Selected putative endoglucanase genes were amplified by high-fidelity PCR using PhusionHF polymerase. The PCR mixes contained 10 µl Phusion HF Buffer, 1 unit of Phusion DNA polymerase (Thermoscientific), 100 µM of dNTPs, 20 ng DNA, 0.2 µM of both the forward primers BG6897 (GCGCCATGGAAATGCTTAAGGTCACTA) for GE39 and BG6899 (GCGCCCATGGAAGACAATAAAGCGTCGGCATAC) for GE40, and the reverse primers BG6898 (CGCCTCGAGTTAT-AAACTTATACTGGACTGATTTG) for GE39 and BG6900 (CGCCTCGAGCTACTTTCCGGCCATCTTCAA) for GE40. MilliQ was added to a total volume of 50 µl. PCR products were checked on 1% agarose gels and products were purified by using a GeneJet PCR purification kit (Fermentas). The PCR products of GE39 were digested with restriction enzymes *Nco*I and *Xho*I and purified PCR products of GE40 were digested with restriction enzymes *Nco*I and *Ava*I and subsequently cloned into the pCDF-1b vector (Figure 2). The recombinant plasmids were introduced to *E.coli* DH5α using heat shock competent cells and then plated on selective media. Plasmids were purified using the GeneJET plasmid miniprep kit according to manufacturer's protocol and were subsequently introduced to *E. coli* BL21(DE3) for enzyme expression. BL21 strains containing the recombinant endoglucanases were cultured overnight in 10 mL LB medium in a 37 °C rotary shaker at 150 RPM. Next morning, cultures were cooled on ice for 10 minutes prior to being transferred to 50 mL LB medium supplemented with 50 µg/ml streptomycin and

grown at 20 °C under constant agitation at 150 RPM. Expression of the endoglucanases was induced using 0.1 mM isopropyl-β-thioglactopyanoside (IPTG) at an OD600 of approximately 1.00.

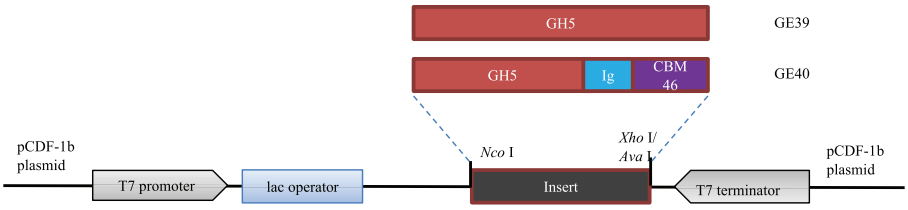


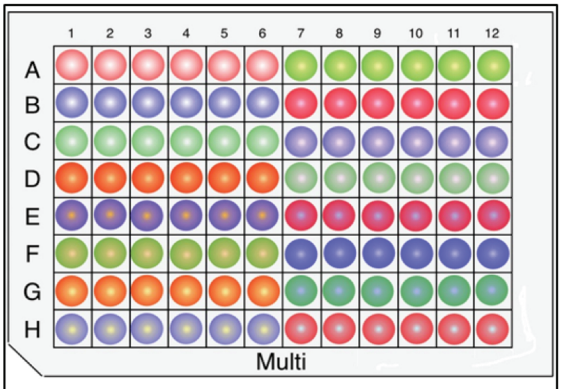
Figure 2 Expression constructs of two putative endoglucanases derived from the metagenome analysis. Gene *GE39* encodes a GH5 catalytic domain and gene *GE40* encodes a GH5 catalytic domain followed by an immunoglobulin-like domain (Ig) and a carbohydrate module of family 46 (CBM). Both genes were expressed in BL21(DE3) using the pCDF-1b plasmid under control of the lac operator.

Extraction and characterization of recombinant predicted endoglucanases GE39 and GE40

The induced *E. coli* BL21(DE3) cultures (50 mL) containing the recombinant endoglucanases GE39 and GE40 were collected after 18h by centrifugation at 4,800 ×*g* for 15 min at 4°C. The cells were resuspended in 5 mL of 200 mM sodium phosphate buffer (pH 6.00) and disrupted using a French press at 1,200 psi. For DNA lysis, DNase I was added (1 mg/mL) to the crude extracts and incubated at room temperature for 15 minutes. The cell debris was removed by centrifugation at 30,000 ×*g* for 15 min at 4 °C. The resulting supernatant was filter-sterilized (0.45 µm) to remove remaining cell debris and protein concentrations of the obtained CFE were determined by the Bradford method using the Bradford Reagent Assay Kit (Sigma) with the bovine serum albumin as the standard (Bradford 1976).

To determine saccharolytic activities of the CFE we used a series of chromogenic substrates in a Glycospot Multi CPH 96-wells filter plate (Kračun et al. 2015) (Figure 3) (Glycospot, Frederiksberg C, Denmark). Substrates were activated by the addition of 200 µL activation solution. Centrifugation (2,700 ×*g*, 10 minutes) was applied to remove the solution followed by a double wash with 100 µL

milliQ. The final reaction mixture in each well consisted of 145 µL sodium phosphate buffer (pH 6.0) and 5 µL of CFE. Plates were then sealed using an aluminium adhesive foil (VWR, Radnor, PA, USA) and incubated at 60 °C in a rotary shaker at 180 RPM. After 24 h the reaction mixture was collected in the product plate by centrifugation (2,700 ×*g*, 10 min) and absorbance was measured at 595 nm (blue) and 517 nm (red) using a plate reader (Biotek Instruments Inc., Winooski, VT, USA). Negative controls consisted of sodium phosphate buffer and CFE from an *E. coli* culture containing empty pCDF1b plasmid. The thermostable endoglucanase, CelTM, (Megazyme, Wicklow, Ireland) from *Thermotoga maritima* was used as positive control at a concentration of 1 µg/mL.



Row/ Column	1-6	7-12
A	CPH-2-hydroxyethylcellulose	CPH-xyloglucan
B	CPH-β-glucan from barley	CPH-amylopectin
C	CPH-β-glucan from yeast	CPH-amylose
D	CPH-curdlan	CPH-xylan
E	CPH-pachyman	CPH-arabinan
F	CPH-pullulan	CPH-pectic galactan
G	CPH-dextran	CPH-galactomannan
H	CPH-chitosan	CPH-rhamnogalacturonan (soybean)

Figure 3 Glycospot's Multi CPH plate. The substrates and positions are shown (adapted from www.glycospot.dk/48unique.html)

To determine exo-activity, we used 3 µg of CFE mixed with 200 µL of a 200 mM sodium phosphate buffer (pH6.00). The reaction was started by adding 10 µL of 50 mM *p*-nitrophenyl-β-D-xylopyranoside (pNPβX) or *p*-nitrophenyl-β-D-glucopyranoside (pNPβG) (Sigma, St. Louis, MO, USA) and incubated at 60 °C for 10 minutes. Negative controls consisted of sodium phosphate buffer and CFE from an *E. coli* culture containing empty pCDF1b plasmid. Reaction was stopped by adding 1 mL of 0.5M bicarbonate and the amount of released pNP was determined by absorbance measurement at 410 nm and subsequently plotting the data against a standard curve generated using pNP as a substrate. One unit (U) of activity was defined as the release of 1 µmol pNP per minute.

Expression constructs of *Geobacillus* endoglucanases

Linear constructs of the metagenome derived GE40 encoding gene as well as *Geobacillus* endoglucanases *M1* (GI:523426779), *M2* (GI:523426040) and *celA* (GI:214003628) were synthetically manufactured by Bio Basic Inc. (Amherst, NY, USA). Constructs were composed of the P_{uppT12} promoter driving expression of the various endoglucanase genes supplemented with the coding sequence for the *Gt*XynA1 (KX962565.1) signal peptide and were separately cloned into the pNW33n vector using restriction enzymes listed in Table 2. Ligation mixes were introduced directly to *G. thermodenitrificans* T12 as previous attempts in cloning the *Geobacillus* constructs to *E.coli* DH5α and TG90 failed to yield correct transformants. Transformation and recovery of the transformed T12 cells performed as described before with minor modifications (Daas et al. 2016).Strain T12 was grown O/N in LB2 after which the cells were diluted in 50 mL fresh LB2 medium (OD_{600} = 0.05) in a 250 mL baffled shake flask. This pre-culture was incubated at 65 °C in a rotary shaker at 180 RPM. When OD_{600} reached 0.95, the cells were pelleted by centrifugation (4800 ×*g*) and washed twice with ice-cold milliQ (50 mL) and twice with 10% glycerol (25 mL and 10 mL, respectively). Competent cells were then aliquoted (65 µL) and incubated with 1 µg plasmid DNA for 2 min. Following electroporation (2 kV, 200 Ω, 25 µF), cells were recovered for 2 h in 1 mL pre-warmed LB2

at 55 °C. Cells were then plated on LB2 agar containing 7 µg/mL chloramphenicol. Colonies were picked after 24h of growth at 55 °C and plasmid presence and integrity was verified by PCR using FW primers BG3665 (5'- GCTCGTTATAGTCGATCGGTTC-3'), or BG3859 (5'-GTTTGCAAGCAGCAGATTACG-3') for *celA*, and RV primer BG3664 (5'-AGGGCTCGCCTTTGGGAAG-3').

Table 8 Overview of the different components of each used construct and the restriction enzymes (RE) used for the insertion of the construct to the pNW33n plasmid. Gene fragment numbers indicate the fragment of the gene used to build the different constructs. AA: amino acid.

				Gene fragment			
Gene origin	**Construct name**	**Promoter**	**Signal peptide**	**Start (AA)**	**End (AA)**	**Construct length (bp)**	**Restriction enzymes**
M1 (GI:523426779)	T12-M1	P_{uppT12}	*Gt*XynA1	1	389	1361	*Acc*65I / *Pst*I
M2 (GI:523426040)	T12-M2	P_{uppT12}	*Gt*XynA1	1	359	1263	*Pst*I / *Hind*III
CelA (GI:214003628)	T12-CelA	P_{uppT12}	*Gt*XynA1	25	392	1294	*Bsp*HI / *Eco*RI
GE39	T12-GE39	P_{uppT12}	*Gt*XynA1	1	488	1674	*Acc*65I / *Pst*I
GE40wt	T12-GE40wt	P_{uppT12}	GE40	1	538	1732	*Acc*65I / *Bam*HI
GE40	T12-GE40	P_{uppT12}	*Gt*XynA1	28	538	1732	*Acc*65I / *Bam*HI
CelK (GI:2978566)	T12-CelK	P_{uppT12}	*Gt*XynA1	28	812	2536	*Acc*65I /*Xba*I
CelC (GI:12584559)	T12-CelC	P_{uppT12}	*Gt*XynA1	1	343	1219	*Bsp*HI / *Eco*RI
CelS (GI:145558928)	T12-CelS	P_{uppT12}	*Gt*XynA1	28	674	2130	*Pst*I / *Hind*III
CelSH (GI:145558928)	T12-CelSH	P_{uppT12}	*Gt*XynA1	28	674	2130	*Pst*I / *Hind*III

C. thermocellum cellulases CelC, CelK and CelS

Linear constructs of *C. thermocellum* cellulases *celK*, *celS* and *celC* genes were synthetically manufactured by Bio Basic Inc. (Amherst, NY, USA). Constructs were composed of the P_{uppT12} promoter driving expression of the various catalytic domains fused to the carbohydrate binding module of each of the *C. thermocellum* cellulases supplemented with the coding sequence for the *Gt*XynA1 (KX962565.1) signal peptide (Figure 4). Constructs were separately cloned into the pNW33n vector using restriction enzymes listed in Table 2. Because of the severe mismatch in codon usage between

G. thermodenitrificans and the gene *celS* from *C. thermocellum*, a codon harmonized variant (*celSH*) was synthesized. Constructs were separately cloned into the pNW33n vector and ligation mixes were introduced to *E.coli* TG90. Recovery of the transformed TG90 cells was done at 30 °C for 2.5 hours in a rotary shaker at 150RPM, as recovery at 37 °C failed to yield correct transformants. Cells were plated on LB agar containing 12.5 µg/mL chloramphenicol. Colonies were picked after 48h of growth at 30 °C and plasmid presence and integrity was verified by PCR using FW primers BG3665 (or BG3859 for *celC*) and RV primer BG3664. Plasmids containing the correct insert sequence were isolated using JETstar Plasmid Purification MAXI Kit (Genomed, Löhne, Germany) according to manufacturer's protocol. Purified plasmids were subsequently introduced to *G. thermodenitrificans* T12 as previously described and positive clones were verified by PCR and sequencing of the plasmid insert.

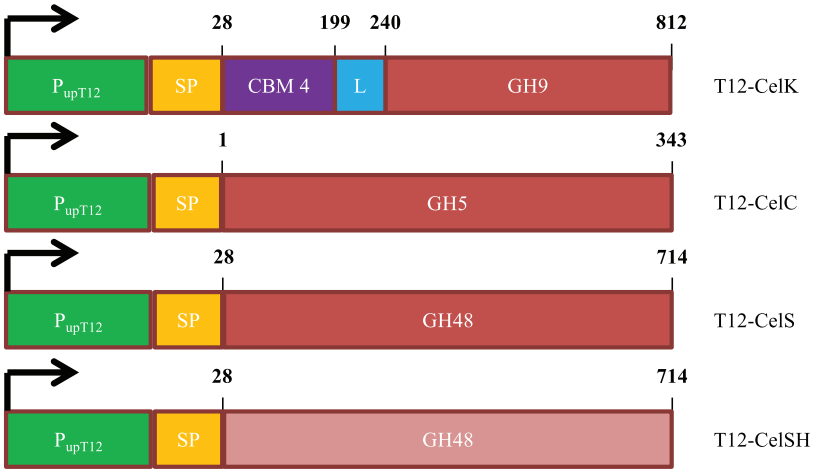


Figure 4 Overview *C. thermocellum* constructs for expression in T12. Numbers indicate start and end amino acid numbers of the original cellulases as deposited in the UniprotKB database.

Endoglucanase activity assays

Cultures (50 mL) of *G. thermodenitrificans* T12 containing the different cellulase constructs were grown in LB2 medium containing 0.5% CMC as substrate. After 18h of growth at 55 °C a 10 µL sample was

taken and spotted on solid LB2 medium containing 0.5% CMC. Plates were incubated for 24 h at 55 °C and subsequently stained for 5 minutes by flooding the plates with 0.1% Congo red dye followed by destaining with a 1 M NaCl solution for 15 min.

The remainder of the T12 cultures were centrifuged (4,800 ×*g*, 10 minutes) after which the cell fractions were disrupted using a French press at 1,200 psi. The obtained cell lysates were centrifuged (30,000 ×*g*, 15 min) to obtain clear CFEs. To determine saccharolytic activities of the CFE we used a series of chromogenic substrates in a Glycospot Multi CPH 96-wells filter plate (Kračun et al. 2015) as described above. Negative controls consisted of sodium phosphate buffer and CFE of a T12 culture containing an empty pNW33n plasmid. The thermostable endoglucanase, CelTM, (Megazyme, Wicklow, Ireland) from *Thermotoga maritima* was used as positive control at a concentration of 1 µg/mL.

Cellulolytic activity of the cultures supernatants was analysed by high performance size exclusion chromatography (HPSEC) on an Ultimate 3000 HPLC system (Thermo Scientific, Sunnyvale, CA, USA) equipped with a set of three TSK-gel columns (6.0 mm x 15.0 cm per column) in series (SuperAW4000, SuperAW3000, SuperAW25000, Tosoh Bioscience, Stuttgart, Germany) in combination with a PWX-guard column (Tosoh Bioscience). HPSEC was controlled by the Chromeleon software (Thermo Scientific). Elution took place at 55°C with 0.2 M sodium nitrate at a flow rate of 0.6 mL/min. The eluate was monitored using a refractive index (RI) detector (Shoko Scientific Co., Yokohama, Japan). Calibration was made by using pullulan series (Polymer Laboratories, Union, NY, USA) with a molecular weight in the range of 0.18-788 kDa.

Nucleotide sequence accession numbers

Nucleotide sequences of GE32, GE33, GE39 and GE40 were deposited in GenBank with accession numbers XXXX to XXXX; nucleotide sequence of scaffold_9 from *G. thermodenitrificans* T81 has been deposited in GenBank with accession number XXXX.

Results

Selection and sequence analysis of metagenome putative cellulases
A total of 82 hits for potential endoglucanases or exoglucanases were identified in the metagenome of 73 *G. thermodenitrificans* strains using HMM profiles against all glycoside hydrolase families known to contain endoglucanases and/or exoglucanases. After manual inspection of the best hits with the uniprotKB database, we obtained a final selection of four (GE32, GE33, GE39 and GE40) potential endoglucanases which were subject to further sequence analysis. Sequence analysis of the four putative endoglucanases obtained from the metagenome screen showed high amino acid sequence identity (>94%) between GE32, GE33 and GE39. Therefore, we assumed these three proteins to have identical activities and we selected GE39 and GE40 for our assays. One scaffold (assigned scaffold 9) contained both the *GE39* and *GE40* genes and was annotated further using BlastP against the NCBI non-redundant protein database. Comparison of this scaffold with the genome of *G. thermodenitrificans* T12 revealed that the two genes reside in the hemicellulose utilization (HUS) locus. Great variation in genetic content was observed between the HUS loci of strain T12 and scaffold 9 (Figure 5). However, the localization of the *GE39* gene on scaffold 9 is clustered with *xylA* and *xylB* which may suggest a role as a β-xylosidase, in analogy to the XynB3 encoding gene of strain T12. Indeed, the GE39 enzyme shows most amino acid sequence identity (36%) to the characterized β-xylosidase (*Pc*Xyl5) from *Phanerochaete chrysosporium* and contains GH5-family conserved glutamine residues at positions E188 and E318 (Figure 6, S2).

For gene GE40 we could not predict its putative function based on this alignment as we could not relate GE40 to any of the T12 genes nor to any other *Geobacillus* derived protein from the NCBI non-redundant database. The amino acid sequence of GE40 is closely related (51% and 31% AA sequence identity) to the well characterized GH5 family endoglucanases from *Bacillus halodurans* (BhCel5b, PDB:4V2X_A) and *Bacillus licheniformis* (BlCel5b, PDB: 4YZT_A), respectively (Figure S2). Sequence comparison shows similar

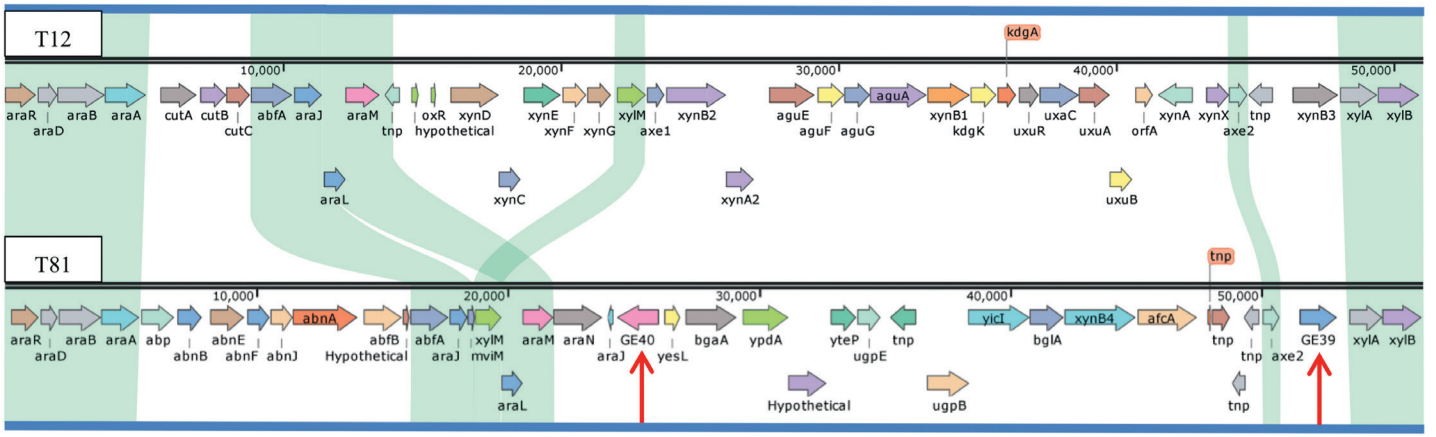


Figure 5 Overview of the partial HUS loci of *G. thermodenitrificans* T12 (top) and T81 (bottom). Annotations were assigned based on protein sequence blast (BlastP) against the National Center for Biotechnology (NCBI) non-redundant protein database. Comparison between the predicted CDSs of the two HUS loci was performed using a localized BlastP analysis in CloneManager v9.51 (Sci-ed software, Denver, CO, USA). Conserved genes are shaded in green, genes coding for endoglucanase GE40 and β-xylosidase GE39 are indicating by red arrows.

