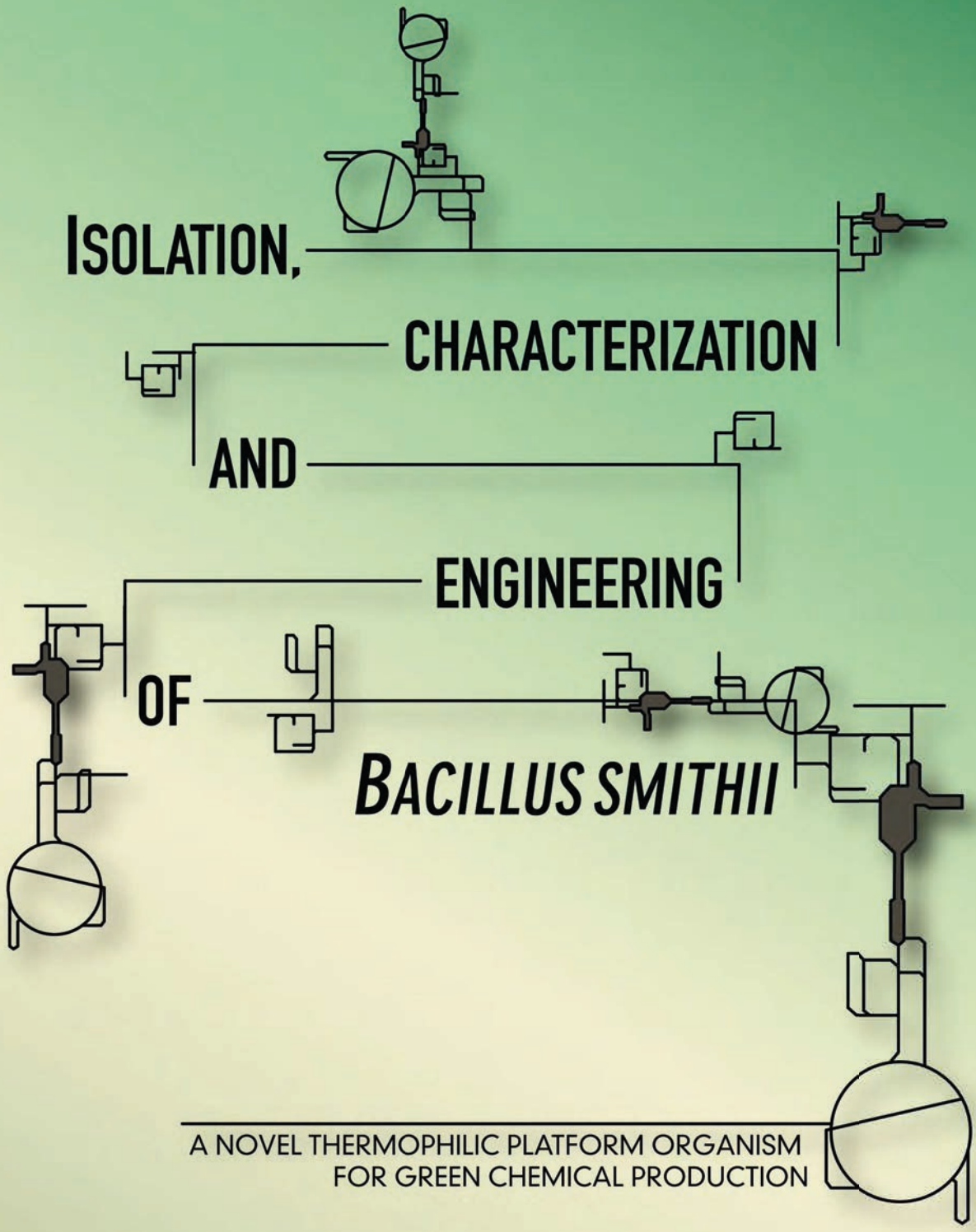

Isolation, characterization and engineering of *Bacillus smithii*

A novel thermophilic platform organism for green chemical production

Elleke F. Bosma

Isolation, characterization and engineering of *Bacillus smithii*

A novel thermophilic platform organism for green chemical production

Elleke F. Bosma

Thesis committee

Promotors
Prof. Dr Willem M. de Vos
Professor of Microbiology
Wageningen University

Prof. Dr John van der Oost
Personal Chair at the Laboratory of Microbiology
Wageningen University

Co-promotor
Dr Ir. Richard van Kranenburg
Corporate Scientist Cell Factories and Team Leader Strain Development
Corbion, Gorinchem

Other members
Prof. Dr Gerrit Eggink, Wageningen University
Prof. Dr Oscar P. Kuipers, University of Groningen
Dr Eric Johansen, Chr. Hansen A/S, Hørsholm, Denmark
Dr Filipe Branco dos Santos, University of Amsterdam

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

Isolation, characterization and engineering of *Bacillus smithii*

A novel thermophilic platform organism for green chemical production

Elleke F. Bosma

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 6 November 2015
at 4 p.m. in the Aula.

Elleke F. Bosma
Isolation, characterization and engineering of *Bacillus smithii* - A novel thermophilic platform organism for green chemical production.
220 pages.