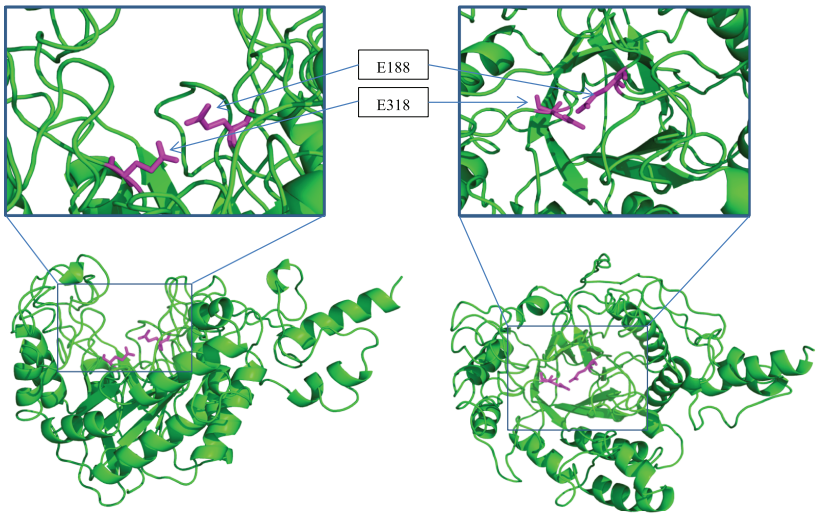


Figure 6 Protein structure of the GE39 endoglucanase protein. The protein model has been predicted using Phyre2 on the basis of 7 templates. 98% of the residues from GE39 were modelled at >90% confidence. The GE39 protein has a GH5 catalytic domain with active site residues (pink sticks) predicted to be E178 and E296 based on multiple alignment.

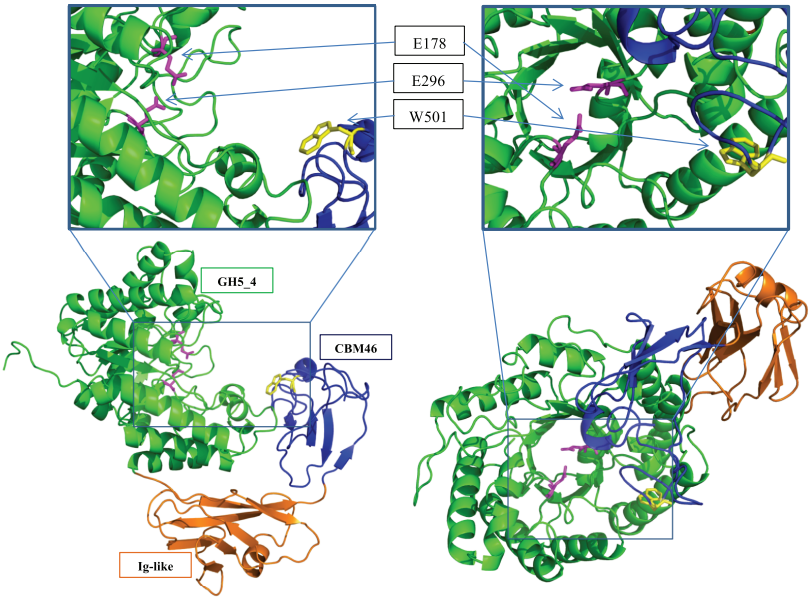


Figure 7 Protein structure of the GE40 endoglucanase protein. The protein model has been predicted based on the crystal structure of a *Bacillus halodurans* endoglucanase (*Bh*Cel5b; PDB:4V2X_A). 92% of the residues from GE40 were modelled at >90% confidence. The GE40 protein contains an N-terminal GH5_4 catalytic domain, an immunoglobulin-like domain and a C-terminal CBM46. Active site residues were to be E178 and E296 (pink sticks). The tryptophan residue at position W501 (yellow stick) was predicted to be involved in ligand binding.

protein architecture of both Cel5b proteins and protein GE40. GE40 is comprised of a GH5_4 catalytic domain, an immunoglobulin-like module and a carbohydrate binding module belonging to family 46 (Figure 7). Deeper analysis of the amino acid sequence of GE40 reveals conserved active site residues of the GH5_4 domain at positions Glu-178 and Glu-296, and a conserved residue active in ligand-binding of the CBM to be Trp-501 (Figure 7) (Venditto et al. 2015; Liberato et al. 2016). The location of Glu-178 and Glu-296 on the C-termini of the fourth and seventh ß-strand respectively, is in accordance with the known position of active site residues in enzymes belonging to the GH5 family (Venditto et al. 2015; Venditto et al. 2014; Henrissat et al. 1995).

Cellulolytic activity of the GE40 and GE39 proteins produced in *E. coli* was measured using the solubilized fraction of the chromogenic substrates in a Glycospot Multi CPH assay plate (Table S1). CFE of the GE40 expressing *E.coli* culture showed high activity towards cellulose and barley derived β-glucan. In contrast, CFE of the GE39 producing *E.coli* culture showed no activity to any of the chromogenic substrates (Figure 8, Table S1). However, it showed activity towards *p*-nitrophenyl-β-d-xylopyranoside (55 U/mg) along with some side activity towards *p*-nitrophenyl-β-d-glucopyranoside (11 U/mg). We therefore conclude that GE39 is a β-xylosidase.

G. thermodenitrificans cellulolytic activity assays

The GE40 metagenome endoglucanase, the *Geobacillus* endoglucanase encoding genes *M1*, *M2* and *celA*, and the *C. thermocellum* cellulase encoding genes (*celK, celC and celS*) were used to create expression constructs for *G. thermodenitrificans* T12. Only the catalytic domains and, if present, the carbohydrate binding modules were used and fused with the *Gt*XynA (extracellular endoxylanase) signal peptide and the P_{uppT12} promoter sequence (derived from the uracil phosphoribosyltransferase encoding gene of strain T12), which have both successfully been used for heterologous protein production in *G. thermodenitrificans* T12 (Daas et al. 2016).

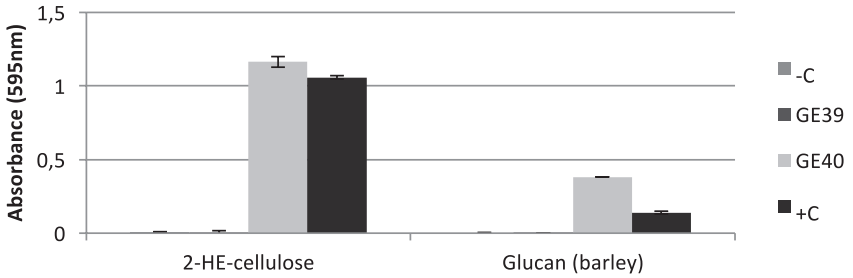


Figure 8 Enzymatic acitivity of CFE from GE39 and GE40 expressing *E. coli* BL21(DE3) cultures. CFE was incubated at a pH of 6.0 on a series of chromogenic substrates and incubated for 24h at 60 °C in a rotary shaker (180 RPM). Degradation of the substrate by cellulolytic activity was measured by colorimetric measurements of the solubilized chromogenic oligosaccharides at 595nm.

For endoglucanase GE40, we created two constructs; T12-GE40wt: the native gene under control of the P_{uppT12} promoter and T12-GE40: as T12-GE40wt but with the *Gt*XynA signal peptide as replacement for the original signal peptide. Also for the expression of *C. thermocellum* exoglucanase *celS* we created two constructs; one with the original sequence (T12-*celS*) and one with a codon harmonized sequence (T12-*celSH*). Constructs were introduced to strain T12 using the pNW33n vector and functional activities were tested *in vivo* using the Congo red assay and HPSEC analysis. Strains harbouring constructs with the *C. thermocellum celC*, *celS* and *celSH* genes lack cellulase activity. However, strains harbouring a construct derived from *C. thermocellum celK*, the *Geobacillus celA* or the *GE40* gene showed activity against amorphous cellulose (Figure 8, Table S1). This activity was also confirmed by HPSEC analysis for constructs T12-CelK and T12-CelA, indicated by the reduced MW of the CMC peak eluting between 8 and 10 minutes (Figure 9).

Discussion

G. thermodenitrificans T12 has been shown to be a potential host for consolidated bioprocessing mainly due to its xylanolytic activity and genetic accessibility (Daas et al. 2016). The requirement for a true CBP organism to hydrolyze both hemicellulose and cellulose

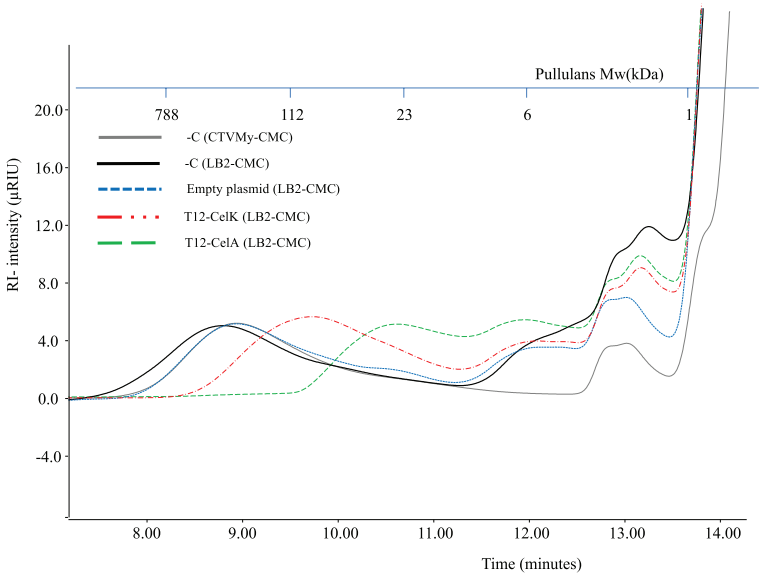


Figure 9 HPSEC analysis of carboxymethylcellulose (CMC) degradation by *G. thermodenitrificans* T12. Cultures were grown for 24h on LB2 with 1% CMC. The degradation of the cellulose is visualized by the change in molecular weight of the peak eluting between 8 and 9 minutes. Both cultures expressing CelA (T12-CelA, green) and CelK (T12-CelK, red) show a clear degradation of cellulose in contrast to the empty plasmid control (Empty plasmid, blue). Negative controls (-C) consisting of non-inoculated media show some effect of yeast extract around an elution time of 12-14 min.

has not been achieved yet for *G. thermodenitrificans*. Although major efforts have been done to find cellulolytic *Geobacillus* spp. (Hussein et al. 2015; Rastogi et al. 2010; Daas et al. 2016; Ng et al. 2009; Bhalla et al. 2014; Bosma et al. 2015; Tai et al. 2004), to date there is no evidence of a strain capable of efficient cellulose conversion. The endoglucanase CelA from *Geobacillus* sp. 70PC53 is to date the only characterized cellulolytic enzyme native to *Geobacillus* spp. (Rastogi et al. 2010; Ng et al. 2009; Bhalla et al. 2014). Two more *Geobacillus* derived enzymes have been proposed to be endoglucanases, M1 and M2 from *Geobacillus* sp. WSUCF1. The study describing the characterization of endoglucanases M1 and M2 shows activity of the M1 and M2 proteins against cellulose (Rastogi et al. 2010). However, in this study we could not detect degradation of CMC, not even under constitutive expression of the *M1* and *M2* genes. Sequence analysis against the UniprotKB database gives highest sequence identities against peptidases. In addition, we could not identify active site

residues based on sequence comparison to other endoglucanases of the GH5 family. Therefore, we assume that these enzymes are most likely peptidases and not true endoglucanases.

We screened the metagenome of 73 *G. thermodenitrificans* strains for genes that match with the HMM profiles of every glycoside hydrolase family known to contain cellulases. A total of 82 proteins matched with one or more profiles. To reduce the number of false positives obtained in this study, a more reliable cut-off threshold can be made using validated proteins of each protein family as a training set against other families. Two putative endoglucanases (GE39 and GE40) of glycoside hydrolase family 5 were identified that have been shown to be located in the HUS locus of the isolation strain. The genetic variation between the HUS locus of T12 and the partial HUS locus encoded by scaffold 9 is remarkable, as previous HUS loci comparison between three *G. thermodenitrificans* strains showed no variation in genetic content (De Maayer et al. 2014). The protein architecture of GE39 contains a GH5 catalytic domain, and active site residues have been identified at positions E188 and E318. The GE39 sequence shows identity to several endoglucanases (Fig. 6) and structure modeling of GE39 revealed a folding similar to a GH5 exo-1,3-β-glucanase from *S. cerevisiae*. Based on protein sequence alignment, we expected GE39 to belong to the GH family 5 endoglucanases. However, protein alignment reveals highest identity (36%) to a GH5 β-xylosidase from *Phanerochaete chrysosporium* (r*Pc*Xyl5) (Huy et al. 2015). For r*Pc*Xyl5 it was shown that it was most active against *p*-nitrophenyl-β-D-xylopyranoside (pNPbX) with minor activities against xylan, which is in agreement with the lack of activity seen in our activity assays on chromogenic xylan and cellulose substrates. Likewise, we found GE39 to be active against *p*-nitrophenyl-β-D-xylopyranoside. In contrast to r*Pc*Xyl5, the GE39 β-xylosidase showed side activity against *p*-nitrophenyl-β-D-glucopyranoside (pNPβX).

Where enzyme GE39 did not show endo-activity, enzyme GE40 was demonstrated to be active against 2-HE-cellulose, CM-cellulose and barley glucan. The protein architecture of GE40 is comprised of a GH5_4 catalytic domain, an immunoglobulin-like module and a carbohydrate binding module of family 46 (Figure 7). Family

46 CBMs, with the ability to bind cellulose and glucan, are always located on the C-terminus of enzymes containing a GH5_4 catalytic domain and an immunoglobulin (Ig)-like module. The Ig-like module is believed to act as a structural hinge, thereby holding the catalytic domain and the CBM in position for optimal enzymatic activity (Liberato et al. 2016). The CBM acts in synergy with the GH5_4 catalytic domain to bind glucans and thereby aid in the hydrolytic cleavage of the substrate (Venditto et al. 2015). Although cellulase activity was clearly demonstrated in the CFE of a *GE40* expressing *E. coli* culture, we could only detect minor activity when *GE40* was expressed in *G. thermodenitrificans* T12 (Figure S3). Furthermore, SDS PAGE analysis of the cell lysate from a *GE40* expressing T12 culture did not show heterologous protein formation and no additional protein bands were visualized (data not shown). We have provided evidence that GE40 is a novel endoglucanase, however, further development of suitable promoters and/or signal peptide sequences is needed to increase protein yield and thereby the extracellular activity required for efficient cellulose conversion.

In an attempt to complement the cellulolytic machinery of *G. thermodenitrificans*, we introduced several cellulases from *C. thermocellum* into strain T12. By removing the dockerin domain we created smaller constructs that were expected to be easier to introduce to *G. thermodenitrificans* T12. The removal of the dockerin domain of CelK and CelS was shown not to reduce the enzymatic activity (Vazana et al. 2010). Therefore, we hypothesize that their removal has not been instrumental in the lack of activity seen in our study. This lack of activity is possibly caused by insufficient protein production as we did not detect heterologous protein by SDS-PAGE analysis on the intracellular protein fraction (data not shown) and no intracellular or extracellular enzyme activity was detected by using chromogenic substrates and HPSEC (data not shown). We therefore hypothesize that problems at the transcriptional- or early translational stage hamper functional expression of these enzymes. Cultures producing T12-CelK, T12-CelA and T12-GE40 did show both extracellular (Figures 8-9, S3) and intracellular cellulolytic activity (data not shown).

Conclusions

This study shows the potential of metagenome mining for the discovery of novel cellulases. Metagenome-derived enzyme GE39 was shown to be a novel GH5 β-xylosidase with some β-glucosidase activity. Enzyme GE40 was shown to be a novel GH5 endoglucanase that is active against amorphous cellulose and barley glucan and had 55% identity to its closest ortholog *Bh*Cel5b from *Bacillus halodurans*. It is the second endoglucanase retrieved from *Geobacillus* and the first found in *G. thermodenitrificans*.

We also demonstrated the ability of *Geobacillus thermodenitrificans* T12 to act as a host for heterologous cellulase expression. Although the degradation of cellulose by strain T12 still requires optimization as activities remained low, the methods described in this study provide a starting point for further development of *Geobacillus spp.* as potential hosts for consolidated bioprocessing.

Supplementary information

Table S1 Overview of measured absorbance of different chromogenic substrates after incubation with the metagenome derived putative cellulases GE39 and GE40 expressed from *E. coli*. The final reaction mixture in each well of the substrate plate consisted of 145 µL sodium phosphate buffer (pH 6.0) and 5 µL of CFE. Plates were then sealed using an aluminum adhesive foil and incubated at 60 °C in a rotary shaker at 180 RPM. After 24 h the reaction mixture was collected in a product plate by centrifugation (2,700 ×g, 10 min) and absorbance was measured at 595 nm (blue) and 517 nm (red) using a plate reader (Biotek Instruments Inc., Winooski, VT, USA). Negative control consisted of sodium phosphate buffer and CFE from an *E. coli* culture containing empty pCDF1b plasmid. The thermostable endoglucanase, CelTM, (Megazyme, Wicklow, Ireland) from *Thermotoga maritima* was used as positive control (+C) at a concentration of 1 µg/mL. Values for the negative control have been subtracted. CFE of the GE40 expressing *E.coli* culture showed high activity towards cellulose and barley derived β-glucan. In contrast, CFE of the GE39 producing *E.coli* culture showed no activity to any of the chromogenic substrates.

Substrate	GE39	GE40	+C
CPH-2-hydroxyethylcellulose	0.002	1.164	1.0705
CPH-β-glucan from barley	0.001	0.382	0.1495
CPH-β-glucan from yeast	0.019	0	0.034
CPH-curdlan	0.017	0.016	0
CPH-pachyman	0.007	0.021	0
CPH-pullulan	0.002	0.003	0
CPH-dextran	0	0.001	0.001
CPH-chitosan	0.1	0	0
CPH-xyloglucan	0.015	0.003	0.043
CPH-amylopectin	0	0.001	0
CPH-amylose	0	0.004	0.008
CPH-xylan	0	0	0.0035
CPH-arabinan	0	0	0.007
CPH-pectic galactan	0.011	0.002	0.002
CPH-galactomannan	0	0	0.1425
CPH-rhamnogalacturonan (soybean)	0	0	0.0225

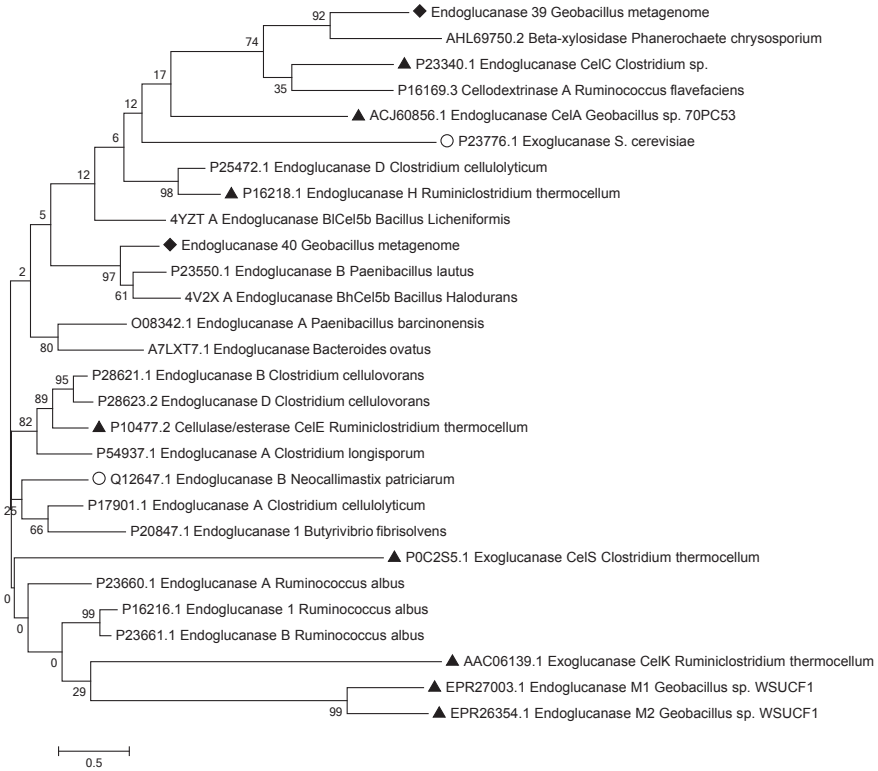


Figure S2 The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al. 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 28 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 146 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016). Open circle: eukaryotic organism; closed triangle: thermophilic organism; closed diamond: sequences obtained in this study.

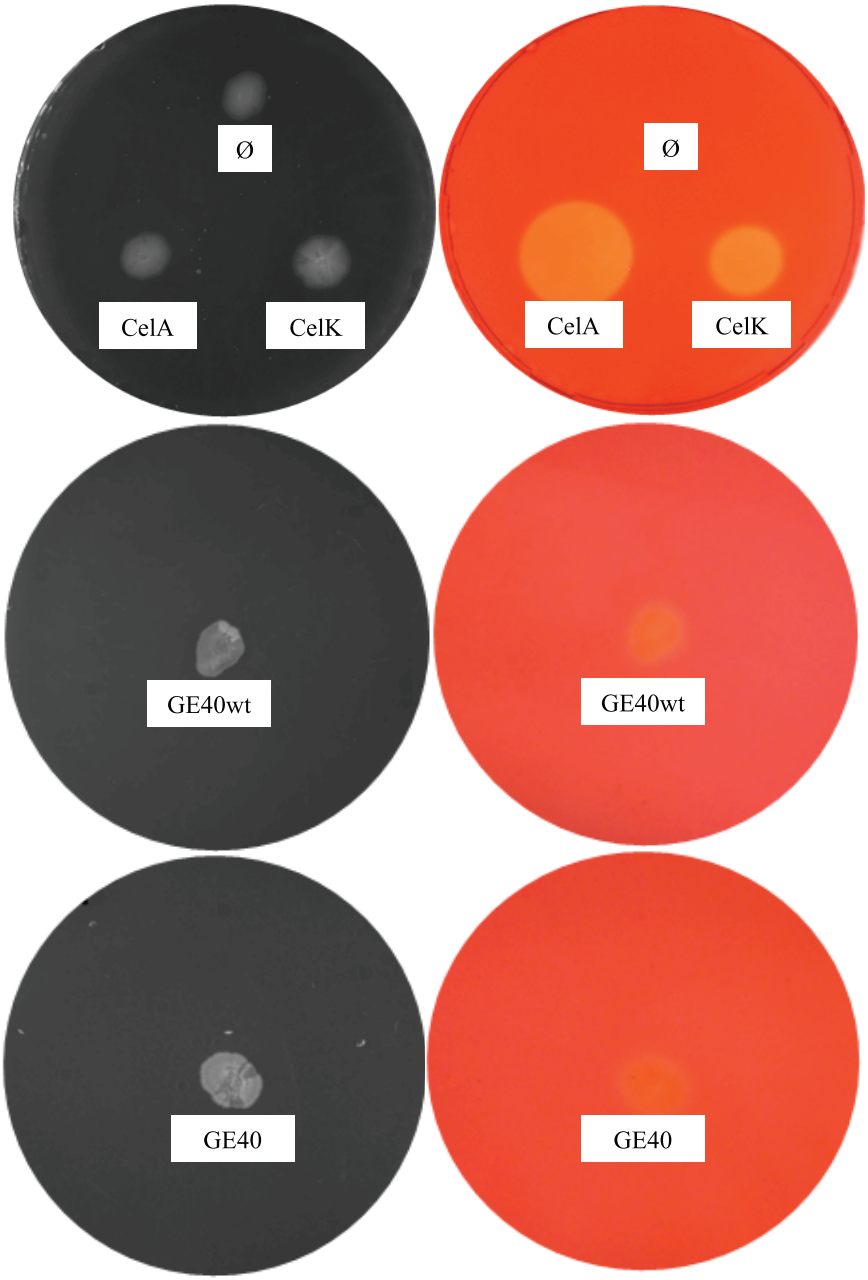
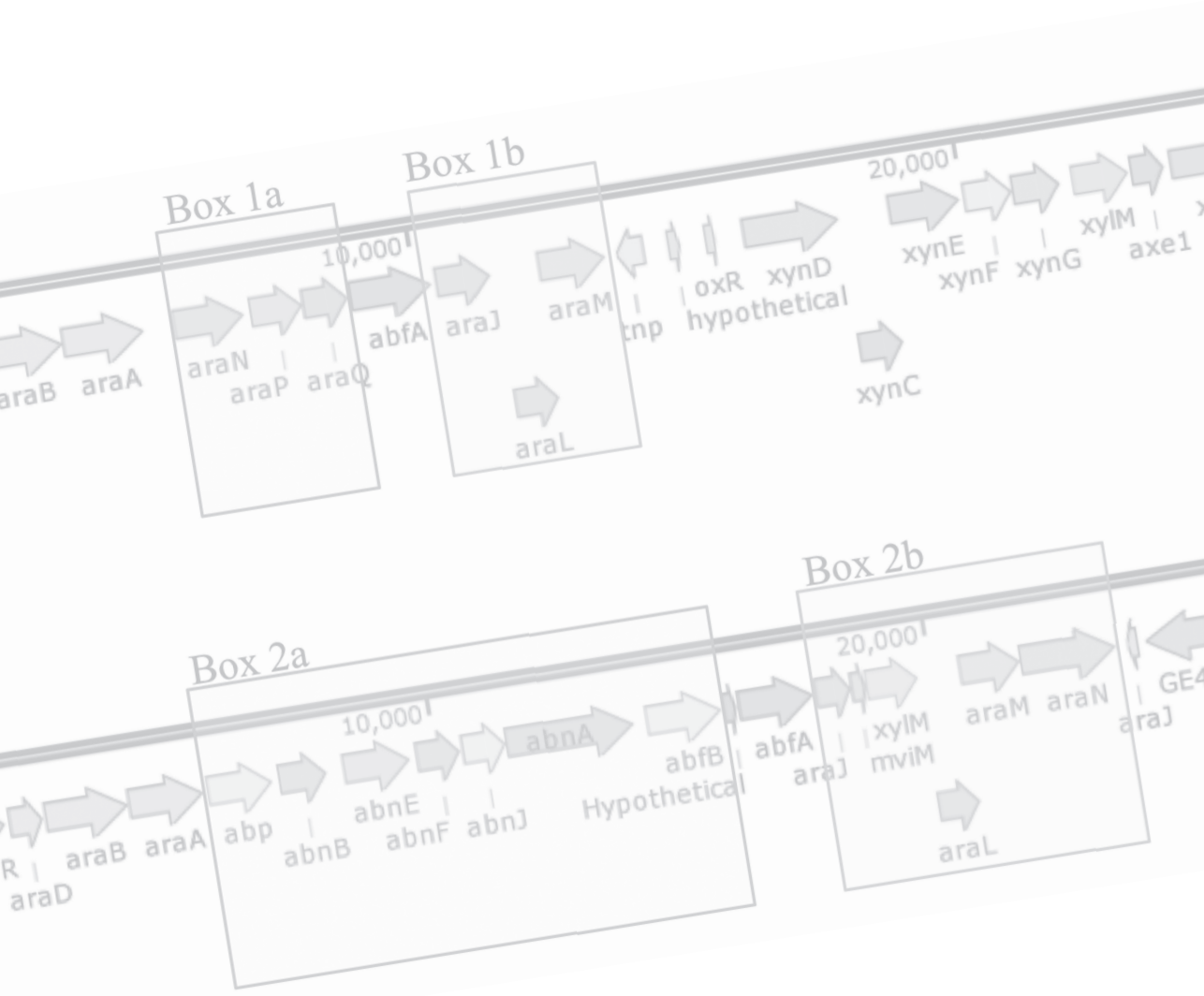
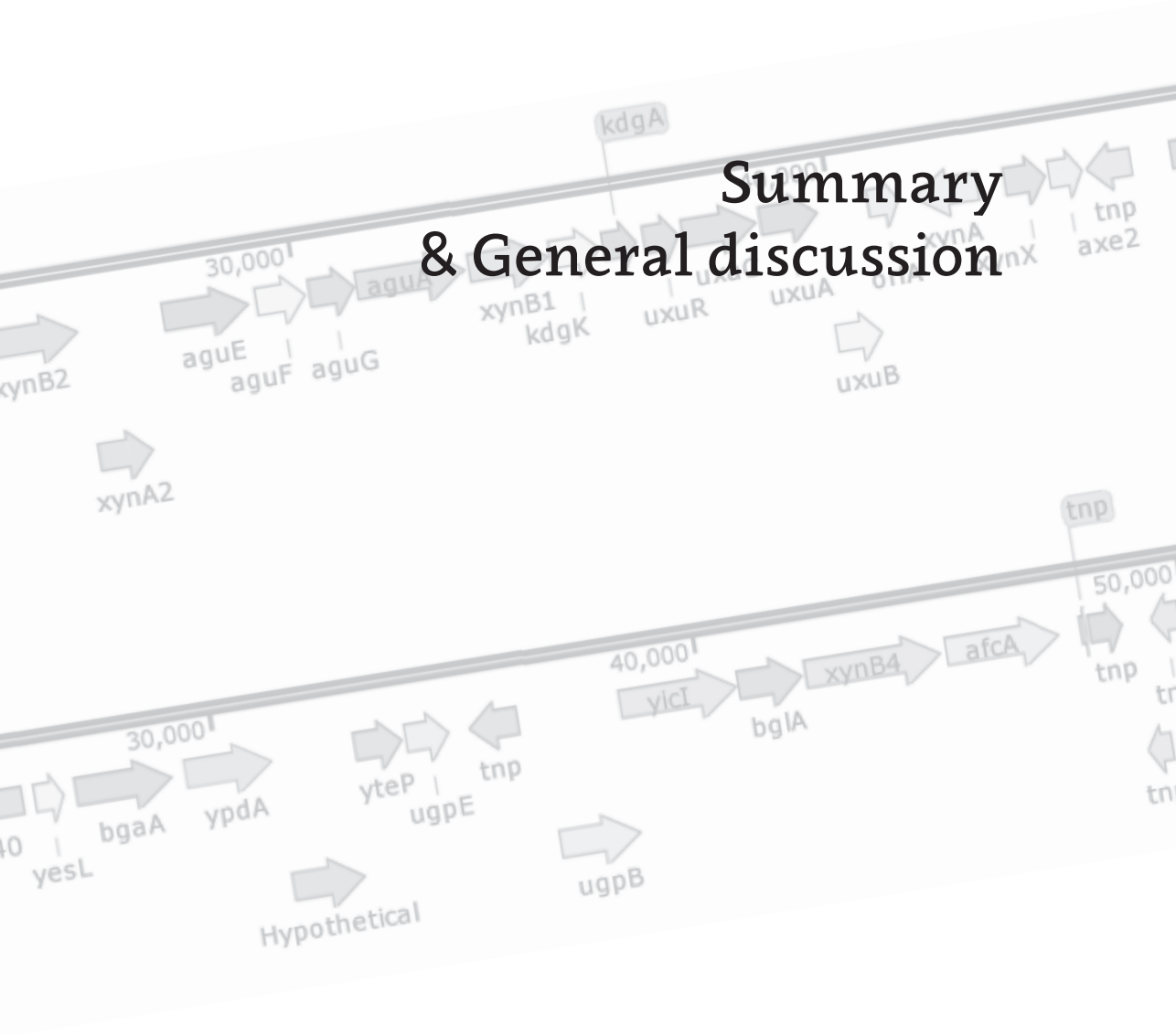


Figure S3 Congo red assays of *G. thermodenitrificans* cultures grown on LB2 medium with 1% carboxymethylcellulose. Each culture produces a different cellulase. Ø: empty plasmid (pNW33n) control; CelA: GH5 endoglucanase CelA (*Geobacillus* 70PC53); CelK: GH9 exoglucanase (*C. thermocellum*); GE40wt: GH5 endoglucanase (*Geobacillus* metagenome derived) containing its native signal peptide; GE40: GH5 endoglucanase (*Geobacillus* metagenome derived).


Box 1a
Box 1b
10,000
20,000
araB
araA
araN
araP
araQ
abfA
araJ
araM
araL
tnp
oxR
xynD
hypothetical
xynC
xynE
xynF
xynG
xylM
axe1
Box 2a
Box 2b
10,000
20,000
R
araD
araB
araA
abp
abnB
abnE
abnF
abnJ
abnA
Hypothetical
abfB
abfA
araJ
xylM
mviM
araL
araM
araN
araJ
GE4

Chapter 6

Summary & General discussion



Summary

The growing demand for consumables and energy, combined with increasing consciousness over environmental issues like global warming, faces us with the challenge to find alternatives for fossil resources. Alternative production methods for energy, like windmills, solar panels and hydroelectricity plants, are far developed and have become economically competitive to fossil resource-based production processes. However, the production of many (bulk) chemicals and products is still dominated by the petroleum industry. One such chemical is lactic acid, a fermentation product of many bacteria and a compound that is gaining interest as a building block for poly lactic acid (PLA). PLA is a polymer used to produce bioplastics, and thereby provides an alternative to petroleum-based plastic production. As described in **Chapter 1**, economically feasible production of lactic acid is envisioned through consolidated bioprocessing (CBP). In a CBP process, pretreated lignocellulosic biomass is hydrolyzed to fermentable sugars and those sugars are subsequently fermented to desired product in one reaction vessel. The organism of choice for this hydrolyzation and fermentation is preferentially a thermophile, capable of enzyme production and lactic acid fermentation.

Species from the genus *Geobacillus* have many of the desired characteristics, and in **Chapter 2** we have enriched and isolated facultative anaerobic (hemi)cellulolytic *Geobacillus* strains from compost samples. By selecting for growth on both cellulose and xylan, 94 strains were isolated. Subsequent screening for lactic acid production was carried out from C6 and C5 sugar fermentations and a selection of the best lactic acid producers was made. The denitrifying *Geobacillus thermodenitrificans* T12 was selected for further research and was rendered genetically accessible with a transformation efficiency of 1.7×10^5 CFU/µg of plasmid DNA. In fermentations on a mixture of glucose and xylose, a total of 20.3 g of lactic acid was produced with a yield of 0.94 g product/g sugar consumed. In addition, we demonstrated that strain T12 is capable of direct conversion of beechwood xylan to mainly lactic acid in minimal media.

Chapter 3 describes the genome sequencing and several features of *G. thermodenitrificans* T12. The genome of strain T12 consists of a 3.64 Mb chromosome and two plasmids of 59 kb and 56 kb. It has a total of 3.676 genes with an average genomic GC content of 48.7%. The T12 genome encodes a denitrification pathway, allowing for anaerobic respiration. The identity and localization of the responsible genes is similar to those of the denitrification pathways found in strain NG80-2. The host-defence systems present comprise a Type II and a Type III restriction-modification system, as well as a CRISPR-Cas Type II system that could potentially be exploited as a genome editing tool for thermophiles. Furthermore, the hemicellulose utilisation (HUS) locus of strain T12 appeared to have orthologues for all the genes that are present in strain T-6 except for the arabinan degradation cluster. Instead, the HUS locus of strain T12 contains genes for both an inositol and a pectate degradation pathway.

The HUS-locus associated gene, *GtxynA1*, encodes an extracellular endoxylanase of strain T12, and belongs to the family 10 glycoside hydrolases (GH10). In **Chapter 4**, we describe the cloning, expression and characterization of GtXynA1. The recombinant endoxylanase was purified to homogeneity and showed activity between 40°C and 80°C, with an optimum activity at 60°C, while being active between pH 3.0 to 9.0 with an optimum at pH 6.0. Its thermal stability was high and *Gt*XynA1 showed 85% residual activity after 1 h of incubation at 60°C. Highest activity was demonstrated towards wheat arabinoxylan (WAX), beechwood xylan (BeWX) and birchwood xylan (BiWX). *Gt*XynA1 can degrade WAX and BeWX producing mainly xylobiose and xylotriose. To determine its mode of action, we compared the hydrolysis products generated by GtXynA1 with those from the well-characterized GH10 endoxylanase produced from *Aspergillus awamori* (*Aa*XynA). The main difference in the mode of action between *Gt*XynA1 and *Aa*XynA on WAX is that *Gt*XynA1 is less hindered by arabinosyl substituents and can therefore release shorter oligosaccharides. The extensive hydrolysis of branched xylans makes this enzyme particularly suited for the conversion of a broad range of lignocellulosic substrates.

The enzymatic conversion of cellulose to glucose requires the synergistic action of three types of enzymes: exoglucanases, endoglucanases and β-glucosidases. The thermophilic, hemicellulolytic *Geobacillus thermodenitrificans* T12 was shown to be a potential candidate for CBP but lacks the desired endo and exoglucanases needed for the conversion of cellulose. In **Chapter 5** we report the heterologous expression of endoglucanases and exoglucanases by *G. thermodenitrificans* T12, in an attempt to complement the enzymatic machinery of this strain and its suitability for consolidated bioprocessing. A metagenome screen was performed on the metagenome of 73 *G. thermodenitrificans* strains using HMM profiles of all known CAZy families that contain endo and/or exoglucanases. Two putative endoglucanases, GE39 and GE40, belonging to glucoside hydrolase family 5 were isolated and expressed in both *E. coli* and *G. thermodenitrificans* T12. Structure modeling of GE39 revealed a folding similar to a GH5 exo-1,3-β-glucanase from *S. cerevisiae*. However, we determined GE39 to be a β-xylosidase having most activity towards *p*-nitrophenyl-β-d-xylopyranoside. Structure modelling of GE40 revealed a protein architecture similar to a GH5 endoglucanase from *B. halodurans*, and its endoglucanase activity was confirmed by enzymatic analysis against 2-HE-cellulose, CM-cellulose and barley β-glucan. In addition, we successfully expressed the earlier characterized *Geobacillus* sp. 70PC53 endoglucanase *celA* and the *C. thermocellum* exoglucanase *celK* in strain *G. thermodenitrificans* T12. The native hemicellulolytic activity and the heterologous cellulolytic activity described in this research provide a good basis for the further development of *Geobacillus thermodenitrificans* T12 as a host for consolidated bioprocessing.

In **Chapter 6**, we provided more insight in the genetic variation of the hemicellulolytic utilization cluster of *G. thermodenitrificans*. This variation is far greater than described before and gives ample opportunities for the further development of *Geobacillus* spp. for hemicellulose degradation. The production of cellulases in *Geobacillus* species is demonstrated to be successful, and we have expanded on that knowledge with the expression of both endo

and exoglucanases from *C. thermocellum*. However, in line with previous studies, direct cellulose fermentation by geobacilli is not yet achieved, most likely due to insufficient cellulase production and/or secretion. With a rapidly expanding genetic toolbox for thermophiles, now including a thermostable Cas9, we expect that the successful development of *Geobacillus* spp. for consolidated bioprocessing is just a matter of time.