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summary in English

ISBN 978-94-6257-507-3

Voor Oma Sietske

TABLE OF CONTENTS

Chapter 1

General introduction and thesis outline

GENERAL INTRODUCTION

A SHORT HISTORY OF MICROORGANISMS AS CELL FACTORIES

In the late 18th century, Antonie van Leeuwenhoek made his famous microscopic observations of “animalcules” (“little animals”), which we now know as bacteria. In the century after that, Louis Pasteur made his ground-breaking discoveries on the spoilage of wine and milk, revealing that bacterial activity was the cause of this phenomenon. He also postulated that if bacteria could spoil wine and milk from the air, they also might be the cause of human disease. This theory was confirmed in the late 19th century by Robert Koch, who showed that *Bacillus anthracis* was the direct causative agent of anthrax. He was also the first one to establish methods to grow pure bacterial cultures in the laboratory as we do nowadays. Although micro-organisms such as bacteria and yeast had been used already for centuries in basic biotechnological processes such as those for making bread, wine, beer and sake, it was only after the discovery of bacteria as organisms and fermentation as a process that people consciously started to exploit bacteria (as well as other microbes such as yeast and fungi) as cell factories. In the late 19th century, beer brewers started to use pure yeast cultures in order to make their process more reliable and the taste of their beer more constant. During World War I, also other fermentation products started to be used as replacements for products for which a shortage occurred, such as acetone as solvent for the explosive cordite and lactic acid as hydraulic fluid – both products were made by bacterial fermentation and it was around this time that the word ‘biotechnology’ was coined for these processes, in which raw material was converted to valuable products by biological systems. From this time onwards, the field of biotechnology in general, and microbial biotechnology in particular, has been ever-expanding and more and more useful products made by microbes were discovered. Biotechnology took an even larger flight after the discovery of the structure of DNA by Watson and Crick in 1953 (Watson & Crick, 1953) and the subsequent discovery of a technique to transfer DNA from one bacterium to another and from eukaryotes to bacteria in the 1970’s (Cohen & Chang, 1973; Morrow et al., 1974). The ability to modify the genome and with that the production capacities of the host organism, created a whole new range of possibilities to develop microorganisms into efficient cell factories for the production of valuable compounds such as medicines, food, chemicals and fuels.

THE BIOREFINERY

Whereas in a traditional oil refinery valuable products are made from raw fossil materials (oil), in a biorefinery valuable compounds are produced from raw renewable materials (biomass) by microbial fermentation (Figure 1). Like in a traditional refinery, in a biorefinery a wide range of products can be made in different, integrated streams of the refinery, in

which waste products of one production stream can be used as substrate or input of another stream. Also comparable to a traditional refinery, the products in a biorefinery can range from fuels that can be directly used, to so-called building block chemicals that can be used in further processes such as polymerization to create for example bio-plastics. Examples of such building blocks are lactic acid and succinic acid, which can be used as a basis for generating polymers such as plastics, nylons, solvents, pharmaceuticals, cosmetics and food ingredients such as preservatives (Bozell & Petersen, 2010). A major advantage of using microbial conversion of biomass instead of chemical conversion is the optical purity that can generally be achieved via microbial metabolism. Examples of products that can only be produced as racemic mixtures via chemical synthesis, but rather in their pure optical forms by microbes are 2,3-butanediol (2,3-BDO), malate and lactate. To create uniform polymers of these molecules, optically pure building blocks are required (Nair & Laurencin, 2007). Whereas this requires an extra purification step when produced chemically, they can be purified relatively easily when produced by microbes containing only one of the stereo-selective enzymes.

With the growing global demand for fuels and chemicals and because of the environmental impact of the use of fossil resources, the production of fuels and chemicals from renewable resources in a biorefinery is generally considered to be an attractive green alternative. The initial focus of the green industry was mainly on bio-fuels (mainly bio-ethanol), but green chemicals are gaining attention. In 2013, the global market size for alcohols was 110 billion USD and is expected to grow with 4.4% until 2020, whereas the market for organic acids and polymers was 3.5 billion USD and 0.6 billion USD, respectively, with expected respective growths of 8.8% and 13.5% until 2020 (Deloitte, 2014). One of the major concerns with the biological processes is that often food resources such as corn or sugar beet are used as raw material, potentially creating competition between fuels and chemicals and the human food chain. Therefore, current research focuses on the development of 'second generation' bio-fuels and chemicals, which are derived from non-food biomass such as non-edible parts of crops, agricultural, forestry and household waste. This is challenging as the pure sugars used in first generation biorefineries can be easily utilized by the microorganisms that are currently used as production hosts, whereas the second generation biomass consists of lignocellulose. Lignocellulose consists of tightly packed sugar chains containing different types of sugars, which is harder to utilize and requires both additional pre-treatment to make these substrates accessible to enzymatic hydrolysis as well as the addition of hydrolytic enzymes to generate fermentable sugars. 'Third generation' alternatives such as syngas fermentation and photosynthetic processes are also considered but these are very recently emerged strategies that require time to be developed (Sheldon, 2011).

In order to make the second generation large scale biorefineries (Figure 1) economically and ethically feasible, the production costs need to be reduced and non-food substrates

need to be used. There are ample possibilities to achieve cost reduction in each step of the biorefinery such as using cheaper substrates; optimizing pre-treatment; reducing enzyme load; reducing operational costs such as required tanks, cooling and contamination costs; optimizing the microorganisms to produce more of the compound of interest and less by-products; and optimizing down-stream processing. For example, different operation modes using separated or simultaneous saccharification of the biomass by hydrolytic enzymes and fermentation by the microbes are possible (Figure 1). This will be further explained in Chapter 2. To shift from food to non-food resources, either the current organisms can be engineered, or different organisms can be used that are either naturally capable to degrade this type of biomass or can do so in a cheaper way together with hydrolytic enzymes. This thesis describes the development of a thermophilic novel potential platform organism. Chapter 2 will elaborate further on how the use of a thermophilic organism can aid in achieving cost reduction and what the challenges are in the development of these novel production organisms.