General discussion

The use of thermophiles in consolidated bioprocessing (CBP) has gained much interest in the last decade. This has boosted research and resulted in substantial gains in knowledge and applicability of those strains. The research outlined in this thesis focusses on relevant physiological and molecular features of *Geobacillus,* and to evaluate its potential as a host for consolidated bioprocessing. This closing chapter places the results from this thesis in a wider context and discusses the value of the work presented and the challenges that remain in using *Geobacillus* for CBP.

The isolation and classification of *Geobacillus* spp.

The interest for *Geobacillus* spp. has increased over the last years, either as a source of useful thermostable enzymes, or as cell factories for bioremediation, biofuel production and degradation of lignocellulose [1, and references therein]. As discussed in Chapter 1, many studies have focused on the isolation of *Geobacillus* spp. from a wide variety of environments, yielding an array of isolates all with their own unique features. In our study, we isolated novel thermophilic strains of *Geobacillus* from compost heaps (35°C and 65°C) consisting solely of plant material. The isolation in this study was performed by plating dilution series directly from the compost. By using this selection method, we directly selected for strains capable of growth on carboxymethylcellulose (CMC) and xylan, as well as their tolerance to aerobic conditions. We surprisingly found only *Geobacillus* spp., or after recent re-classification, *Geobacillus* and *Parageobacillus* spp. (Aliyu et al. 2016). In addition to plating dilution series of the compost sample, we also performed shake flask enrichments on cellulose. The primary enrichment was started by adding compost to medium using cellulose (Whatman filter paper) as sole carbon source. Enrichments were transferred each 4 to 8 days based on visual determination of breakdown, and in all nine consecutive transfers we noticed degradation of the cellulose paper, in both anaerobic and aerobic cultures. Anaerobic and aerobic enrichments were used to obtain pure isolates by dilution plating

according to the same protocol as described for the pure compost sample (Chapter 2). Again, only *Geobacillus* spp. isolates were obtained indicating the stringent selection of the used isolation method by which potential anaerobic species will be eliminated due to oxygen exposure. Another crucial factor in obtaining only *Geobacillus* isolates is the time of sampling during the composting process. Samples used in this study were obtained from both a thermophilic stage of composting, and from the matured stage of composting from which strain T12 was isolated (Chapter 2). It is well known that the microbial community of compost changes as the composting process progresses from its mesophilic stage to the thermophilic stage. DNA-based microbial community analysis tends to give a biased indication of the species present as DNA from dormant and even dead cells is included (Li et al. 2014). Meta-transcriptome data generated from a thermophilic composting process reveals the most abundant genera to be *Symbiobacterium* and *Thermaerobacter* for Clostridiales, *Geobacillus* and *Ureibacillus* for Bacillales and *Thermopolyspora* and *Thermobispora* for Actinomycetales. *Symbiobacterium* is known to be cultivatable only in the presence of *Bacillus sp.* and requires an anoxic environment with several still unknown growth factors. *Thermaerobacter* isolates have been reported to grow at more alkaline conditions, although some have been demonstrated to grow at pH levels as low as 5.0. However, *Thermaerobacter* spp. require rich media usually containing high amounts of yeast extract and tryptone (Yabe et al. 2009). *Ureibacillus* are aerobic and were found unable to grow and produce organic acids from all major biomass-derived sugars (Fortina et al. 2001). Genera from the *Actinomycetales* order are among the most abundant species in thermophilic compost and although they have previously been isolated from environments like the compost used in our study, we did not isolate any strains from this order. Considered all the above, it is not surprising that we isolated *Geobacillus*. However, it is striking that *G. thermodenitrificans* is highly overrepresented (79%) within our selection. Although hard to define, it seems that this species has developed a certain niche within the sampled compost that gives it an advantage over the

other members of this genus. Deeper analysis of the genomic region containing most of the hydrolytic enzymes, the hemicellulose utilization (HUS) locus, might give clues towards this advantage of *G. thermodenitrificans*.

The identification of isolated strains in our study was based on 16S RNA encoding genes. This method is widely accepted as a valid tool for the identification at the species level although in the case of *Geobacillus* it is impossible to distinguish *G. kaustophilus* and *G. thermoleovorans*. Recently these two species have indeed been merged as *G. thermoleovorans* (Aliyu et al. 2016). The taxonomy of *Geobacillus* spp. is, like many other genera, prone to name adjustments and reclassifications. Illustrative in this respect is the renaming of several *Bacillus* spp. to *Geobacillus* spp. in 2001 (Nazina et al. 2001). Not only were bacilli now renamed to *Geobacillus*, a *Saccharococcus* strain was included as well. Just recently, several *Geobacillus* spp. have been reclassified as *Parageobacillus* and are considered a separate phylogenic cluster (Aliyu et al. 2016). The renaming of *G. thermoglucosidasius* to *P. thermoglucosidans* and the taxonomic revision of several *Geobacillus* spp. to *Anoxybacillus* spp. has also led to more obscurity in that many researchers stick to the original naming and others do not. Strain identification based on 16S and *recN* are valuable methods, if identification based on available knowledge is desirable, as was the case in this study. However, from time to time a disruptive approach can reclassify the current order of genera and species.

The selection and characterization of *Geobacillus thermodenitrificans* T12

From the compost culture collection, we selected *G. thermodenitrificans* T12 based on reproducible xylan degradation and genetic accessibility for further research. Recently, Bosma *et al.* (Bosma et al. 2015) isolated several *G. thermodenitrificans* strains that were amenable to electroporation, albeit with low efficiencies (CFU/µg DNA ≤ 10). This and other studies on *Geobacillus* spp. have demonstrated that genetic accessibility is highly strain specific, and no universal protocol is available for any *Geobacillus* spp. (Kananavičiute & Čitavičius 2015;

Bosma et al. 2015). Genetic modification of some *(Para)Geobacillus* strains has been reported with the most notable effort by Cripps et al. (Cripps et al. 2009), who achieved functional expression of a heterologous pyruvate decarboxylase in *P. thermoglucosidans*. In addition, a mutant strain was developed that lacked the *ldhL* and *pfl* genes, encoding L-lactate dehydrogenase and pyruvate-formate lyase, respectively, and produced ethanol as its major end-product. More recently, a modular Clostridia-based genetic toolkit was redesigned for its use in *Parageobacillus*. In 30 days a mutant strain (in-frame mutations of *ldh* and *pfl*) was created using allele-coupled exchange which alleviates the use of extensive double cross-over screening (Sheng et al. 2017).

Recent reclassification of *G. thermoglucosidans* as a member of the genus *Parageobacillus*, leaves *G. kaustophilus* HTA426 as the only *Geobacillus* strain that is both genetically accessible and fully sequenced. Transformation of strain HTA426 is achieved by conjugative transfer with *E. coli* as a donor, and a mutant strain, MK244, in which two sets of restriction/modification (R/M) system encoding genes have been disrupted, can be transformed with even greater efficiencies (H. Suzuki et al. 2013). The achievements with this strain give rise to its use as a host for the heterologous expression of proteins and the construction of gene libraries. Its potential for CBP, however, is limited due to the lack of hydrolytic enzymes encoded on its genome. Strain HTA426 lacks the HUS locus present in many other geobacilli and is thereby not able to hydrolyze hemicellulose.

As described in this thesis, *G thermodenitrificans* T12 was shown to be genetically accessible. The efficiency of transformation was greatly improved by adjusting the concentration of K_2HPO_4 the growth medium. The change in salt concentration of the growth medium has a severe effect on the morphology of the cells; cells grown in high salt concentrations (2.50 g/L K_2HPO_4) being easy to pellet and cells grown in low salt concentration (LB medium w/o added salts) remaining unpelletable even after prolonged (up to 1 hour) centrifugation of the culture. Growing the cells in medium with intermediate concentrations of K_2HPO_4 gave a firmer pellet that yielded highest transformation efficiencies. When the difference in

salt concentration was mimicked by adjusting NaCl concentrations, we did not detect a notable difference in pellet formation, indicating that K_2HPO_4 is instrumental to our observations. The concentration of K_2HPO_4 seems to influence the cell wall composition causing changes in the cell's morphology, and as such of the pelleting features. However, the mechanism behind this morphological change was not elucidated. The efficiency of transformation was even further increased when using a pNW33n-cured T12 strain. We hypothesized that one or more R/M systems were inactivated, *e.g.*, by insertion of a mobile element, but resequencing of the genome yielded no clues in this direction. In addition, also the CRISPR/Cas system was not interrupted by mobile elements.

The CRISPR/Cas Type II-C system on the genome of T12 was an important discovery as it can be the key to moving metabolic engineering of *Geobacillus*, and thermophiles in general, to the next level. Comparison of the CRISPR array between T12 and other *Geobacillus* spp. harboring a similar CRISPR/Cas system revealed differences in the spacer sequences. Therefore, we hypothesize that the system is not a remnant of homologous gene transfer but is in fact an active CRISPR/Cas system. The Cas9-based CRISPR/Cas toolkit has increased the possibilities of genome engineering in multiple strains, and is now widely used in various organisms (Mougiakos et al. 2016). For now, this discovery is the start in the development of a Cas9 toolkit for thermophiles, and exciting developments are to be expected in the near future.

Hemicellulolytic activity

Hemicellulosic substrates consist of complex polysaccharides that differ in composition and structure based on their origin (Mäki-Arvela et al. 2011). The most relevant hemicelluloses are xylans which constitute about 20-50 % of the biomass of woods and grasses (Gírio et al. 2010; Murciano Martínez et al. 2015). Due to the abundance of different structures and sugars present in hemicellulose, a wide variety of enzymes are needed to fully degrade this substrate to fermentable sugars. Chemical pretreatment is often applied to reduce the complexity of the substrate. In general, two types of pretreatments

are described in literature: acidic and alkaline pretreatment. Acidic pretreatment is used to hydrolyze the hemicellulose fraction and thereby generating improved access for enzymes active against cellulose. On the other hand, alkaline pretreatments are mainly used to solubilize lignin which results in better enzyme accessibility towards cellulose and hemicellulose (Murciano Martínez et al. 2015) (Figure 1). The hemicellulosic polysaccharides obtained after alkaline pretreatment consist in general of glucuronoxylans (hardwoods), arabinoxylans (cereals) and arabinoglucuronoxylans (agricultural residues). Glucuronoxylans contain about one residue of methyl glucuronic acid per ten xylose residues (Fig. 1a) and the concerted action of both GH10 endoxylanases and GH67 α-glucuronidases is sufficient for complete breakdown of this substrate. In Chapter 4 we demonstrated the ability of *Gt*XynA1 (a GH10 endoxylanase) to hydrolyze xylans substituted with glucuronyl residues and identified *aguA* (GTHT_01356) to encode an intracellular GH67 α-glucuronidase. It is known that short xylo-oligosaccharides substituted with glucuronyl residues can be transported, via the AguEFG transporter, into the cell by *G. stearothermophilus* T-6 (Shulami et al. 1999), and we assume a similar mechanism for strain T12 based on protein sequence identities (Chapter 3 and 4). The substituted xylo-oligomers are degraded intracellular by both GH67 α-glucuronidases, GH10 endoxylanases and β-xylosidases of the GH39, GH43 and GH52 families (Shulami et al. 2014) (Chapter 3). The degradation of arabinoxylans is like that of glucuronoxylans, in that short-substituted oligosaccharides can be taken up by the cell and are degraded further by intracellular enzymes. In the case of arabinoxylans the arabinose side-groups are cleaved of by an intracellular GH51 α-L-arabinofuranosidase AbfA (GTHT_01376) Arabinoglucuronoxylans, the main hemicellulose fraction of grasses, wheat bran and corn residues, can differ greatly in their degree of branches on the backbone, depending on the plant part they have been extracted from (Biely et al. 2016). The substituents of arabinoglucuronoxylans are arabinose, glucuronic acids, and in mild alkali treatments a portion of ferulic acid esters persists in the xylan structure. Ferulic acid was shown not to hinder the action of GH10 endoxylanases (Kormelink et al. 1993) but the

dense substitution of most arabinoglucuronoxylans results in a high degree of double substituted xylosyl residues on the xylan backbone. Although GH10 is not hindered by these double substitutions (Chapter 4**),** it is unknown whether those oligosaccharides can be taken up by the cell and subsequently if the α-glucuronidase AguA and α-L-arabinofuranosidase AbfA are able to cleave such densely packed residues. Previous studies show GH51 activity towards single substituted residues, with double substitutions being reduced by GH43 α-L-arabinofuranosidases, although some GH51 enzymes were reported to act on di-arabinosylated xylose residues (Cartmell et al. 2011).

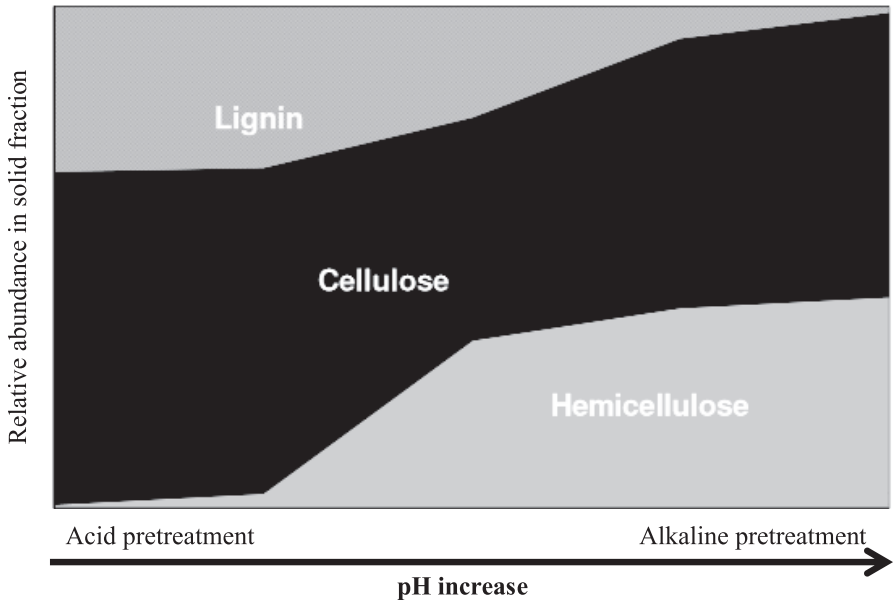


Figure 1 Relative abundance of lignin, cellulose and hemicellulose in the solids remaining after acid or alkaline pretreatment of lignocellulosics (Murciano Martinez, 2016)

The variability among the hemicellulose utilization (HUS) loci, of the *Geobacillus* spp. sequenced to date, is mainly found at the 5' end of the locus (De Maayer et al. 2014). The T12 HUS locus extends those reported before, most likely reflecting the variability in *G. thermodenitrificans* hemicellulose degradation

potential. The additional cluster for pectate degradation in the HUS locus of T12 probably enables this strain to degrade galactan and rhamnogalacturonan I (Chapter 2), two components of pectin. Pectins are commonly found in the cell wall of cereals, fruits and vegetables where they crosslink cellulose and hemicellulose fibers. The degradation of pectin thus improves the accessibility of cellulases and hemicellulases to their substrates (Kohli & Gupta 2015). We hypothesize that the pectate degradation cluster of T12 enhances its hydrolytic activity on fruits and vegetables, but might also positively influence its ability to hydrolyze other plant derived substrates as they contain pectin as well (Hoondal et al. 2002).

Recently, De Maayer et al. (2014b) suggested the HUS locus to be an insertion point for the centralization of sugar metabolic pathways in *Geobacillus spp*. Our finding of a pectate degradation cluster in the T12 HUS locus is in line with this theory. We also revealed major variation among the HUS loci of strain T81 and strain T12, both identified as *G. thermodenitrificans*, which contrasts with the previous genome comparison of the HUS loci from *G. thermodenitrificans* NG80-2, DSM465 and G11MC16 which revealed no difference in their clusters (De Maayer et al. 2014; Brumm et al. 2015). The genetic variability between strain T12 and strain T81 (Chapter 5) demonstrates more diversity among *Geobacillus* strains than reported to date (De Maayer et al. 2014; Brumm et al. 2015). The single scaffold (scaffold 9) of strain T81that was analyzed in Chapter 5 showed remarkable differences in gene localization and pathways present between the HUS loci of strains T12 and T81. Whereas strain T12 has genes encoding for arabinose transport (Figure 2, box 1a; Figure 3, cluster K), strain T81 has genes for a complete arabinan degradation and transport cluster (Figure 2, box 1b) like the arabinan cluster previously found in *G. stearothermophilus* T-6 (De Maayer et al. 2014). Strain T81 has all the genes encoding for an alternative pentose phosphate pathway (Figure 2, box 2b) whereas strain T12 lacks the *araN* gene (Figure 2, box2a; Figure 3, cluster I*). In addition, the novel endoglucanase GE40 (chapter) was identified on the T81 HUS locus and is located upstream of a sugar ABC transport operon (*ypdA*, *Hypthetical_2*, *yteP*, *ugpE* and *ugpB*). However, the substrate specificity of this ABC transporter is unknown.

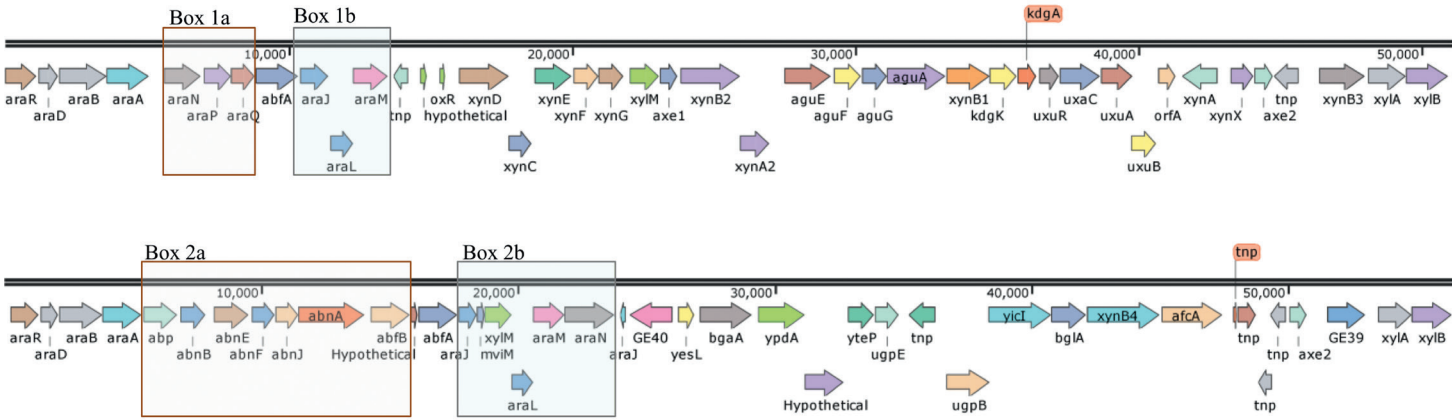


Figure 2 Schematic diagram of the 3' end of the hemicellulose utilization (HUS) locus of *G. thermodenitrificans* T12 (top) and T81 (bottom). Each arrow represents a gene in the locus. Box 1a: arabinose transport, cluster K; box 1b: incomplete alternative pentose sugar metabolism, cluster I; box 2a: arabinan degradation and transport; box 2b: complete alternative pentose sugar metabolism, cluster I.