Figure 1. The biorefinery. Different modes of separate or simultaneous saccharification and fermentation are depicted with step 3 and 4. Figure adapted from U.S. Department of Energy Genome Programs: http://genomics.energy.gov.

Criteria for an ideal production organism

For many products, there are natural producers that are very efficient in producing that specific compound, but these organisms often have drawbacks such as high nutrient or oxygen requirements, as will be further discussed in Chapter 2. Furthermore, using a specific organism for a specific product requires different cultivation conditions for each organism and each product, which is very inconvenient as the process needs to be optimized for each organism. When a single platform organism is used that can be engineered to make different products, culture conditions will be rather constant and experience in culturing the organism can be more widely applied. There are many non-natural products with high market values that can only be produced by engineered organisms (Ro et al., 2006), creating the absolute requirement for genetic accessibility and a platform organism. For the production of green chemicals and fuels in a biorefinery, ideally one such organism is used. The choice of organism has a large influence on the production costs and therefore, the organism should be as ideal as possible. A list of characteristics of such a platform organism would look as follows:

1. Genetically accessible
   - To obtain maximum yields and high titres in order to maximize profit, minimize waste streams and cheapen downstream processing costs.
2. High productivity, titre and yield.
   - This most likely has to be engineered, stressing the importance of point 1.
   - For a high productivity, a fast metabolism is required.
   - To obtain high productivity, titre and yield, also tolerance against high product and substrate concentrations is important.
3. Flexible in substrate: capable of utilizing a wide range of carbon sources (either raw lignocellulose or in an efficient SSF process), as well as flexibility in this due to for example seasonal changes in biomass composition.
4. Robust against contamination and infection (meaning that fast growth or highly specific conditions are advantageous, as well as phage-resistance) as well as against fluctuations in pH, temperature and gas supply (see also points 4, 5 and 6).
5. Thermophilic
   - Lowers contamination risks and no need for expensive sterilisation.
   - Many thermophiles have a large growth temperature range, which provides robustness against temperature fluctuations.
   - Reduces cooling costs.
   - Increases product and substrate solubility.
   - Enables direct product removal as a cheap downstream processing of volatiles produced above their evaporation temperature.

- In the case of the desire to use an efficient simultaneous saccharification and fermentation (SSF) process it should be a moderate thermophile rather than a hyperthermophile, as commercially available hydrolytic enzymes also have their optimum at these temperatures (50-60°C). Also, at extremely thermophilic temperatures caramelization and Maillard reactions might become problematic, making a moderate thermophile preferred.

6. In case of organic acid production: capable of low-pH growth and production (acid-resistant)
   - In case the pH is well below the product's pKa, produced organic acids can be directly recovered in the acid form, decreasing downstream processing costs and waste streams from base addition.
   - Lowers contamination risks and no need for expensive sterilisation.
   - Generally provides robustness against pH fluctuations.
7. Facultatively anaerobic
   - Compared to aerobic organisms: reduces or eliminates aeration costs. In case of aerated processes, provides more robustness against fluctuations in oxygen supply.
   - Compared to anaerobic organisms: easier to handle, no need for specific anaerobic cultivation facilities or special (reducing) treatment of the media and no inhibition of SSF-enzymes by reduced medium conditions (Podkaminer et al., 2012).
8. Simple and cheap nutritional requirements
   - Reduces medium costs

It is questionable whether a culturable organism exists that has all the above features. Especially the combination between high temperature and organic acid resistance at low pH is very rare. Also, many organisms are capable of growth at low pH but they are not able to produce high titres of organic acids when the pH is below the product's pKa value, as they are not capable of maintaining normal cellular pH under these conditions (Cotter & Hill, 2003). The currently most used platform organisms are *E. coli* and *S. cerevisiae*. Although remarkable results have been achieved with these organisms, they still have drawbacks such as limited substrate-utilizing capacities, their mesophilic nature and oxygen requirements. Especially the use of a thermophilic organism could further reduce the costs of the process compared to mesophiles, as will be discussed in more detail in Chapter 2.

The metabolically versatile family of *Bacillaceae*

The main focus of this thesis are thermophilic bacilli, belonging to the family *Bacillaceae*, which is part of the phylum Firmicutes, the class Bacilli and the order Bacillales (Figure 2). Altogether, facultatively thermophilic bacilli offer many advantages of the points listed above: they have the advantages listed for thermophiles and facultative anaerobes, are

known to degrade a wide range of carbon sources; they are robust against culture condition fluctuations as many of them have a wide temperature and pH growth range; and many of them are able to grow on minimal or relatively simple and cheap media, especially when compared to for example LAB. More detail and various examples of thermophilic bacilli currently used for green chemical production will be provided in Chapter 2.