In contrast to strain T12, strain T81 lacks the genes required for extracellular xylan degradation (*xynA*) and intracellular xylo-oligosaccharide degradation (*xynA2*, *xynB2*, *axe1* and *xylM*) indicating its inability to degrade xylans. Additionally, strain T81 lacks both pathways for arabinose transport and the pathway for arabinose metabolism. An overview of the genes, located on the partial HUS locus of strain T81, and their predicted function is given in Table S1. To further analyze the genetic variability among HUS loci of *G. thermodenitrificans*, we sequenced the genomic DNA of all 73 *G. thermodenitrificans* isolates obtained in this study separately, and performed blast analysis of the HUS locus derived genes from strain T12 against the obtained 73 (gapped) genomes. We could identify all orthologs of genes present in the HUS locus of T12 and by dividing the HUS associated genes into clusters we attempted to predict whether clusters are functionally present or disrupted due to missing genes or low sequence identity (Figure 3, Table S2). Multiple strains seem to lack the clusters associated with extracellular xylan degradation, xylo-oligosaccharide transport and intracellular xylan degradation and are likely not capable of xylan degradation. The pectate degradation cluster (Q), present in T12, can be found in about one third of our isolates, further demonstrating the variability of the HUS loci even among strains from the same sampling site and the same species. Comparison between HUS locus associated genes from *G. stearothermophilus* T-6 and *G. thermodenitrificans* T12 revealed over 70% sequence identity for all orthologs. Therefore, we assume that sequence identity based comparison of HUS loci among *G. thermodenitrificans* strains is a reliable method for detecting orthologs. Our analysis is limited to predicting the presence or absence of orthologs with high sequence identity to T12 genes, thus, neither the presence of genes with other activities, nor the presence of orthologs with low sequence identity to T12 genes was determined. When we analyzed the HUS locus of T81 (chapter 5), we demonstrated that β-xylosidase GE39 has a similar function as XynB3, although their protein sequences showed only 11% amino acid sequence identity. Therefore, clusters presented as “likely non-functional” or “non-functional” might potentially be

functional due to the presence of genes with low sequence identity to their T12 orthologs. In addition, cluster B, encoded by *xynB3* and *axe2*, is presented as "likely functional" (Figure 3) because we detected an *xynB3* ortholog with 69% sequence identity, however, this gene is not located at its usual position upstream of *axe2*. We have not determined the position of the *xynB3* ortholog and in our analysis, we did not include GE39 as a functional protein of cluster B. Further analysis of each HUS locus and the visualization of their gene distribution might give more insight in the genetic variability, possible leading to the confirmation of more functional clusters due to orthologs with low sequence identity and the presence of novel genes.

Xylanolytic *Geobacillus* spp. rely on the detection of free xylose for the induction of their core xylanolytic machinery (Shulami et al. 2014). Instrumental in this machinery is the production of the extracellular endoxylanse (XynA) which can cleave the backbone of xylans, thereby generating shorter oligosaccharides that will further induce the gene expression on the HUS locus. The production of *Gt*XynA1 and its orthologs in other *Geobacillus* spp. seems to be controlled by a variety of repressors (Shulami et al. 2014). This control of production is mainly due to the repression by CodY, XylR and CcpA. The repression by CodY is induced in rapidly growing cells in high nutrient medium (Sonenshein 2005) and was demonstrated in our study to reduce the production of *Gt*XynA1 when T12 was grown in media containing 5 g/L yeast extract (Chapter 2). Growth rate and maximum cell densities increase with increasing concentrations of yeast extract (data not shown) although genome analysis suggests the presence of complete pathways to produce all amino acids and vitamins, except D-biotin (Chapter 3). Further optimization of the growth medium is highly desirable from an industrial point of view and reduction of the amount of yeast extract will further reduce the effect of CodY on xylanase production. XylR is a transcription regulator that can bind the promoter region of the *G. stearothermophilus* T-6 endoxylanase encoding gene (*xynA*) and is thought to have the same function in strain T12 where multiple XylR binding-sites have been identified in the *gtxynA1* promoter region

(data not shown). The repression by XylR is prevented by xylose, but glucose-6-phosphate acts as an inducer of XylR repression and is therefore thought to play a role in catabolite repression. Carbon catabolite repression is in mostly regulated by the carbon catabolite protein A (CcpA). CcpA was shown to repress the expression of *xynA* in *G. stearothermophilus* strain T-6 but, although the CcpA protein is also encoded on the genome of *G. thermodenitrificans* strain T12, we were not able to detect binding sites of CcpA neither in the GtXynA1 promoter region nor in the *gtxynA1* gene itself (data not shown). Although we have demonstrated the co-utilization of glucose and xylose, the catabolite repression towards the expression of *gtxynA1* is a potential bottleneck for any process in which hemicellulose and cellulose are degraded and fermented. A XylR defective mutant (M-7) was shown to have less repression of xylanase production in the presence of glucose, although catabolite repression still occurs *via* CcpA. Recently, the heterologous overexpression of *xynA* from *G. thermodenitrificans* NG80-2 in *P. thermoglucosidans* C56-YS93 was demonstrated by using a xylose inducible promoter (P_{xylA}). An increase in xylanase activity was observed in the supernatant indicating successful protein secretion and activity. However, the production of extracellular xylanase is downregulated in the presence of glucose, as the P_{xylA} promoter region used contains both a CcpA and a XylR binding site (Pogrebnyakov et al. 2017). In our study, the overexpression of *gtxynA1* in strain T12, using a constitutive promoter (P_{uppT12}), did not result in the degradation of xylan in high nutrient medium, suggesting that no additional active xylanases were secreted from the cells (data not shown). One of the potential bottlenecks in the overexpression of *xynA* is weak promoter activity, as no additional XynA product could be visualized on SDS-PAGE.

Cellulolytic activity of *Geobacillus* spp.

Cellulose is a polysaccharide that is composed of β-1,4-linked D-glucose units that form hydrogen-bonded crystalline fibers. Due to its crystallinity and insolubility it is a substrate that is highly recalcitrant to enzymatic degradation. The enzymatic degradation requires the concerted action of several enzymes;

Cluster	Function	Cluster	Function	Cluster	Function
A	Xylose metabolism	G	Intracellular xylo-oligosaccharide degradation	M	L-Arabinose transport 2
B	Intracellular xylo-oligosaccharide degradation	H	Xylo-oligosaccharide transport	N	Inositol metabolism
C	Extracellular xylan degradation	I*	Alternative pentose sugar metabolism pathway	O	Ribose transport
D	Glucuronic acid metabolism	J	Intracellular arabinofuranose degradation	P	Peptide degradation and transport
E	Intracellular aldotetraouronic acid degradation	K	L-Arabinose transport 1	Q	Pectate degradation, transport and metabolism
F	Aldotetraouronic acid transport	L	Arabinose metabolism		

Figure 3: Overview of the genetic variability of *G. thermodenitrificans* HUS loci. All genes present on the HUS locus of strain T12 were blasted against each of the 73 *G. thermodenitrificans* isolates and genes were clustered based on their predicted function. Clusters with 100% of included genes having ≥75% identity to their T12 orthologs are colored green, clusters with ≤50% of included genes having <75% identity to their T12 orthologs are colored yellow, clusters with ≥50% of included genes having between 50-75% identity to their T12 orthologs are colored orange and clusters with ≥50% of their included genes having <50% identity to their T12 orthologs are colored red. * indicates that this clusters is not complete in T12 and is therefore never marked green in our analysis as we only compare T12 genes to the dataset. Included genes per cluster are given in Table S2.

endoglucanases, exoglucanases and β-glucosidases. Endoglucanases cleave the cellulose chains in a random manner, thereby opening the structure and generating cello-oligosaccharides of various lengths. Exoglucanases act on these oligosaccharides by removing cellobiose from the reducing- and nonreducing ends. Finally, the β-glucosidases split the cellobiose in two glucose molecules. *Geobacillus thermodenitrificans* contains several intracellular 6-phospho-β-glucosidases that require the direct uptake of cellobiose into the cell. In strain T12 no endoglucanases or exoglucanases were found. However, we did identify a 6-phospho-β-glucosidase (GTHT_01331) with a downstream putative cellobiose specific phosphotransferase uptake mechanism (GTHT_01332 through GTHT_01334) just upstream of the HUS locus on the genome of T12. Furthermore, three β-glucosidases were identified scattered over the genome (GTHT_01847, GTHT_02694 and GTHT_02696). In our study, during the isolation of novel strains (Chapter 2), we noticed growth on carboxymethylcellulose (CMC) by several isolates.

This growth on CMC led to the metagenome analysis described in (Chapter 5) where we screened the metagenome of 73 *G. thermodenitrificans* strains using HMM profiles of all known CAZy families that contain endo and/or exoglucanases. Surprisingly, we could only identify a single strain (T81) containing an endoglucanase. The growth on CMC by the other strains was most likely facilitated by β-glucosidases and other enzymes with activity towards CMC, for instance after cell lysis. Cell lysis is often seen in *Geobacillus* when grown on nutrient-limited media and all *Geobacillus* spp. sequenced to date are predicted to contain β-glucosidases. An additional possibility is the cell lysis induced release of several peptidases with activity towards CMC. In Chapter 5, we discussed the potential wrong annotation of endoglucanases M1 and M2 from *Geobacillus* WSUCF1 which seem to be peptidases (Rastogi et al. 2010). Orthologs of these enzymes are present in strain T12 with high sequence identity (>90%) and appear to be widely present in *Geobacillus* spp. based on BlastP analysis. We were not able to detect activity of these enzymes against CMC but our analysis cannot exclude that M1 and M2 have some activity towards CMC. The release of these enzymes might,

therefore, result in growth due to the release of sugars from CMC. When *G. thermodenitrificans* T71, another isolate from our study, was grown on CMC, we could detect the degradation of CMC by HPSEC. However, when we screened the genome of T71 we did not detect cellulases, implying that this strain either contains proteins that have cellulolytic activity towards CMC, although these proteins are not true cellulases, or that the *Geobacillus* spp. in our study contain cellulases that are not covered by any of the Hidden Markov models for endo and exoglucanases. In strain T71, we did identify orthologs for proteins M1 and M2, previously demonstrated to have at least some activity against CMC (Rastogi et al. 2010).

The attempts to complete the cellulolytic enzyme-set of *Geobacillus* have resulted in strains with increased activity towards mainly amorphous cellulose, but the engineering of a true cellulolytic strain has not been achieved yet (H Suzuki et al. 2013; Bartosiak-Jentys et al. 2013). Previous efforts in heterologous expression of Cel5A, an endoglucanase from *Thermotoga maritima*, resulted in a *P. thermoglucosidans* strain expressing active endoglucanases, although at very low yields as no protein could be visualized by SDS-PAGE (Bartosiak-Jentys et al. 2013). The expression of the catalytic domain of a GH48 exoglucanase in the same strain yielded only marginal activity. Another attempt was performed with *G. kaustophilus* HTA426, in which the expression of a heterologous endoglucanase (WP_010885255.1) from *Pyrococcus horikoshii* rendered this strain capable of CMC degradation (H Suzuki et al. 2013). Although both studies describe the successful introduction of endoglucanases, they do not report the production of fermentation products from cellulose. One of the rare cases in which direct conversion of cellulose to product was achieved was reported for *Bacillus subtilis* (Zhang et al. 2011). Zhang et al. overexpressed the native GH5 endoglucanase of *B. subtilis* enabling it to degrade CMC. Further improvements were obtained by directed evolution of the signal peptide fused to the endoglucanase, resulting in a strain capable of not only growing on CMC but also capable of producing lactic acid directly from it. However, the heterologous expression in *B. subtilis* of both a GH5 and GH9 from *C. phytofermentans* did not result in growth on CMC (Zhang et al. 2011).

In our experiments, we demonstrated the degradation of CMC by T12 strains expressing (heterologous) cellulases. However, in accordance with previous cellulase expression in *Geobacillus* and *Bacillus* spp., no organic acid production from cellulose was detected (Chapter 5). Equal production (rates) of organic acids were seen in fermentations using cellobiose and glucose (data not shown), excluding the cellobiose transporter as limiting factor in the conversion of CMC to lactic acid. As overproduction of cellulases could not be verified by SDS-PAGE, it is more likely that functional expression and consequently cellulolytic activity is limiting fermentative growth on cellulose. Common causes for low heterologous enzyme activity are low gene expression, hampered protein production and secretory problems in the case of secreted proteins (Gupta & Shukla 2016). Levels of gene expression are directly linked to promoter activity and the effect of different promoters for heterologous protein production in *Geobacillus* spp. has only been addressed in a few studies (Bartosiak-Jentys et al. 2013; H Suzuki et al. 2013; Daas et al. 2016; Cripps et al. 2009; Lin et al. 2014). The *G. stearothermophilus* lactate dehydrogenase promoter (P_{ldh}) was long assumed the strongest promoter, and has been used for the expression of proteins in *P. thermoglucosidans* (Bartosiak-Jentys et al. 2012). However, when using P_{ldh} its expression levels are influenced by cultivation conditions and corresponding redox potentials (Reeve et al. 2016), which makes this promoter not universally useful. Protein expression controlled by the promoter of the cellobiose-specific PTS system operon ($P_{\beta glu}$, which is inducible by cellobiose, showed lower activity in comparison to the *ldh* promoter, and expression was less strong in media containing glucose or xylose. Another promising promoter described is the P_{up2n38} promoter, derived from the upstream region of the uracil phosphoribosyltransferase encoding gene of *P. thermoglucosidans* NCIMB 11955. The activity of a heterologous endoglucanase expressed in *P. thermoglucosidans* was five times higher in comparison to the $P_{\beta glu}$ and was shown to be constitutively expressed. In our study the P_{uppT12} promoter, the T12 counterpart of P_{up2n38}, gave promising results in expression of the *pheB* reporter gene (Chapter 2) and was therefore used for the expression

of various cellulases (Chapter 5). Recently, an expansion of the *(Para) geobacillus* genetic toolbox was presented that makes use of the P_{rplS} promoter (Reeve et al. 2016). This promoter has been derived from the *P. thermoglucosidans* genome where it acts as a promoter of the ribosomal protein RplS. P_{rplS} was demonstrated to exhibit strong constitutive expression under various growth conditions in other *Bacilli* and is active in both *E. coli* and *P. thermoglucosidans*. The *rplS* promoter is approximately two times as strong as P_{ldh} and is considered the strongest *Geobacillus* promoter to date (Reeve et al. 2016).

Protein secretion in bacteria proceeds primarily via the Sec pathway (Kang et al. 2014). The secretion signal peptide (SP) of various proteins of *Geobacillus* spp. have been analyzed and predicted to have a Sec dependent signal peptide (Bartosiak-Jentys et al. 2013; H Suzuki et al. 2013). In our study, we used mainly the (Sec dependent) signal peptide originating from the extracellular endoxylanase *gt*XynA (GTHT_01347; GenBank: KX962565.1) as this signal peptide was shown to be useful for hydrolytic protein secretion (Daas et al. 2016; Bartosiak-Jentys et al. 2013; H Suzuki et al. 2013). Recently, a library of 173 genome-derived signal peptides of *B. subtilis* were screened for their influence on the heterologous expression of cellulases from *C. thermocellum* (Thanh Bien et al. 2014). It was found that these signal peptides caused negligible differences in the amounts of enzyme secreted and in their activity towards cellulose. In addition, previous research suggests there is no uniformly applicable signal peptide for protein expression (Brockmeier et al. 2006). It is still not possible to predict the effect of different signal peptides on the level of secretion, and more secretome analysis are needed to generate insight in the crucial sequence elements and compatibility of the SPs with the mature proteins (Anné et al. 2014; Delic et al. 2014). We demonstrated activity of both CelK, a *C. thermocellum* derived exoglucanase, and CelA, an endoglucanase from *Geobacillus* sp. 70PC53, which both were secreted using the Sec pathway. When analyzing the intracellular protein fraction by SDS-PAGE, we were unable to detect heterologous protein suggesting low protein expression levels. However, it does show that the Sec pathway for protein secretion was suitable for

these enzymes. Although the cause of the lack of activity of several other glucanases in this study remains elusive, it most likely reflects hampered translation, protein misfolding, proteolytic degradation or a combination thereof.

Towards CBP

The exploitation of *Geobacillus* for consolidated bioprocessing is still in its infancy, but several studies report the direct conversion of polysaccharides, especially starch, to valuable fermentation products. Using *G. stearothermophilus*, raw potato starch was directly converted to lactic acid at a rate of 1.8 g/L with an optical purity of 98% (Smerilli et al. 2015). Fermentation by *P. thermoglucosidans* of palm kernel cake (PKC) hydrolysate, rich in mannan-derived oligosaccharides, resulted in an ethanol yield of 0.47 g/g (92% of the theoretical maximum), thereby showing its superiority to *S. cerevisiae* (0.43 g/g ethanol yield) on purely monomeric sugars derived from PKC. The requirement of *S. cerevisiae* for additional β-mannosidase, to reduce remaining oligosaccharides to fermentable monosaccharides, further strengthens the added benefit of using *P. thermoglucosidans* in the fermentation of polysaccharide containing substrates. Although these efforts certainly show the potential of (*Para*)*geobacillus*, the fermentation of lignocellulosic substrates requires a more complex array of enzymatic activities. The heterologous expression of cellulases was partially successful with both endoglucanases and an exoglucanase successfully expressed in *G. thermodenitrificans* T12. In theory, we complemented the cellulolytic machinery of T12, although a co-culture of strains expressing endoglucanases and exoglucanases did not result in the production of organic acids from cellulose.

Strain T12 harbors many of the features that are required as a CBP organism; however, it is still limited in its use due to several factors. The degradation of hemicellulose is a key feature, but is limited to poor media without glucose. The presence of glucose in a CBP process is inevitable as the expression of heterologous cellulases will result in the generation of cellobiose and glucose, leading to negative feedback on the expression of xylanases through the catabolite

repressor protein CcpA. It was reported for *P. thermoglucosidans* T-6 that the expression of the extracellular xylanase Xyn10A was reduced by 70% when grown in the presence of glucose (Shulami et al. 2014). Furthermore, a XylR mutant (M-7), showed slightly higher xylanase expression upon the addition of glucose, proving XylR to be involved in catabolite repression. This implies that the influence of both CcpA and XylR do affect xylanase expression and should therefore both be targeted to overcome catabolite repression. We demonstrated the direct conversion of beechwood xylan to lactic acid in poor media using strain T12 (Chapter 2), and we could not detect any potential binding sites for CcpA in the nucleotide sequence of the *gtxynA* promoter region. However, we did not determine the effect of glucose or other sugars in relation to xylanase expression.

Concluding remarks

This study shows the potential of *G. thermodenitrificans* as a host for consolidated bioprocessing. The hemicellulolytic activity has been demonstrated, but needs further optimization to utilize a wider range of substrates and specially to disrupt the regulatory systems that inhibit enzymatic activity under certain conditions. The introduction of cellulolytic activity has so far been unsuccessful, however, with the rapidly expanding genetic toolbox of thermophilic bacilli, the successful exploitation of cellulose comes within reach. The discovery of a thermostable CRISPR/Cas Type II system, in *G. thermodenitrificans* T12, has proven strain T12 to be an asset in the further development of genetic tools for *Geobacillus* and thermophiles in general. The discovery of novel traits, and especially the insights gained in the saccharolytic capacity of *Geobacillus*, provides opportunities to assemble desired sets of hydrolytic clusters required for the degradation of a substrate of choice. In this way, a genetically accessible *Geobacillus* strain could be used as a chassis organism to develop customized CBP strains.

Table S1 The partial HUS locus of *G. thermodenitrificans* T81

Gene name	Product size (aa)	Subcellular Localization	Predicted protein function	Closest non-*Geobacillus* ortholog	Strain	AA id (%)	Bitscore	E-value
araR	365	Cytoplasm	Arabinose metabolism transcriptional repressor	*Anoxybacillus geothermalis*	-	91	1791	0.00E+00
araD	229	Cytoplasm	L-ribulose-5-phosphate 4-epimerase	*Anoxybacillus geothermalis*	-	93	1141	2.00E-63
araB	565	Cytoplasm	L-ribulokinase	*Anoxybacillus flavithermus*	-	96	2837	0.00E+0
araA	497	Cytoplasm	L-arabinose isomerase	*Anoxybacillus flavithermus*	-	96	2559	0.00E+00
abp	434	Cytoplasm	Alpha-galactosidase	*Bacillus litoralis*	-	75	1741	0.00E+00
abnB	316	Membrane	Intracellular endo-alpha-L-arabinanase	*Bacillus sp.*	TS-3	99	1666	0.00E+00
abnE	454	Membrane	Arabino-oligosaccharides-binding protein	*Domibacillus indicus*	-	72	1778	0.00E+00
abnF	301	Membrane	Sugar ABC transporter permease	*Bacillus sp.*	MRMR6	79	1257	9.00E-173
abnJ	294	Membrane	Sugar ABC transporter permease	*Mycobacterium abscessus*	-	76	1183	1.00E-161
abnA	852	Extracellular	Extracellular arabinase	*Bacillus sp.*	UNC41MFS5	72	3339	0.00E+00
abfB	506	Cytoplasm	Alpha-L-arabinofuranosidase	*Paenibacillus ihumii*	-	81	2263	0.00E+00
Hypothetical_1	89	-	Hypothetical protein	*Bacillus niaini*	-	48	230	5.00E-23
abfA	503	Cytoplasm	Alpha-L-arabinofuranosidase	*Anoxybacillus geothermalis*	-	96	2607	0.00E+00
araJ (partial)	237	Cytoplasm	NAD-dependent oxidoreductase	*Anoxybacillus geothermalis*	-	87	1125	3.00E-153
araJ (partial) (*mviM* annotation)	93	Cytoplasm	NAD-dependent oxidoreductase	*Anoxybacillus geothermalis*	-	82	390	2.00E-44
xylM	351	Cytoplasm	Aldose 1-epimerase (mutarotase)	*Anoxybacillus geothermalis*	-	87	1600	0.00E+00
araL	274	Cytoplasm	Haloacid dehydrogenase	*Anoybacillus geothermalis*	-	83	1217	2.00E-167
araM	405	Cytoplasm	Glycerol-1-phosphate dehydrogenase	*Anoybacillus geothermalis*	-	83	1819	0.00E+00
araN	641	Cytoplasm	Glycosyl hydrolase	*Parageobacillus caldoxylosilyticus*	-	80	2852	0.00E+00

araJ (partial)	73	Cytoplasm	NAD-dependent oxidoreductase	*Anoxybacillus geothermalis*	-	89	344	7.00E-36
GE40	542	Extracellular	Endoglucanase	*Bacillus cellulosilyticus*	DSM 2522	58	1691	0.00E+00
yesL	214	Cytoplasm	Hypothetical protein	*Parageobacillus thermogluosidans*	C56-YS93	70	808	4.00E-107
bgaA	679	Cytoplasm	Beta-galactosidase	*Bacillus hemicellulosilyticus*	JCM 9152	69	2624	000E+00
ypdA	602	Cytoplasm	Sensor histidine kinase	*Bacillus sp.*	SA1-12	53	1672	000E+00
Hypothetical_2	511	Cytoplasm	Response regulator	*Bacillus sp.*	SA1-12	43	1061	2.00E-136
yteP	333	Membrane	Sugar ABC transporter permease	*Bacillus okuhidensis*	-	77	1288	1.00E-176
ugpE	305	Membrane	Sugar ABC transporter permease	*Bacillus okuhidensis*	-	78	1244	1.00E-170
Tnp_1	334	Cytoplasm	Transposase	*Anoxybacillus sp.*	SK3-4	98	1806	0.00E+00
ugpB	559	Membrane	ABC transporter protein	*Bacillus sp.*	SA1-12	78	2374	0.00E+00
yicI	813	Cytoplasm	Alpha-xylosidase	*Bacillus lentus*	-	68	2922	0.00E+00
bglA	448	Cytoplasm	Beta-glucosidase A	*Bacillus bogoriensis*	-	69	1748	0.00E+00
xynB4	933	Cytoplasm	Beta-xylosidase	*Bacillus psychrosaccharolyticus*	-	62	3058	0.00E+00
afcA	779	Cytoplasm	Alpha-fucosidase	*Bacillus sp.*	J37	65	2788	0.00E+00
Tnp_2 (partial)	46	Cytoplasm	Transposase	*Bacillus smithii*	7_3_47FAA	91	209	1.00E-20
Tnp_3 (partial)	244	Cytoplasm	Transposase	*Bacillus smithii*	7_3_47FAA	71	907	1.00E-20
Tnp_4	172	Cytoplasm	Transposase	*Anoxybacillus suryakundensis*	-	86	820	3.00E-110
Tnp_5	176	Cytoplasm	Transposase	*Anoxybacillus suryakundensis*	-	91	855	2.00E-115
Axe2	221	Cytoplasm	Acetyl xylan esterase	*Parageobacillus thermoglucosidans*	C56-YS93	100	1170	4.00E-162
GE39	493	Cytoplasm	Beta-xylosidase	*Parageobacillus thermantarticus*		99	2657	0.00E+00
xylA	442	Cytoplasm	Xylose isomerase	*Parageobacillus thermoglucosidans*	-	99	2353	0.00E+00
xylB		Cytoplasm	Xylose kinase	*Parageobacillus thermantarticus*	-	99	2647	0.00E+00

Table S2 Annotations of the genes located in the HUS locus of *G. thermodenitrificans* T12. The gene locus tag, gene names and their respective cluster have been indicated.

Locus tag	name		Cluster
GTHT_01340	*npd*		
GTHT_01341	*xylB*	Xylulose kinase	**A**
GTHT_01342	*xylA*	Xylose isomerase	
GTHT_01343	*xynB3*	Xylan beta-1,4-xylosidase	**B**
GTHT_01344	*tnp*	Transposase	
GTHT_01345	*axe2*	Acetyl xylan esterase	
GTHT_01346	*xynX*	NGG1-interacting factor 3	
GTHT_01347	*xynA*	Endo-1,4-beta-xylanase (extracellular)	**C**
GTHT_01348	*orfA*	Integral membrane protein	
GTHT_01349	*uxuB*	Fructuronate reductase	**D**
GTHT_01350	*uxuA*	Mannonate dehydratase	
GTHT_01351	*uxaC*	Uronate isomerase	
GTHT_01352	*uxuR*	GntR family transcriptional regulator	
GTHT_01353	*kdgA*	2-dehydro-3-deoxy-phosphogluconate aldolase	
GTHT_01354	*kdgK*	5-dehydro-2-deoxygluconokinase	
GTHT_01355	*xynB1*	Xylan 1,4-beta-xylosidase	**E**
GTHT_01356	*aguA*	Xylan alpha-(1->2)-glucuronosidase	
GTHT_01357	*aguG*	Putative ABC transporter permease protein YesQ	**F**
GTHT_01358	*aguF*	Putative ABC transporter permease protein YurN	
GTHT_01359	*aguE*	Aldotetraouronic acid-binding protein	
GTHT_01360	*xynA2*	Endo-1,4-β-xylanase (intracellular)	**G**
GTHT_01362	*xynB2*	Xylan 1,4-β-xylosidase	
GTHT_01363	*axe1*	Acetyl xylan esterase	
GTHT_01364	*xylM*	Aldose 1-epimerase	
GTHT_01365	*xynG*	Xylo-oligosaccharide transport system permease protein	
GTHT_01366	*xynF*	Xylo-oligosaccharide transport system permease protein	**H**
GTHT_01367	*xynE*	Xylo-oligosaccharide-binding protein	
GTHT_01368	*xynC*	Putative response regulatory protein	
GTHT_01369	*XynD*	Putative sensor-like histidine kinase	
GTHT_01370	*ox*	Oxidoreductase like	
GTHT_01371	*tnp*	Transposase	
GTHT_01372	*tnp*	Transposase	
GTHT_01373	*araM*	Glycerol-1-phosphate dehydrogenase [NAD(P)+]	**I***
GTHT_01374	*araL*	Haloacid dehydrogenase	
GTHT_01375	*araJ*	NAD-dependent oxidoreductase	
GTHT_01376	*abfA*	Intracellular Alpha-L-arabinofuranosidase	**J**
GTHT_01377	*cutC*	L-arabinose ABC transporter permease	**K**
GTHT_01378	*cutB*	L-arabinose ABC transporter permease	
GTHT_01379	*cutA*	L-arabinose transporter (sugar-binding protein)	

Table S2 continued.

Locus tag	name		Cluster
GTHT_01380	*araA*	L-arabinose isomerase	L
GTHT_01381	*araB*	Ribulokinase	
GTHT_01382	*araD*	L-ribulose-5-phosphate 4-epimerase UlaF	
GTHT_01383	*araR*	HTH-type transcriptional repressor PurR	
GTHT_01384	*araH*	Xylose transport system permease protein XylH	M
GTHT_01385	*araG*	Xylose import ATP-binding protein XylG	
GTHT_01386	*araE*	D-xylose-binding periplasmic protein	
GTHT_01387	*araT*	Two component transcriptional regulator, AraC family	
GTHT_01388	*araS*	Two-component sensor histidine kinase	
GTHT_01389	*araP*	Sugar ABC transporter sugar-binding protein	
GTHT_01390	*tnp*	transposase	
GTHT_01391	*fbaA*	putative fructose-bisphosphate aldolase	
GTHT_01392	*iolA*	Methylmalonate semialdehyde dehydrogenase [acylating] 2	
GTHT_01393	*iolC*	5-dehydro-2-deoxygluconokinase	
GTHT_01394	*iolB*	5-deoxy-glucuronate isomerase	
GTHT_01395	*iolE*	Inososе dehydratase	N
GTHT_01396	*iolD*	3D-(3,5/4)-trihydroxycyclohexane-1,2-dione hydrolase	
GTHT_01397	*iolG*	Inositol 2-dehydrogenase	
GTHT_01398	*rbsR*	HTH-type transcriptional repressor PurR	
GTHT_01399	*rbsB*	D-ribose-binding periplasmic protein	
GTHT_01400	*rbsC*	Ribose transport system permease protein RbsC	
GTHT_01401	*rbsA_1*	Ribose import ATP-binding protein RbsA	
GTHT_01402	*tnp*	transposase	
GTHT_01403	*rbsA_2*	Ribose import ATP-binding protein RbsA	
GTHT_01404	*mviM*	putative oxidoreductase ORF334	
GTHT_01405	*mivM*	Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase	
GTHT_01406	*pcp*	Pyrrolidone-carboxylate peptidase	O
GTHT_01407	-	-	
GTHT_01408	*oppA_2*	Periplasmic oligopeptide-binding protein	
GTHT_01409	*appF*	Oligopeptide transport ATP-binding protein OppF	
GTHT_01410	*dppD_2*	Dipeptide transport ATP-binding protein DppD	
GTHT_01411	*oppC*	Oligopeptide transport system permease protein OppC	
GTHT_01412	*oppB*	Oligopeptide transport system permease protein OppB	
GTHT_01413	-	Peptidase	
GTHT_01414	-	HTH-type transcriptional regulator	
GTHT_01415	*hutI_2*	Imidazolonepropionase	

Table S2 continued.

Locus tag	name		Cluster
GTHT_01416	*kduI1*	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase	
GTHT_01417	*kduD*	2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase	
GTHT_01418	*agaR_1*	Putative aga operon transcriptional repressor	
GTHT_01419	-	putative response regulatory protein	**P**
GTHT_01420	*ypdA_1*	Sensor histidine kinase YpdA	
GTHT_01421	*pel*	Pectate lyase	
GTHT_01422	*yesR*	Unsaturated rhamnogalacturonyl hydrolase YesR	
GTHT_01423	-	Putative ABC transporter periplasmic-binding protein YcjN	
GTHT_01424	-	Putative ABC transporter permease protein ORF2	
GTHT_01425	-	Trehalose/maltose transport system permease protein MalF	
GTHT_01426	*kdgK_3*	Fructokinase-1	
GTHT_01427	*eda_2*	2-dehydro-3-deoxy-phosphogluconate aldolase	
GTHT_01428	*uxaA_1*	D-galactarate dehydratase	
GTHT_01429	*uxaB*	Altronate oxidoreductase	
GTHT_01430	*rbsR_2*	HTH-type transcriptional repressor PurR	
GTHT_01431	*uxaC_2*	Uronate isomerase	
GTHT_01432	*gfo*	Inositol 2-dehydrogenase 3	
GTHT_01433	-	-	
GTHT_01434	-	-	
GTHT_01435	*piv_3*	Pilin gene-inverting protein	
GTHT_01436	*gltT_1*	C_4-dicarboxylate transport protein 2	
GTHT_01437	*Man2c1*	Alpha-mannosidase 2C1	
GTHT_01438	-	Putative arabinogalactan oligomer transport system permease protein GanQ	
GTHT_01439	*lacF_2*	Lactose transport system permease protein LacF	
GTHT_01440	*cycB*	Cyclodextrin-binding protein	
GTHT_01441	*araR_2*	HTH-type transcriptional repressor PurR	
GTHT_01442	*echD*		

References

Abdel-Rahman, M.A. & Sonomoto, K., 2016. Opportunities to overcome the current limitations and challenges for efficient microbial production of optically pure lactic acid. *Journal of Biotechnology*, 236, pp.176–192.

Abdel-Rahman, M.A., Tashiro, Y. & Sonomoto, K., 2010. Lactic acid production from lignocellulose-derived sugars using lactic acid bacteria: Overview and limits. *Journal of Biotechnology*, 156(4), pp.286–301.

Adiguzel, A., Ozkan, H., Baris, O., Inan, K., Gulluce, M., Sahin, F., 2009. Identification and characterization of thermophilic bacteria isolated from hot springs in Turkey. *J. Microbiol. Meth*, 79, pp.321–328.

Akinosho, H. et al., 2014. The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Frontiers in chemistry*, 2(August), p.66.

Aliyu, H. et al., 2016. Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. *Systematic and Applied Microbiology*, 39(8), pp.527–533.

Anand, A., Kumar, V. & Satyanarayana, T., 2013. Characteristics of thermostable endoxylanase and β-xylosidase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1 and its applicability in generating xylooligosaccharides and xylose from agro-residues. *Extremophiles*, 17(3), pp.357–366.

Anné, J. et al., 2014. Protein secretion biotechnology in Gram-positive bacteria with special emphasis on Streptomyces lividans. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8), pp.1750–1761.

Arantes, V. & Saddler, J.N., 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for biofuels*, 3(4), pp.1–11.

Argyros, D.A. et al., 2011. High ethanol Titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Applied and Environmental Microbiology*, 77(23), pp.8288–8294.

Assareh, R. et al., 2012. Characterization of the newly isolated *Geobacillus* sp. T1, the efficient cellulase-producer on untreated barley and wheat straws. *Bioresource Technology*, 120, pp.99–105.

Ban, X. et al., 2016. Expression and biochemical characterization of a thermostable branching enzyme from *Geobacillus thermoglucosidans*. *Journal of Molecular Microbiology and Biotechnology*, 26(5), pp.303–311.

Banat, I.M., Marchant, R. & Rahman, T.J., 2004. *Geobacillus debilis* sp. nov., a novel obligately thermophilic bacterium isolated from a cool soil environment, and reassignment of *Bacillus pallidus* to *Geobacillus pallidus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 54(6), pp.2197–2201.

Bartosiak-Jentys, J. et al., 2013. Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus* thermoglucosidasius. *Microbiology (United Kingdom)*, 159(PART7), pp.1267–1275.

Bartosiak-Jentys, J., Eley, K. & Leak, D.J., 2012. Application of pheB as a reporter gene for *Geobacillus* spp., enabling qualitative colony screening and quantitative analysis of promoter strength. *Applied and Environmental Microbiology*, 78(16), pp.5945–5947.

Beaugrand, J. et al., 2004. Impact and efficiency of GH10 and GH11 thermostable endoxylanases on wheat bran and alkali-extractable arabinoxylans. *Carbohydrate Research*, 339(15), pp.2529–2540.

Beauprez, J.J., De Mey, M. & Soetaert, W.K., 2010. Microbial succinic acid production: Natural versus metabolic engineered producers. *Process Biochemistry*, 45(7), pp.1103–1114.

Becker, J. et al., 2015. Top value platform chemicals: Bio-based production of organic acids. *Current Opinion in Biotechnology*, 36, pp.168–175.

Berlin, A., 2013. No barriers to cellulose breakdown. *Science*, 342(6165), pp.1454–6.

Bhalla, A. et al., 2014. Novel thermostable endo-xylanase cloned and expressed from bacterium *Geobacillus* sp. WSUCF1. *Bioresource Technology*, 165(C), pp.314–318.

Biely, P., Singh, S. & Puchart, V., 2016. Towards enzymatic breakdown of complex plant xylan structures: State of the art. *Biotechnology Advances*, 34(7), pp.1260–1274.

Blanchard, K., Robic, S. & Matsumura, I., 2014. Transformable facultative thermophile *Geobacillus stearothermophilus* NUB3621 as a host strain for metabolic engineering. *Applied Microbiology and Biotechnology*, 98(15), pp.6715–6723.

Blumer-Schuette, S.E. et al., 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. - PubMed - NCBI. *Current Opinion in Biotechnology*, 19, pp.210–217.

Blumer-Schuette, S.E. et al., 2014. Thermophilic lignocellulose deconstruction. *FEMS Microbiology Reviews*, 38(3), pp.393–448.

Boetzer, M. et al., 2012. Toward almost closed genomes with GapFiller. *Genome Biology*, 13(6), p.R56.

Boetzer, M. & Pirovano, W., 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC bioinformatics*, 15(1), p.211.

Bosma, E.F. et al., 2015. Isolation and screening of thermophilic bacilli from compost for electrotransformation and fermentation: Characterization of *Bacillus smithii* ET 138 as a new biocatalyst. *Applied and Environmental Microbiology*, 81(5), pp.1874–1883.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1–2), pp.248–254.

Breccia, J.D. et al., 1998. Purification and characterization of a thermostable xylanase from *Bacillus amyloliquefaciens*. *Enzyme*, 229(97), pp.42–49.

Brock, W.H., 1995. Justus von Liebig. Gatekeeper of chemistry. *Chemical Society Reviews*, 24(6).

Brockmeier, U. et al., 2006. Systematic Screening of All Signal Peptides from *Bacillus subtilis*: A Powerful Strategy in Optimizing Heterologous Protein Secretion in Gram-positive Bacteria. *Journal of Molecular Biology*, 362(3), pp.393–402.

Brumm, P.J. et al., 2015. Genomic analysis of six new *Geobacillus* strains reveals highly conserved carbohydrate degradation architectures and strategies. *Frontiers in Microbiology*, 6(MAY).

Canakci, S. et al., 2012. Cloning, purification and characterization of an alkali-stable endoxylanase from thermophilic *Geobacillus* sp. 71. *World Journal of Microbiology and Biotechnology*, 28(5), pp.1981–1988.

Cartmell, A. et al., 2011. The structure and function of an arabinan-specific α-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. *Journal of Biological Chemistry*, 286(17), pp.15483–15495.

Cha, M. et al., 2013. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. *Biotechnology for biofuels*, 6(1), p.85.

Chaisson, M.J. & Tesler, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC bioinformatics*, 13, p.238.

Chamkha, M., Mnif, S. & Sayadi, S., 2008. Isolation of a thermophilic and halophilic tyrosol-degrading *Geobacillus* from a Tunisian high-temperature oil field. *FEMS Microbiology Letters*, 283(1), pp.23–29.

Chen, Z. et al., 2012. Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *Journal of Biological Chemistry*, 287(30), pp.25335–25343.

Chen, Z. & Wan, C., 2017. Biological valorization strategies for converting lignin into fuels and chemicals. *Renewable and Sustainable Energy Reviews*, 73, pp.610–621.

Chikhi, R. & Medvedev, P., 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, 30(1), pp.31–37.

Chung, D. et al., 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), pp.8931–6.

Chylinski, K. et al., 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research*, 42(10), pp.6091–6105.

Couñago, R. & Shamoo, Y., 2005. Gene replacement of adenylate kinase in the gram-positive thermophile *Geobacillus stearothermophilus* disrupts adenine nucleotide homeostasis and reduces cell viability. *Extremophiles*, 9(2), pp.135–144.