The *Bacillaceae* family contains a very diverse collection of organisms, ranging from the Gram positive model organism *Bacillus subtilis* to pathogens such as *B. anthracis*, probiotic species such as *Bacillus coagulans* and compost-degrading thermophiles such as *Geobacillus caldoxylosilyticus.* The *Bacillaceae* are very diverse in habitat, ranging from the intestinal tract to soil, compost and hot springs, and the family contains both mesophilic and strictly thermophilic organisms, as well as 'facultatively thermophilic' organisms that grow at mesophilic as well as moderately thermophilic temperatures. Examples of species belon-

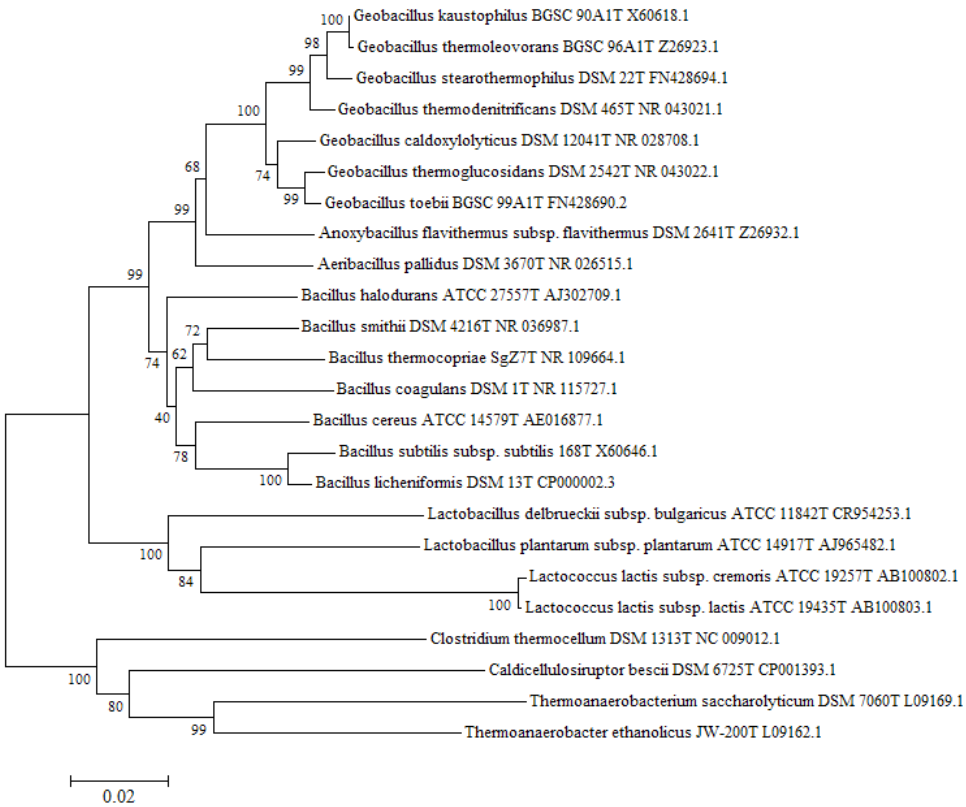

Figure 2. Phylogenetic tree based on 16S rRNA gene sequences. The tree shows the position of several Bacillus type strains, as well as several other industrially important Firmicutes, especially thermophilic Clostridia type strains and Lactic Acid Bacterium type strains. The sequences were aligned with Mega6 software using Clustal, after which the Neighbour Joining method was used to create the phylogram and bootstrap analysis was performed using 1000 replicate analyses. GenBank accession numbers are indicated behind the strain names.

ging to the latter group are *B. coagulans, Bacillus licheniformis* and *Bacillus smithii* (Nakamura et al., 1988; O'Donnell et al., 1980). Remarkably, many of the strict thermophiles such as *Geobacillus thermoglucosidans, Geobacillus toebii* and *G. caldoxylosilyticus* are found in warm environments such as self-heating compost heaps and hot springs, as well as in temperate or even cold environments such as soil or sea water. This wide range of habitats might be caused by the wide-spread ability of bacilli to sporulate: as their spores are able to survive extremely harsh conditions, it is likely that DNA of many different bacilli species can be found in many different environments (Zeigler, 2014). In industrial applications, sporulation is often undesirable as it is a possible way for genetically modified organisms (GMOs) to survive outside the contained factory environment. On the other hand, for many probiotic applications with wild-type, safe organisms, sporulation is very useful as it allows the cells to pass the low-pH stomach and reach the intestinal tract.

Besides in temperature, the family *Bacillaceae* is also highly diverse in metabolism and physiology. Differences in secondary metabolism causes species such as *B. anthracis* and some *Bacillus cereus* strains to be pathogenic, and several others to be very good producers of anti-bacterial compounds (bacteriocins) (Abriouel et al., 2011). Whereas some species are strictly aerobic such as *Geobacillus kaustophilus*, many are facultative anaerobes capable of growth under microaerobic conditions (Hussein et al., 2015). There are no strictly anaerobic *Bacillaceae*; the related class of *Clostridia* contains the strict anaerobes within the phylum Firmicutes.

As described in Chapter 2 and 3, most currently used organisms are not well capable of utilizing lignocellulose or all lignocellulose-derived sugars and these should be engineered to do so. This is achieved by inserting either hydrolytic enzymes for direct biomass degradation, or by inserting genes to degrade certain sugars, mainly C5 sugars, so all components of lignocellulose can be degraded in cooperation with externally added hydrolytic enzymes in an SSF process. Organisms naturally capable of degrading lignocellulose or lignocellulose-derived sugars are generally more efficient at doing this than engineered microbes. An extensive discussion of the enzyme systems and capabilities of different organisms has been reviewed in (Blumer-Schuette et al., 2013). Especially fungi are very good raw biomass degraders, but many of them are not genetically accessible or hard to culture due to mycelia. Therefore, they are mainly used as a source for enzymes. Within the Firmicutes, the anaerobic clostridia are the best raw biomass degraders with species of *Clostridium,. Thermoanaerobacter(ium)* and *Caldicellulosiruptor*. Many of these species are either only cellulolytic or hemicellulolytic and are used in co-cultures of either wild-type (Svetlitchnyi et al., 2013; Wongwilaiwalin et al., 2010) or engineered (Argyros et al., 2011) strains. Within the bacilli, several *Geobacilli* are hemicellulolytic and no reports exist on truly cellulolytic species, although some cellulolytic enzymes are encoded in the genome (Brumm et al., 2015; De Maayer et al., 2014; Hussein et al., 2015). However, a very common

feature of Firmicutes is the ability to utilize a wide range of carbon sources including C_5 and C_6 sugars. Especially in the thermophilic species this creates the ideal organism for green chemical production in an SSF setting. In this thesis, the focus is not on organisms containing hydrolytic enzyme systems but rather on those capable of utilizing both the C_5 and C_6 fraction from biomass in an efficient SSF process. In Chapter 3 of this thesis, several different species belonging to the genera *Aeribacillus, Anoxybacillus, Bacillus, Caldibacillus* and *Geobacillus* will be described, after which the rest of the thesis will focus on *Bacillus smithii*.

THESIS OUTLINE

As introduced briefly above in **Chapter 1**, the use of microbes for the production of chemicals and fuels has a long history, but becomes increasingly relevant nowadays as an alternative for the large-scale production of fuels and chemicals. This thesis describes the development of a novel thermophilic platform organism for the production of green chemicals starting from its isolation, characterization and genome sequencing, to the development of genetic tools and subsequent transcriptomics and metabolomics studies of the mutant strains.

Chapter 2 provides an introduction to green chemicals and their production in a biorefinery via microbial fermentation. The impact of different production systems, substrates and organisms is discussed and extremophiles are introduced as production hosts. Whereas extremophiles have a long tradition as a source of enzymes, their application as cell factories for the production of green fuels and chemicals has emerged only recently. This Chapter provides an overview of thermophiles currently used for green chemical production, discussing the advantages and challenges when using these organisms. As one of the major challenges is the development of genetic tools for thermophilic platform organism, this Chapter also provides an overview of currently available transformation, integration and counter-selection tools for the construction of gene deletions in these organisms. As an extension of this, the diversity and understanding of the metabolism of thermophiles, as well as further options for optimization of their genetic toolkit is further discussed in Chapter 7.

Chapter 3 describes the isolation and screening of facultatively anaerobic, thermophilic bacilli from compost in order to select a potential new thermophilic platform organism for green chemical production. The isolate with most potential for this was selected based on sugar utilization, organic acid production and genetic accessibility. The isolate meeting all selection criteria was *B. smithii* ET 138, for which further characterization is described in the form of fermentation on C_5 and C_6 carbon sources and using different pH-values. In this Chapter, genetic accessibility is also described for two other strains of *B. smithii*, among which type strain DSM 4216^T. For both the isolate and the type strain, the optimization of the electroporation protocol is described. This is the first report of *B. smithii* being genetically accessible.

To identify metabolic potential and to perform metabolic engineering, the genomes of *B. smithii* strains DSM 4216^T and ET 138 were sequenced. **Chapter 4** describes the full genome sequence of *B. smithii* DSM 4216^T. After standard annotation via the RAST pipeline, several pathways commonly present in bacilli could not be found in the *B. smithii* genome. A second analysis based on protein domains was performed in an attempt to identify these pathways. In this Chapter, the differences in the genome between *B. smithii* and related organisms are discussed.

The further development of *B. smithii* into a platform organism is described by the establishment of genetic tools in **Chapter 5**. As L-lactate is the main product of the species under micro-aerobic conditions, a clean gene deletion system was developed via which the *ldhL* gene was deleted from both *B. smithii* strain ET 138 and strain DSM 4216^T. The removal of this gene required a laborious screening procedure and therefore, a counter-selection system was developed based on the toxicity of high X-gal concentration in the presence of a *lacZ* gene. Using this system, the genes encoding sporulation-specific sigma factor *sigF* and pyruvate dehydrogenase subunit E1-α *pdhA* were consecutively removed from ET 138 Δ*ldhL*. In this Chapter, also an initial characterization of the mutant strains is discussed, showing the elimination of L-lactate production and a strong decrease in anaerobic growth and production capacities after *ldhL* removal, a sporulation-deficient phenotype after *sigF* deletion and elimination of acetate production and creation of acetate auxotrophy after *pdhA* deletion.

In **Chapter 6**, the characterization of the mutant strains is extended to transcriptomics, metabolomics, enzyme assays and overexpression of potential NAD^+-regenerating rescue pathways. When comparing the *ldhL* and *ldhL-sigF* mutants to the wild-type, transcriptomics and metabolomics data suggested a bottleneck in glycolysis at the NADH-generating glyceraldehyde-3-phosphate dehydrogenase step. Also, a cycle around pyruvate was observed via pyruvate dehydrogenase and a strongly overexpressed glyoxylate shunt. When heterologously overexpressing NAD^+-generating D-lactate dehydrogenase, the anaerobic growth and production capacities of the *ldhL*-mutant could largely be restored. However, when the native pyruvate carboxylase and NAD^+-generating malate dehydrogenase were overexpressed in an attempt to direct metabolism towards dicarboxylic acid production, hardly any effect was observed. Subsequent enzyme assays show an increased activity of malic enzyme, re-creating the cycle around pyruvate observed earlier in the transcriptomics data without overexpressions. Altogether, this Chapter provides insight into the metabolism of *B. smithii* and indicates directions for further engineering steps.