Cripps, R.E. et al., 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering*, 11, pp.398–408.

Currie, D.H. et al., 2013. Functional heterologous expression of an engineered full length CipA from *Clostridium thermocellum* in *Thermoanaerobacterium saccharolyticum*. *Biotechnology for biofuels*, 6(1), p.32.

Daas, M.J.A. et al., 2016. Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost. *Biotechnology for Biofuels*, 9(1), p.210.

Davis, R. et al., 2013. *Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons*, NREL.

DeFlaun, M.F. et al., 2007. Isolation and characterization of a *Geobacillus thermoleovorans* species from an ultra-deep South African gold mine. *Systematic and Applied Microbiology*, 30, pp.152–162.

Delic, M. et al., 2014. Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxidants & redox signaling*, 21(3), pp.414–37.

Desai, S.G., Guerinot, M.L. & Lynd, L.R., 2004. Cloning of L-lactate dehydrogenase and elimination of lactic acid production via gene knockout in *Thermoanaerobacterium saccharolyticum* JW/SL-YS485. *Applied Microbiology and Biotechnology*, 65(5), pp.600–605.

Dokuzparmak, E. et al., 2016. Purification and Characterization of a Novel Thermostable Phytase from the Thermophilic *Geobacillus* sp. TF16. *International Journal of Food Properties*, pp.1–13.

Van Dyk, J.S. & Pletschke, B.I., 2012. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. *Biotechnology Advances*, 30(6), pp.1458–1480.

Ebringerova, A., Hromadkova, Z. & Heinze, T., 2005. Hemicellulose. *Advances in Polymer Science*, 186, pp.1–67.

Eddy, S.R., 2011. Accelerated profile HMM searches. *PLoS Computational Biology*, 7(10).

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), pp.1792–1797.

Elvers, D. et al., 2016. Technology Trends in Biodegradable Polymers: Evidence from Patent Analysis. *Polymer Reviews*, 56(4), pp.584–606.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *1. Evolution*, 39(4), pp.783–791.

Feng, L. et al., 2007. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), pp.5602–7.

Finn, R.D. et al., 2016. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, 44(D1), pp.D279–D285.

Fitzpatrick, M. et al., 2010. Bioresource Technology A biorefinery processing perspective : Treatment of lignocellulosic materials for the production of value-added products. *Bioresource Technology*, 101(23), pp.8915–8922.

Fong, J.C.N. et al., 2006. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles*, 10(5), pp.363–372.

Fortina, M.G. et al., 2001. Reclassification of *Saccharococcus caldoxylosilyticus* as *Geobacillus caldoxylosilyticus* (Ahmad et al. 2000) comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 51(6), pp.2063–2071.

Gao, W. et al., 2009. Characterization of thermo-stable endoinulinase from a new strain *Bacillus smithii* T7. *Applied Biochemistry and Biotechnology*, 157(3), pp.498–506.

Garvey, M. et al., 2013. Cellulases for biomass degradation: Comparing recombinant cellulase expression platforms. *Trends in Biotechnology*, 31(10), pp.581–593.

Gaurav, N. et al., 2017. Utilization of bioresources for sustainable biofuels: A Review. *Renewable and Sustainable Energy Reviews*, 73, pp.205–214.

Gerasimova, J. & Kuisiene, N., 2012. Characterization of the novel xylanase from the thermophilic *Geobacillus thermodenitrificans* JK1. *Microbiology*, 81(4), pp.418–424.

Gírio, F.M. et al., 2010. Hemicelluloses for fuel ethanol: A review. *Bioresource Technology*, 101(13), pp.4775–4800.

Gold, N.D. & Martin, V.J.J., 2007. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *Journal of Bacteriology*, 189(19), pp.6787–6795.

Goldman, G.H., 2011. *Routes to Cellulosic Ethanol*, ISBN: 978-0-387-92739-8, pp. 153-163

Van Gool, M.P. et al., 2013. Two novel GH11 endo-xylanases from *Myceliophthora thermophila* C1 act differently toward soluble and insoluble xylans. *Enzyme and Microbial Technology*, 53(1), pp.25–32.

Van Gool, M.P. et al., 2011. Screening for distinct xylan degrading enzymes in complex shake flask fermentation supernatants. *Bioresource Technology*, 102(10), pp.6039–6047.

Gupta, S.K. & Shukla, P., 2016. Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. *Critical Reviews in Biotechnology*, 36(6), pp.1089–1098.

Hasunuma, T. et al., 2013. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresource Technology*, 135, pp.513–522.

Hendriks, A.T.W.M. & Zeeman, G., 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology*, 100(1), pp.10–18.

Henrissat, B. et al., 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*, 92(15), pp.7090–7094.

Hoondal, G. et al., 2002. Microbial alkaline pectinases and their industrial applications: A review. *Applied Microbiology and Biotechnology*, 59(4–5), pp.409–418.

Huang, K. et al., 2016. Cloning, expression, and characterization of a thermostable L-arginase from *Geobacillus thermodenitrificans* NG80-2 for L-ornithine production. *Biotechnology and Applied Biochemistry*, 63(3), pp.391–397.

Hult, K. & Berglund, P., 2003. Engineered enzymes for improved organic synthesis. *Current Opinion in Biotechnology*, 14(4), pp.395–400.

Hussein, A.H., Lisowska, B.K. & Leak, D.J., 2015. The Genus *Geobacillus* and Their Biotechnological Potential. *Advances in Applied Microbiology*, 92, pp.1–48.

Huy, N.D. et al., 2015. Putative endoglucanase PcGH5 from *Phanerochaete chrysosporium* is a β-xylosidase that cleaves xylans in synergistic action with endo-xylanase. *Journal of Bioscience and Bioengineering*, 119(4), pp.416–420.

Hyatt, D. et al., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11, p.119.

Imanaka, T. et al., 1982. Transformation of *Bacillus stearothermophilus* with plasmid DNA and characterization of shuttle vector plasmids between *Bacillus stearothermophilus*. *Transformation*, 149(3).

Irfan, M. et al., 2016. Cloning, purification and characterization of a cellulase-free xylanase from *Geobacillus thermodenitrificans* AK53. *Applied Biochemistry and Microbiology*, 52(3), pp.277–286.

Jang, Y.S. et al., 2012. Bio-based production of C2-C6 platform chemicals. *Biotechnology and Bioengineering*, 109(10), pp.2437–2459.

Jinek, M. et al., 2012. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. *Science*, 337(August), pp.816–822.

Jinek, M. et al., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343(6176).

Joe Shaw, A. et al., 2008. End-product pathways in the xylose fermenting bacterium, *Thermoanaerobacterium saccharolyticum*. *Enzyme and Microbial Technology*, 42(6), pp.453–458.

St. John, F.J., Rice, J.D. & Preston, J.F., 2006. Characterization of XynC from *Bacillus subtilis* subsp. *subtilis* strain 168 and analysis of its role in depolymerization of glucuronoxylan. *Journal of Bacteriology*, 188(24), pp.8617–8626.

Jones, D.T., Taylor, W.R. & Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, 8(3).

Kananavičiūtė, R. & Čitavičius, D., 2015. Genetic engineering of *Geobacillus* spp. *Journal of Microbiological Methods*, 111, pp.31–39.

Kang, Z. et al., 2014. Molecular engineering of secretory machinery components for high-level secretion of proteins in *Bacillus* species. *Journal of Industrial Microbiology and Biotechnology*, 41(11), pp.1599–1607.

Kelley, L.A. et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), pp.845–858.

Kohli, P. & Gupta, R., 2015. Alkaline pectinases: A review. *Biocatalysis and Agricultural Biotechnology*, 4(3), pp.279–285.

Kormelink, F.J.M. et al., 1993. Mode of action of the xylan-degrading enzymes from *Aspergillus awamori* on alkali-extractable cereal arabinoxylans. *Carbohydrate Research*, 249(2), pp.355–367.

Kračun, S.K. et al., 2015. A new generation of versatile chromogenic substrates for high-throughput analysis of biomass-degrading enzymes. *Biotechnology for Biofuels*, 8(1), p.70.

Kubicek, C.P. & Kubicek, E.M., 2016. Enzymatic deconstruction of plant biomass by fungal enzymes. *Current Opinion in Chemical Biology*, 35, pp.51–57.

Kumar, S., Stecher, G. & Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7).

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), pp.680–685.

Lan Thanh Bien, T. et al., 2014. Secretion of heterologous thermostable cellulases in *Bacillus subtilis*. *The Journal of General and Applied Microbiology*, 60(5), pp.175–182.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), pp.357–359.

Lee, Y.E., Lowe, S.E. & Zeikus, J.G., 1993. Regulation and characterization of xylanolytic enzymes of *Thermoanaerobacterium saccharolyticum* B6A-RI. *Applied and Environmental Microbiology*, 59(3), pp.763–771.

Lee R. Lynd, Paul J. Weimer, Willem H. van Zyl, I.S.P., 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiology and Molecular Biology Reviews*, 66(3), pp.506–577.

Lentini, V., Gugliandolo, C. & Maugeri, T.L., 2007. Identification of enzyme-producing thermophilic bacilli isolated from marine vents of Eolian Islands (Italy). *Annals of Microbiology*, 57(September), pp.355–361.

Lever, M., 1972. A new reaction for colorimetric determination of carbohydrates. *Analytical Biochemistry*, 47(1), pp.273–279.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Li, R. et al., 2014. Variations of culturable thermophilic microbe numbers and bacterial communities during the thermophilic phase of composting. *World Journal of Microbiology and Biotechnology*, 30(6), pp.1737–1746.

Li, S. et al., 2012. Technology prospecting on enzymes: Application, marketing and engineering. *Computational and Structural Biotechnology Journal*, 2(3).

Liao, H. & Kanikula, A., 1990. Increased efficiency of transformation of*Bacillus stearothermophilus* by a plasmid carrying a thermostable kanamycin resistance marker. *Current Microbiology*, 21(5), pp.301–306.

Liberato, M. V et al., 2016. Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Scientific reports*, 6(October 2015).

Ligon, B.L., 2002. Louis Pasteur: a controversial figure in a debate on scientific ethics. *Seminars in pediatric infectious diseases*, 13(2).

Lin, P.P. et al., 2014. Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metabolic Engineering*, 24, pp.1–8.

Liu, B. et al., 2012. Characterization of a Recombinant Thermostable Xylanase from Hot Spring. *Journal of Microbiology and Biotechnology*, 22(10), pp.1388–1394.

Losordo, Z. et al., 2016. Cost competitive second-generation ethanol production from hemicellulose in a Brazilian sugarcane biorefinery. *Biofuels, Bioproducts and Biorefining*, 10(5), pp.589–602.

Lynd, L.R. et al., 2005. Consolidated bioprocessing of cellulosic biomass: An update. *Current Opinion in Biotechnology*, 16(5), pp.577–583.

De Maayer, P. et al., 2014. Comparative analysis of the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC genomics*, 15(1), p.836.

Mäki-Arvela, P. et al., 2011. Synthesis of sugars by hydrolysis of hemicelluloses- A review. *Chemical Reviews*, 111(9), pp.5638–5666.

Makky, E., 2009. Avicelase production by a thermophilic *Geobacillus stearothermophilus* isolated from soil using sugarcane bagasse. *World Academy of Science, Engineering and Technology*, 33(9), pp.487–491.

Marchant, R. et al., 2008. Thermophilic bacteria in cool temperate soils: Are they metabolically active or continually added by global atmospheric transport? *Applied Microbiology and Biotechnology*, 78(5), pp.841–852.

Marcolongo, L. et al., 2015. Properties of an alkali-thermo stable xylanase from *Geobacillus thermodenitrificans* A333 and applicability in xylooligosaccharides generation. *World Journal of Microbiology and Biotechnology*, 31(4), pp.633–648.

Mazzoli, R., 2012. Development of Microorganisms for Cellulose-Biofuel Consolidated Bioprocessings: Metabolic Engineers' Tricks. *Computational and Structural Biotechnology Journal*, 3(4), pp.1–9.

Meng, Y. et al., 2012. Bioresource Technology Efficient production of L-lactic acid with high optical purity by alkaliphilic Bacillus. *Bioresource Technology*, 116, pp.334–339.

Menon, V. & Rao, M., 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science*, 38(4), pp.522–550.

Mishima, M. et al., 2009. Cultivation characteristics of denitrification by thermophilic *Geobacillus* sp strain TDN01. *Journal of General and Applied Microbiology*, 55(2), pp.81–86.

Mitra, S. et al., 2015. Cloning, overexpression, and characterization of a novel alkali-thermostable xylanase from *Geobacillus* sp. WBI. *Journal of Basic Microbiology*, 55(4), pp.527–537.

Morange, M., 2007. How biochemistry has evolved as a discipline: From the beginnings. *Biochemist*, 29(5).

Mougiakos, I. et al., 2016. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends in Biotechnology*, 34(7), pp.575–587.

Murciano Martínez, P. et al., 2015. Importance of acid or alkali concentration on the removal of xylan and lignin for enzymatic cellulose hydrolysis. *Industrial Crops and Products*, 64, pp.88–96.

Muyzer, G. et al., 1995. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Archives of Microbiology*, 164(3), pp.165–172.

Nakayama, N. et al., 1992. A new shuttle vector for *Bacillus stearothermophilus* and Escherichia coli. *Biotechnology Letters*, 14(8), pp.649–652.

Narumi, I. et al., 1992. A newly isolated *Bacillus stearothermophilus* KI041 and its transformation by electroporation. *Biotechnology Techniques*, 6(1), pp.83–86.

Narumi, I. et al., 1993. Construction of a new shuttle vector pSTE33 and its stabilities in *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Escherichia coli*. *Biotechnology Letters*, 15(8), pp.815–820.

Nazina, T.N., Tourova, T.P. & Poltaraus, A.B., 2001. Taxonomic study of aerobic thermophilic bacilli:descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. *International Journal of Systematic and Evolutionary Microbiology*, 51 (2), pp.433–446.

Ng, I.S. et al., 2009. A novel endo-glucanase from the thermophilic bacterium *Geobacillus* sp. 70PC53 with high activity and stability over a broad range of temperatures. *Extremophiles*, 13(3), pp.425–435.

Nisha, M. & Satyanarayana, T., 2016. Characteristics and applications of recombinant thermostable amylopullulanase of *Geobacillus thermoleovorans* secreted by Pichia pastoris. *Applied Microbiology and Biotechnology*, 101(6).

Okano, K. et al., 2009. Efficient production of optically pure D-lactic acid from raw corn starch by using a genetically modified L-lactate dehydrogenase gene-deficient and alpha-amylase-secreting *Lactobacillus plantarum* strain. *Applied and Environmental Microbiology*, 75(2), pp.462–467.

Olson, D.G. et al., 2012. Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology*, 23(3), pp.396–405.

van der Oost, J. et al., 2016. Thermostable Cas9 Nucleases. PCT/EP2016/062817 (filed June 12, 2015)

Ou, M.S. et al., 2009. Thermophilic *Bacillus coagulans* requires less cellulases for simultaneous saccharification and fermentation of cellulose to products than mesophilic microbial biocatalysts. *Applied Biochemistry and Biotechnology*, 155(1–3), pp.379–385.

Palazzolo, M.A. & Kurina-Sanz, M., 2016. Microbial utilization of lignin: available biotechnologies for its degradation and valorization. *World Journal of Microbiology and Biotechnology*, 32(10).

Pédelacq, J.-D. et al., 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24(1), pp.79–88.

Petersen, T.N. et al., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8(10), pp.785–6.

Pinzón-Martínez, D.L. et al., 2010. Thermophilic bacteria from Mexican thermal environments: isolation and potential applications. *Environmental technology*, 31(8–9), pp.957–966.

PlasticsEurope, 2016. *Plastics – the Facts 2016*, http://www.plasticseurope.org/documents/document/20161014113313-plastics_the_facts_2016_final_version.pdf

Pogrebnyakov, I., Jendresen, C.B. & Nielsen, A.T., 2017. Genetic toolbox for controlled expression of functional proteins in *Geobacillus* spp. *PLoS ONE*, 12(2).

Poli, A. et al., 2011. *Geobacillus galactosidasius* sp. nov., a new thermophilic galactosidase-producing bacterium isolated from compost. *Systematic and Applied Microbiology*, 34(6), pp.419–423.

Rastogi, G. et al., 2010. Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresource Technology*, 101(22), pp.8798–8806.

Rastogi, G. et al., 2009. Isolation and characterization of cellulose-degrading bacteria from the deep subsurface of the Homestake gold mine, Lead, South Dakota, USA. *Journal of Industrial Microbiology and Biotechnology*, 36(4), pp.585–598.

Ravindran, R. & Jaiswal, A.K., 2016. A comprehensive review on pre-treatment strategy for lignocellulosic food industry waste: Challenges and opportunities. *Bioresource Technology*, 199, pp.92–102.

Reeve, B. et al., 2016. The *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synthetic Biology*, 5(12), pp.1342–1347.

Roberts, R.J. et al., 2015. REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, 43(D1), pp.D298–D299.

De Rossi, E. et al., 1994. New shuttle vector for cloning in *Bacillus stearothermophilus*. *Research in Microbiology*, 145(8), pp.579–583.

Ryckeboer, J. et al., 2003. Microbiological aspects of biowaste during composting in a monitored compost bin. *Journal of Applied Microbiology*, 94(1), pp.127–137.

Rydzak, T. et al., 2012. Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. *BMC microbiology*, 12(1), p.214.

Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec.Biol.Evol.*, 4, pp.406–425.

Schuster, B.G. & Chinn, M.S., 2013. Consolidated Bioprocessing of Lignocellulosic Feedstocks for Ethanol Fuel Production. *Bioenergy Research*, 6(2), pp.416–435.

Shallom, D. & Shoham, Y., 2003. Microbial hemicellulases. *Current Opinion in Microbiology*, 6(3), pp.219–228.

Shaw, A.J., Hogsett, D.A. & Lynd, L.R., 2009. Identification of the [FeFe]-hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium saccharolyticum* and demonstration of increased ethanol yield via hydrogenase knockout. *Journal of Bacteriology*, 191(20), pp.6457–6464.

Sheng, L. et al., 2017. Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnology for Biofuels*, 10(1), p.5.

Shoham, Y. et al., 1992. Delignification of wood pulp by a thermostable xylanase from *Bacillus stearothermophilus* strain T-6. *Biodegradation*, 3(2–3), pp.207–218.

Shulami, S. et al., 2007. A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Applied and Environmental Microbiology*, 73(3), pp.874–884.

Shulami, S. et al., 2014. Multiple regulatory mechanisms control the expression of the *Geobacillus stearothermophilus* gene for extracellular xylanase. *Journal of Biological Chemistry*, 289(37), pp.25957–25975.

Shulami, S. et al., 2011. The L-arabinan utilization system of *Geobacillus stearothermophilus*. *Journal of Bacteriology*, 193(11), pp.2838–2850.

Shulami, S., Gat, O. & Sonenshein, A.L., 1999. The Glucuronic Acid Utilization Gene Cluster from *Bacillus stearothermophilus* T-6. *Journal of Bacteriology*, 181(12), pp.3695–3704.

Sindhu, R., Binod, P. & Pandey, A., 2016. Biological pretreatment of lignocellulosic biomass - An overview. *Bioresource Technology*, 199, pp.76–82.

Singh, V., Braddick, D. & Dhar, P.K., 2017. Exploring the potential of genome editing CRISPR-Cas9 technology. *Gene*, 599, pp.1–18.

Sizova, M. V. et al., 2011. Cellulose-and xylan-degrading thermophilic anaerobic bacteria from biocompost. *Applied and Environmental Microbiology*, 77(7), pp.2282–2291.

Smerilli, M. et al., 2015. Direct fermentation of potato starch and potato residues to lactic acid by *Geobacillus stearothermophilus* under non-sterile conditions. *Journal of Chemical Technology and Biotechnology*, 90(4), pp.648–657.