Chapter 7 summarizes the work described in this thesis, places the research in a broader context and provides an overview of future challenges and possibilities for the use and study of thermophilic bacilli in general and *B. smithii* in particular. Some of the differences in metabolism between *B. smithii* and other bacteria as well as the diversity of metabolism found in thermophiles as described in several chapters of this thesis will be further discussed. Also, further engineering steps to redirect metabolism towards succinate in *B. smithii* are discussed based on the results described in Chapter 6, as well as requirements for improving the genetic toolbox described in Chapter 5 in order to be able to perform this engineering.

Chapter 2

Sustainable production of bio-based chemicals by extremophiles

This chapter has been published as:
Bosma, E.F., van der Oost, J., de Vos, W.M., van Kranenburg, R. 2013. Sustainable production of bio-based chemicals by extremophiles. Current Biotechnology, 2(4), 360-379.

ABSTRACT

To improve microbial fermentation as an efficient way to sustainably produce green chemicals from renewable resources, novel production organisms are being explored. Extremophiles in general, and moderate thermophiles in particular, offer important advantages over well-known mesophilic biotechnology hosts, such as *Escherichia coli* and *Saccharomyces cerevisiae.* These advantages include robust performance at temperatures that allow high substrate and product solubility, decreased contamination risk and growth conditions that match with the optimal conditions for enzymes used in saccharification processes, allowing efficient simultaneous saccharification and fermentation. In this review we will evaluate the use of extremophiles for the production of bio-based chemicals, with the main focus on the potential of these organisms for efficient production of bulk products such as platform chemicals. Examples include the application of thermophilic *Bacillus* and *Clostridium* species for organic acid production. Furthermore, we will discuss the development of genetic tools for biotechnologically-relevant extremophiles, as this is crucial for optimizing their specific production capacities, as well as for developing them as industrial platform organisms.

KEYWORDS

Biorefinery, Building block chemicals, Extremophiles, Genetic tool development, Green chemicals, Halophiles, Metabolic engineering, Sustainability, Thermophiles

INTRODUCTION

Nowadays, the vast majority of fuels and chemicals are derived from fossil resources. However, there is a trend to focus on a future where energy and fuels are expected to be supplied by several sustainable sources such as sunlight, water, wind and biomass (FitzPatrick et al., 2010). For the production of chemicals, however, the alternatives for petrochemical resources are rather limited. Whereas autotrophic systems are being considered for the long term, the near future will most likely see lignocellulosic biomass as a resource for the production of chemicals as a feasible alternative to petrochemicals (Lan & Liao, 2013; Oliver et al., 2013; Sheldon, 2011). Biomass can offer a sustainable and reliable alternative to petrochemicals since it is renewable, regionally available, reduces the carbon footprint of chemicals and fuels, and it can contribute to a more stable and profitable agricultural economy (Sheldon, 2011). Moreover, the production of chemicals from biomass is economically interesting since it has an extremely large added value compared to energy production, while it demands only a fraction of the resources (FitzPatrick et al., 2010). It has been estimated that in the United States, only 3% of the fossil resources are currently used for the production of chemicals, against 97% for energy usage. However, the added value of this 3% chemicals is $375 billion, against $520 billion for the remaining 97% directed at energy usage (FitzPatrick et al., 2010).

A wide range of green building block chemicals can be produced from biomass by microbial fermentation. Building block chemicals are simple, versatile compounds that serve as a basis for many different bulk chemicals such as plastics or nylons. Examples of bio-based building blocks are organic acids such as lactic acid and succinic acid. A shortlist of the most valuable green building block chemicals was made in 2004 by the US Department of Energy (Werpy et al., 2004) and updated in 2010 (Bozell & Petersen, 2010). All chemicals on the DOE-list have the potential to directly replace current oil-based products. Several bio-based chemicals such as lactic acid have a long tradition of sustainable production by industrial biotechnology, but in order to offer a large-scale alternative to petrochemicals they need to become more cost-competitive and switch from food to non-food substrates. Part of the process efficiency and costs are determined by the choice of the organism used for fermentative conversion of biomass to products. In this regard, a highly interesting and useful group of organisms are the extremophiles. Extremophiles have already been used for a long time in applications such as bioremediation and detoxification and especially as a source of enzymes that are stable under harsh conditions, so-called ‘extremozymes’ (Kumar et al., 2011; Morozkina et al., 2010). The use of extremophiles as whole-cell biocatalysts is a relatively new field and initially focussed mainly on fuel production, which has recently been reviewed elsewhere (Chang & Yao, 2011; Taylor et al., 2009). Since a few years, extremophiles, and especially thermophiles, have also been developed for the industrial production

of chemicals in a biorefinery approach. The biorefinery concept is similar to that of an oil refinery, but based on renewable biomass instead of fossil resources. The biomass can be converted to chemicals and fuels via combinations of microbial fermentation and chemical processing in which several production processes can be integrated to minimize waste and optimize revenues (Wellisch et al., 2010). In this review, we will discuss the development of the use of extremophiles for the cost-efficient, sustainable production of bio-based chemicals in a biorefinery.