Sonenshein, A.L., 2005. CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria. *Current Opinion in Microbiology*, 8(2), pp.203–207.

Studholme, D.J., 2015. Some (bacilli) like it hot: Genomics of *Geobacillus* species. *Microbial Biotechnology*, 8(1), pp.40–48.

Sun, Y. & Cheng, J., 2002. Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresource Technology*, 83(1), pp.1–11.

Suzuki, H., Wada, K., et al., 2013. A Ternary Conjugation System for the Construction of DNA Libraries for *Geobacillus kaustophilus* HTA426. *Bioscience, Biotechnology, and Biochemistry*, 130492(11), pp.3–5.

Suzuki, H., Murakami, A. & Yoshida, K.I., 2012. Counterselection System for *Geobacillus kaustophilus* HTA426 through: Disruption of *pyrF* and *pyrR*. *Applied and Environmental Microbiology*, 78(20), pp.7376–7383.

Suzuki, H., Yoshid, K. ichi & Ohshima, T., 2013. Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus kaustophilus* HTA426. *Applied and Environmental Microbiology*, 79(17), pp.5151–5158.

Tai, S.K. et al., 2004. Isolation and characterization of a cellulolytic *Geobacillus thermoleovorans* T_4 strain from sugar refinery wastewater. *Extremophiles*, 8(5), pp.345–349.

Takaku, H. et al., 2006. Microbial communities in the garbage composting with rice hull as an amendment revealed by culture-dependent and -independent approaches. *Journal of Bioscience and Bioengineering*, 101(1), pp.42–50.

Takami, H. et al., 2004. Genomic characterization of thermophilic *Geobacillus* species isolated from the deepest sea mud of the Mariana Trench. *Extremophiles*, 8(5), pp.351–356.

Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.

Tang, Y.J. et al., 2009. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnology and Bioengineering*, 102(5), pp.1377–1386.

Taylor, M.P. et al., 2009. Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends in biotechnology*, 27(7), pp.398–405.

Taylor, M.P., Esteban, C.D. & Leak, D.J., 2008. Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. *Plasmid*, 60(1), pp.45–52.

Thompson, A. et al., 2008. Heterologous expression of pyruvate decarboxylase in *Geobacillus thermoglucosidasius*. *Biotechnology letters*, 30, pp.1359–1365.

Tian, W. et al., 2013. Succession of bacterial communities during composting process as detected by 16S rRNA clone libraries analysis. *International Biodeterioration and Biodegradation*, 78, pp.58–66.

Tomás-Pejó, E., Oliva, J.M. & Ballesteros, M., 2008. Realistic approach for full-scale bioethanol production from lignocellulose: A review. *Journal of Scientific and Industrial Research*, 67(11), pp.874–884.

Tsakraklides, V. et al., 2012. Carbon catabolite repression in *Thermoanaerobacterium saccharolyticum*. *Biotechnology for biofuels*, 5(1), p.85.

Turner, P., Mamo, G. & Karlsson, E.N., 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial cell factories*, 6(9).

US DOE. 2005. Genomics:GTL Roadmap, DOE/SC-0090, U.S. Department of Energy Office of Science. pp. 27

Vazana, Y. et al., 2010. Interplay between *Clostridium thermocellum* family 48 and family 9 cellulases in cellulosomal versus noncellulosomal states. *Applied and Environmental Microbiology*, 76(10), pp.3236–3243.

Venditto, I. et al., 2014. Crystallization and preliminary X-ray diffraction analysis of a trimodular endo-β-1,4-glucanase (Cel5B) from *Bacillus halodurans*. *Acta Crystallographica Section F:Structural Biology Communications*, 70, pp.1628–1630.

Venditto, I. et al., 2015. Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and β-1,3-1,4-glucans through distinct mechanisms. *Journal of Biological Chemistry*, 290(17), pp.10572–10586.

Verma, D., Anand, A. & Satyanarayana, T., 2013. Thermostable and alkalistable endoxylanase of the extremely thermophilic bacterium *Geobacillus thermodenitrificans* TSAA1: Cloning, expression, characteristics and its applicability in generating xylooligosaccharides and fermentable sugars. *Applied Biochemistry and Biotechnology*, 170(1), pp.119–130.

Walker, B.J. et al., 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 9(11).

Wang, L. et al., 2006. Isolation and characterization of a novel thermophilic *Bacillus* strain degrading long-chain n-alkanes. *Extremophiles*, 10(4), pp.347–356.

Waterhouse, A.M. et al., 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), pp.1189–1191.

Wei, H. et al., 2014. Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-seq. *Frontiers in Microbiology*, 5(APR).

Wilson, G.G. & Murray, N.E., 1991. Restriction and modification systems. *Annual review of genetics*, 25, pp.585–627.

Wu, L.J. & Welker, N.E., 1989. Protoplast transformation of *Bacillus stearothermophilus* NUB36 by plasmid DNA. *Journal of general microbiology*, 135(5), pp.1315–24.

Wyman, C. et al., 2004. Hydrolysis of Cellulose and Hemicellulose. ISBN: 978-1-4200-3082-2, Chapter 43, pp. 994-1033

Xiao, Y., Poovaiah, C. & Coleman, H.D., 2016. Expression of Glycosyl Hydrolases in Lignocellulosic Feedstock: An Alternative for Affordable Cellulosic Ethanol Production. *Bioenergy Research*, 9(4), pp.1290–1304.

Xiao, Z. et al., 2012. Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain. *Biotechnology for biofuels*, 5(1), p.88.

Xu, L., Shen, Y. & Bao, X., 2010. Progress and strategies on bioethanol production from lignocellulose by consolidated bioprocessing (CBP) using *Saccharomyces cerevisiae*. *Shengwu Gongcheng Xuebao/Chinese Journal of Biotechnology*, 26(7), pp.870–879.

Yabe, S. et al., 2009. *Thermaerobacter composti* sp. nov., a novel extremely thermophilic bacterium isolated from compost. *Journal of General and Applied Microbiology*, 55(5), pp.323–328.

Yang, X. et al., 2013. Efficient production of L-lactic acid by an engineered *Thermoanaerobacterium aotearoense* with broad substrate specificity. *Biotechnology for biofuels*, 6(1), p.124.

Yin, Y. et al., 2012. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, 40(W1), pp.W445–W451.

Zabed, H. et al., 2016. Fuel ethanol production from lignocellulosic biomass: An overview on feedstocks and technological approaches. *Renewable and Sustainable Energy Reviews*, 66, pp.751–774.

Zeigler, D.R., 2014. The *Geobacillus* paradox: Why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology (United Kingdom)*, 160(PART 1), pp.1–11.

Zeldes, B.M. et al., 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Frontiers in Microbiology*, 6(NOV).

Zhang, X.Z. et al., 2011. One-step production of lactate from cellulose as the sole carbon source without any other organic nutrient by recombinant cellulolytic *Bacillus subtilis*. *Metabolic Engineering*, 13(4), pp.364–372.

Zhang, Y.P. & Lynd, L.R., 2005. Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), pp.7321–5.

Zheng, Z. et al., 2013. Advances of consolidated bioprocessing based on recombinant strategy. *Shengwu Gongcheng Xuebao/Chinese Journal of Biotechnology*, 29(10), pp.1354–1362.

Acknowledgements

These final pages, the acknowledgements, are likely the most-read in any thesis and thereby among the most important pages written. Many people have contributed to this work and I would like to thank and acknowledge you all for your help in the completion of this thesis. It is fascinating to look back at the time I started this project, even the way towards it, as the many people I've met influenced the choices made and, along the four years of my PhD, changed me on a professional and personal level. Some people in particular have contributed to this work and my development, and to those I would like to address my personal gratitude.

Richard, thank you for your guidance throughout these four years. We first met in 2011 during my MSc thesis at the Laboratory of Microbiology and I am grateful for the confidence you expressed in me by asking me for this PhD project. Your calm and convincing manor of steering the project in the right direction has been of immense value and, in these 4+years, you have had the most influence on how I have developed as a scientist. In the last few months of the project, I fired many pages to review at you, but no matter if send on late evenings or weekends, you were always willing to review and respond on a very short notice. I really appreciated that, probably more than I ever showed. Recently, you were appointed special professor of Bacterial Cell Factories. A well-deserved title and I am sure many interesting breakthroughs will follow in the coming years. I hope our roads will cross again, and I look forward to it!

John, a few weeks after my MSc graduation, in the middle of my USA road trip, I received a message from you saying that a PhD project became available and that you and Richard wanted me to jump in. Just a week after I returned home I started on this project and enjoyed it ever since. Thank you for being such an energetic, inspiring and ever enthusiastic group leader. The Bacterial Genetics group is truly a wonderful place to do a PhD and much of that we owe to you. I

enjoyed walking into your office not only for scientific discussions, but also to discuss the latest developments in the Eredivisie and on which squad would crown themselves champions this year. Fey… feyn…. no, I'll just leave it at that.

Willem, thanks a lot for your input in the project. I enjoyed the project and PhD meetings, where you are always able to pinpoint the strengths and weaknesses in any project. The vast amount of novel ideas and research directions fired at me during these meetings was overwhelming, but helped me to succeed in my PhD. For 2.5 years, I participated in the Daily Board, where I very much enjoyed your style of leading meetings. Your role as one of my promoters has been taken over by Richard, but your role in this project has been of excellent value and I am grateful for our collaboration.

To my paranymphs, **Daan** and **Jorrit**, thank you for all the great times we had during these years. The discussions on work related matters as well as on a personal note were of irreplaceable value and I enjoyed working with you both. Besides work, our skiing adventures are truly memorable, although I worry for the GoPro shots taken of my first descents. Despite my current appearance as "Bambi-on-ice", you know it will take just a couple more slopes before I outpace you both. I appreciate that you accepted to be my paranymphs, let's hold the "tall-men-pact" for many more years.

Many thanks also to all collaborators in this project. **Slav**, you started this project in 2011 and we worked together for a few months at the start of my PhD. Thank you for all the help and support. **Serve** and **Brenda**, thank you both for your valuable input during project meetings and our joint efforts on the xylanase and *Clostridium* work. **Bastienne**, thank you for helping us out with the genome assemblies and discussions on how to tackle much of the work now presented in Chapters 3 and 6. **Bart N.**, we have extensively used your expertise on our metagenome search for novel enzymes and, presented in Chapter 5, we managed to find and characterize them. Thank you for all your efforts and contributions. Also thanks to **Peter S.** and **Jasper**

K. for the valuable input in our discussions on this matter. **Patricia**, we successfully collaborated on the biochemical characterization of the *Geobacillus* endoxylanase. I enjoyed our collaboration, and it is great to see that we managed to publish this work!

I also would like to thank the "C3-acids" partners from Corbion, DSM, TU-Delft and WUR for their valuable input during the project meetings. Special thanks to **Peter, Mirjam, Tony, Piet, Maaike, David, Ruud, Edwin** and **Patricia**. Also to **Paulien** and **Brigit**, thanks for all the effort that has been put into the pretreated bagasse samples; unfortunately, we couldn't manage to follow-up with fermentation data on these experiments.

To all colleagues that shared an office with me, thank you so much for the wonderful time. First, **Tom**, you co-supervised my thesis project and later joined me in this PhD project. It was great working with you on the project, organizing the weekly drinks at the "Beerstübe 0.022", and joining the WeDays where we even made it to the soccer finals. Thanks also for the joyful times in and outside working hours. **Elleke**, thank you for the great times and the opportunity to do my MSc thesis in your project. This thesis ultimately led to my first publication and indirectly to my own PhD project. It wasn't always easy for you to concentrate in the office with Tom and me and I have now experienced how difficult any sound can be at the last stretch of the PhD. **Alex**, thanks for giftwrapping every single item on my desk! After these four years, I still haven't unpacked it all. Good luck with your future project in collaboration with Youri, nice to see that you value the potential of *Geobacillus* too ;P **Jeroen**, **Yannis** and **Joyshree**, thank you for all the great times and discussions at work meetings, keep up the excellent work and all the best in finalizing your PhD's.

Many students contributed to the work presented in this thesis and I am grateful that these students chose my project for their thesis. **Rob**, **Sofia**, **Ioanna**, **Laurens**, **Bart**, **Leandra** and **Fons**, thank you all for your contributions. Working with you has enriched my PhD

and has led to many impressive results, in some cases even to co-authorships. I wish you all the very best in your future careers!

Also thanks to **Jasper**, thanks for the great times we had on conferences, the PhD trip and many nights out. It is an honor to be paranymph at your defense. Together with **Bas**, **Nico** and **Daan** we had a blast in Canada after the PhD trip. Also thanks to these guys for making this such an awesome experience. **Melvin**, thanks for all your help on diverse topics, I enjoyed our discussions and talks and I'm looking forward to seeing you again in the Bay Area :P **Franklin**, it was great working with you in the lab. You are the first person I met that can ruin a "rookworst", impressive stuff. **Becca**, thank you for all the great times and for proofreading Chapter 2. To all the colleagues of the Bacterial Genetics group and Microbiology in general; thank you for the friendly and professional work environment that made my PhD a wonderful time.

Wim, Sjon, Steven, Monica, Philippe, Ton, Tom and Tom, thanks to all of you for your support on all kinds of issues, HPLC's, computers, fermenters, etc. etc. You all have been of major help to me throughout these years. **Anja** and **Carolien**, thanks for all the support and for always helping me out with all kinds of questions.

Also, a big thanks to the friends that I've met during my studies and PhD. Especially **Mark** and **Jaap**, you both moved to Leeuwarden but we still occasionally meet for a beer. It's nice to discuss both work and non-work with you. Good luck to both of you in the final months of your PhD's.

Besides all the people directly linked to work, I would also like to thank some people that have, maybe unintentionally, helped me throughout my PhD. My friends, **Joop, Tom, Vis, Jeep** and **Jonk**, we have been friends for about 18 years now and it's great to see everyone achieve so much. Thanks for all the great times and distraction from my project. Special thanks to Tom for designing the cover and invitation cards.

Thanks to my parents, **Jan** & **Caroline** for being so supportive and interested. You taught us to be independent and gave us the freedom to explore our own potential. That this would lead to family members flying all over the map is an unforeseen effect, but an effect you might enjoy pretty much now retirement kicked in and you have all this time to visit us ;) Also thanks to my brother (**Pieter)** and sisters **(Suzanne** and **Kristel**) for your interest and support these years. Special thanks to Kristel also for proofreading parts of this thesis.

Thanks to my parents-in-law **Hans** and **Riet**, for being so interested in my project and for always making me feel welcome in Raalte. I enjoy our late-night discussions when drinking "one last beer".

Marjolein, these last words I will address to you, being the most important person to me for 9 years already. Thank you for everything, your interest and support, your endless patience, your humor and your wittiness. Thank you for all the things we have been through and for the adventures to come. Thank you for being you.

About the author

Martinus (Tijn) J.A. Daas was born on August 10, 1988 in Tilburg, the Netherlands. In 2005, he completed his secondary school (HAVO) at Pius X College in Bladel and started his studies on Biotechnology at the University of Applied Sciences (INHolland) in Delft. He included a minor in Fermentation Sciences in the Bachelor program and performed his thesis at the University of Wageningen, under supervision of Dr Catarina I.F. da Silva, where he studied the production of collagen-like polymers by the yeast *Pichia pastoris*. His bachelor graduation was concluded with an internship at Food & Biobased Research in Wageningen, under supervision of Dr Matthé Wagemaker, were he participated in the EU-PEARLS project by characterizing the expression of genes involved in the rubber synthetic pathway of *Taraxacum koksagiz*. After obtaining his Bachelor's degree, Tijn continued his studies in the Master's program of Biotechnology at Wageningen University, specializing in Molecular and Cellular Biotechnology. He concluded this study with a thesis at the Laboratory of Microbiology of Wageningen University. During this thesis, supervised by Dr Elleke Bosma and Eng. Tom van de Weijer, Tijn worked on the characterization of *Bacillus smitthii* and development of an efficient transformation protocol for this novel thermophile. He completed his Master's degree in 2012 with an internship at Dyadic NL (now DuPont) with research on the optimization of fermentation processes using the fungus *Myceliopthora thermophila*. After obtaining his Master's degree, Tijn started a PhD at the Laboratory of Microbiology in the Bacterial Genetics group under supervision of Prof. Dr John van der Oost and Prof. Dr Richard van Kranenburg. Most of his work on the isolation and development of *Geobacillus thermodenitrificans* for consolidated bioprocessing is presented in this thesis.

Publications

Bosma EF, van de Weijer AHP, Daas MJA, van der Oost J, de Vos WM, van Kranenburg R: **Isolation and screening of thermophilic bacilli from compost for electrotransformation and fermentation: Characterization of *Bacillus smithii* ET 138 as a new biocatalyst.** *Applied and Environmental Microbiology* 2015, **81**(5):1874–1883.

Daas MJA, van de Weijer AHP, de Vos WM, van der Oost J, van Kranenburg R: **Isolation of a genetically accessible thermophilic xylan degrading bacterium from compost**. *Biotechnology for Biofuels* 2016, **9**(1):1–13.

Daas MJA, Vriesendorp B, van de Weijer AHP, van der Oost J, van Kranenburg R: **Complete genome sequence of *Geobacillus thermodenitrificans* T12, a potential host for biotechnological applications.** *Current Microbiology* 2017, (Submitted)

Daas MJA, Martínez PM, van de Weijer AHP, van der Oost J, de Vos WM, Kabel MA, van Kranenburg R: **An extracellular endo-xylanase from the thermophile *Geobacillus thermodenitrificans* T12 is key for its xylan degradation.** *BMC Biotechnology* 2017, **17**:44.

Daas MJA, van de Weijer AHP, de Vos WM , van der Oost J, van Kranenburg R: **Engineering *Geobacillus thermodenitrificans* to introduce cellulolytic activity; expression of native and heterologous cellulase genes.** *Microbial Cell Factories* 2017, (Submitted)

Patent

Van der Oost J, Daas MJA, Kengen SWM, de Vos WM: **Thermostable Cas9 nucleases.** PCT/EP2016/062817 (Filed June 6, 2015)

Overview of completed training activities

Discipline specific activities

Meetings & Conferences

-	# BE-Basic symposium I (NL)	2012
-	# BE-Basic symposium II (NL)	2013
-	Molecular Microbiology Fall meeting (NL)	2014
-	ALW Molecular genetics meeting (NL)	2014
-	# BE-Basic symposium III (NL)	2014
-	* Conference on Metabolic Engineering (NL)	2015
-	BaCell meeting (NL)	2015
-	# BE-Basic symposium IIII (NL)	2015
-	* BioMicroWorld (E)	2015
-	# ECO-BIO (NL)	2016
-	# Symposium on Biotechnology for Fuels and Chemicals (USA)	2016
-	* Microbial Biotechnology symposium (NL)	2016

poster presentation, * oral presentation

Courses &Training

-	BE-BIC Workshop (NL)	2013
-	Microbial physiology and fermentation technology (NL)	2014
-	Training Enzyme assays FCH (NL)	2015
-	Food & Biorefinery Enzymology (NL)	2015

General courses

-	VLAG PhD week	2012
-	Competence assessment	2013
-	Techniques for writing and presenting a scientific paper	2013
-	Scientific writing	2016
-	Career perspectives CCP	2016
-	Last stretch of the PhD programme	2016

Optionals

-	Preparation of research proposal	2012
-	Bacterial Genetics group meetings	weekly
-	PhD study tour (USA & Canada)	2013
-	PhD meetings	montly
-	Microbiology Seminars	montly
-	Daily Board dept. of Microbiology	2013-2016

The work described in this thesis was performed at the Laboratory of Microbiology, Wageningen University, the Netherlands.

The research described in this thesis was financially supported by BE-Basic and Corbion as part of the BE-Basic C3-acids project.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design: Tom Lavrijssen
Thesis lay-out: Ferdinand van Nispen |www.*my*-thesis.nl
Printed: GVO drukkers & vormgevers B.V., Ede |www.gvo.nl