EXTREMOPHILES IN A BIOREFINERY CONCEPT

In a biorefinery, chemicals and fuels are produced by micro-organisms that ferment pre-treated biomass into one or more products, which are then separated and purified in downstream processing. In current first generation bio-based processes, mostly purified sugars from edible biomass (such as corn, tapioca, sugar cane, and sugar beet) are used as substrate since they yield readily fermentable sugars (glucose and sucrose). However, in order to make bio-based processes more sustainable, substrates should not interfere with the food and feed chain. Thus, ideally the substrate is derived either from cheap, non-food biomass such as corn stover, sugar cane bagasse or wheat straw (Kim & Dale, 2004), or from non-edible energy crops such as switchgrass or miscanthus (Robbins et al., 2012). These substrates form the basis of the second generation bio-based processes that do not compete with food supply. These non-food substrates consist mainly of lignocellulose, which is made up of cellulose, hemicellulose and lignin. Cellulose consists of glucose polymers, whereas hemicellulose consists of a mixture of pentoses and hexoses, mainly xylose, glucose, arabinose, galactose and mannose. Lignin consists of polyphenolic rings that cannot be degraded by most bacteria and is often even inhibitory to the microbial fermentation process as it may generate toxic phenolics during pre-treatment of the biomass. Pre-treatment of lignocellulosic material is required to make the cellulose and hemicellulose accessible for saccharolytic enzymes. The method of pre-treatment has a large influence on the digestibility of the biomass and different types of biomass require different types of pre-treatment (Leu & Zhu, 2013). Pre-treatment methods can involve physical treatment such as milling and irradiation (Peng et al., 2013), chemical treatment such as lime-treatment (Chang et al., 1998), or a combination of physical and chemical methods such as alkaline treatment with steam explosion (Playne, 1984). An extensive description of pre-treatment methods and the effects of these on biomass composition and digestibility is out of the scope of this review, but these topics have been recently reviewed elsewhere (Agbor et al., 2011; Leu & Zhu, 2013; Zhao et al., 2012a). Because of its toxicity, lignin is either removed prior to fermentation, or pre-treatment is performed in such a way that the lignin is

depolymerised or degraded (Lynd et al., 2002; Taylor et al., 2012). The phenolics generated by lignin during pre-treatment are hydrophobic compounds that easily dissolve into the cell membrane, thereby increasing membrane fluidity and permeability and causing cell leakage (Heipieper et al., 1991). The relatively rigid membranes of extremophiles might render some of these organisms more resistant to such membrane-disruptive molecules compared to mesophilic organisms. Other toxic compounds that can be released during pre-treatment are aldehydes, organic acids and ketones (Taylor et al., 2012). An ideal production organism should be resistant to these compounds as much as possible in order not to inhibit the fermentative production process. In well-known mesophilic production organisms such as *E. coli*, several resistance mechanisms to pre-treatment inhibitors are known, which besides general stress responses also include degradation or transformation pathways using specific reductases and dehydrogenases to transform ketones and aldehydes initially into their corresponding alcohols and organic acids, and subsequently into harmless metabolites (Kroutil et al., 2004; Lasko et al., 2000). In extremophiles, these mechanisms have not yet been extensively studied, although there are indications for degradation pathways that target for example phenolic compounds and furans (Taylor et al., 2012; Walton et al., 2010). Also, Gram positive bacteria might have a higher resistance to such compounds because of their cell envelope architecture that consists of a rigid cell wall covering a single membrane (Taylor et al., 2012). Inhibitors and mechanisms of resistance in bacteria and yeast have recently been reviewed elsewhere (Taylor et al., 2012).

After pre-treatment, the (hemi)cellulose polymers need to be saccharified to monomeric sugars by enzymes and fermented into products by the micro-organisms. For these steps there are several options of either separate or simultaneous saccharification and fermentation. Consolidated bioprocessing (CBP) is the most integrated variant, in which the organism produces all enzymes necessary to hydrolyse the pre-treated biomass and no additional enzymes are added (Kambam & Henson, 2010; Lynd et al., 2002). A partly integrated variant is simultaneous saccharification and fermentation (SSF), in which saccharification and fermentation are performed in one reactor, but enzymes are still added to hydrolyse the polymeric biomass into monomeric sugars which are then simultaneously fermented by the microbes within the same reactor (Kambam & Henson, 2010; Lynd et al., 2002). In some literature, a distinction is made between SSF, in which hemicellulose and cellulose fractions are being separated, and SSCF (simultaneous saccharification and co-fermentation), in which these two fractions are utilized simultaneously (Kambam & Henson, 2010; Olofsson et al., 2008). Whether SSF or SSCF is used depends on for example biomass composition, pre-treatment methods and on the organism used for fermentation: organisms that are not able to ferment pentose sugars such as many industrially applied yeast strains are more suited for an SSF-setting. In this review we will use the most commonly used term ‘SSF’ for both processes. The least integrated and least cost-efficient option is separated hydrolysis

and fermentation (SHF) in which two separate steps are needed for the action of enzymes and microbes, for example because enzymes and microbes have different optimum temperatures or pH (Kambam & Henson, 2010; Lynd et al., 2002). The choice of production organism can make a significant difference in the costs, since the addition of enzymes in SSF and SHF is one of the main cost drivers in the process of bio-based chemical production (Buschke et al., 2012; Kambam & Henson, 2010). Commercially available hydrolytic enzymes work optimally around 50°C and pH 4-5 (Kambam & Henson, 2010), which does not match the optimum of currently used mesophilic organisms such as *Escherichia coli* or *Saccharomyces cerevisiae*. For SSF, the use of an organism matching the optimum temperature and pH of the hydrolytic enzymes will reduce costs. It has been shown that when a moderately thermophilic *Bacillus coagulans* with the same optimum temperature and pH as the enzymes is used, a three times lower enzyme dosage is needed in SSF, and productivities are significantly increased compared to when using a mesophilic production organism since fermentation time is reduced (Ou et al., 2009)E. Another advantage of SSF compared to SHF is the reduced risk of enzyme product inhibition, since the hydrolysis products are immediately consumed by the microbes in the SSF setting (Weber et al., 2010). For CBP, it is also necessary that the organism of choice utilizes the polymeric substrates. The ability to utilize polymeric substrates is not strictly related to growth conditions, but extremophiles are known for their metabolic diversity and several moderate and extreme thermophiles are known to be able to convert polymeric substrates into fuels and chemicals in a CBP-setting, such as *Thermoanaerobacterium thermosaccharolyticum* on xylan (Saripan & Reungsang, 2013), *Clostridium clariflavum* on xylan and cellulose and the mixture of these two substrates (Sizova et al., 2011), *Clostridium thermocellum* on cellulose (Tripathi et al., 2010) or *Caldicellulosiruptor* spp. on several steam-explosion-pre-treated lignocellulosic substrates (Svetlitchnyi et al., 2013). Also in an SSF or SHF setting, the organism should have a wide substrate range in order to efficiently utilize all sugars present in the biomass. Current production organisms such as *E. coli* or *S. cerevisiae* are often not naturally capable of utilizing the C_5-fraction. Although they have been successfully engineered to utilize C_5-sugars, organisms that are naturally able to do this are usually more efficient, especially in simultaneous utilization of C_5- and C_6-sugars. Several thermophiles have been shown to efficiently utilize both the C_5- and C_6-sugars derived from lignocellulosic biomass, such as the simultaneous fermentation of C_5- and C_6-sugars by *Thermoanaerobacterium saccharolyticum* (Shaw et al., 2008) and *Bacillus coagulans* (Ou et al., 2011).

Next to cost reduction via SSF, a thermophilic organism offers more advantages such as reduced cooling costs after pre-treatment of the feedstock when operating at a higher temperature compared to mesophiles (Kambam & Henson, 2010; Taylor et al., 2009). In addition, rates of chemistry processes are generally accelerated at higher temperatures, and substrate and product solubility are often increased (Taylor et al., 2011). Furthermore,

contamination risks are lowered since most common contaminants are mesophiles (Ouyang et al., 2013; Qin et al., 2009; Sakai & Yamanami, 2006; Zhao et al., 2010). The lowered contamination risk is not only an advantage of thermophiles, but also of other extremophiles such as halophiles, which are organisms tolerant to high salt concentrations (Tan et al., 2011) since all these organisms grow at unusual and extreme conditions that prevent most common contaminants to grow. Another advantage of using extremophiles is that these organisms are highly robust and generally capable of withstanding fluctuations in conditions such as pH and temperature which might occur in a large scale reactor. Preferably, a production organism should also be anaerobic or facultatively anaerobic, since aeration adds significantly to the process costs. In addition, oxygen transfer in large scale reactors is low, leading to oxygen-limited conditions which limit productivities (Weusthuis et al., 2011). Many of the ethanol-producing thermophiles used in CBP settings are strict anaerobes, and most chemical-producing extremophiles are facultative anaerobes.

In summary, besides producing high amounts of product at high yields, the ideal biorefinery-organism should be able to utilize all sugars present in biomass, either directly from biomass in a CBP process, or with the help of enzymes in an SSF process at elevated temperatures. The use of anaerobic thermophiles offers ample advantages over most current production organisms. Hence, it is crucial that these organisms are further explored and developed in order to improve the production of green chemicals and fuels by microbial fermentation. One of the reasons why mesophilic hosts such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Bacillus subtilis* and *Corynebacterium glutamicum* are currently used for green chemical and fuel production is that for these species genetic tools are well established, enabling their use as platform organisms in a biorefinery. The use of extremophiles as platform organisms can only be established if genetic tools are available. The development of genetic tools for biotechnologically relevant extremophiles is an emerging research field, which will be discussed later in this review. First, several examples will be discussed of extremophiles for the production of green building block chemicals (Table 1) and polymers, including their advantages, shortcomings and future directions.

EXAMPLES OF GREEN BUILDING BLOCK CHEMICAL PRODUCTION BY EXTREMOPHILES

Lactic acid

Lactic acid has been produced at industrial scale for many years. The majority of lactic acid is produced biotechnologically by microbial fermentation, and mainly targeted at the food and pharmaceutical market as an acidulant, flavour and preservative (Hofvendahl &

Hahn–Hägerdal, 2000). Nowadays, it is becoming a bulk building block, in particular for the chemical production of the green polymer poly-lactic acid (PLA) that can be applied in bioplastics. Chemically, lactic acid can be produced by the hydrolysis of lactonitrile de-

rived from petrochemical-based acetaldehydes (Wee et al., 2006). A major advantage of microbial production of lactic acid over its chemical synthesis is that the enzyme-catalysed microbial production yields a significantly higher optical purity. Whereas chemical synthesis always yields racemic mixtures, microbial production can yield purities of over 99.5% of either L(+)- or D(-)-lactic acid by using either L- or D-lactate dehydrogenase, respectively (Figure 1) (Wee et al., 2006). For the production of PLA it is necessary to have an optically pure lactic acid monomer (Nair & Laurencin, 2007), which makes microbial fermentation the preferred production process. In order to make lactic acid more economically viable for the bulk chemical market, further cost reductions are needed. Also, in order to expand the market, non-food substrates should be used, for which it is not favourable to use the traditional lactic acid bacteria (LAB). Although LAB are efficient producers of lactate from the currently

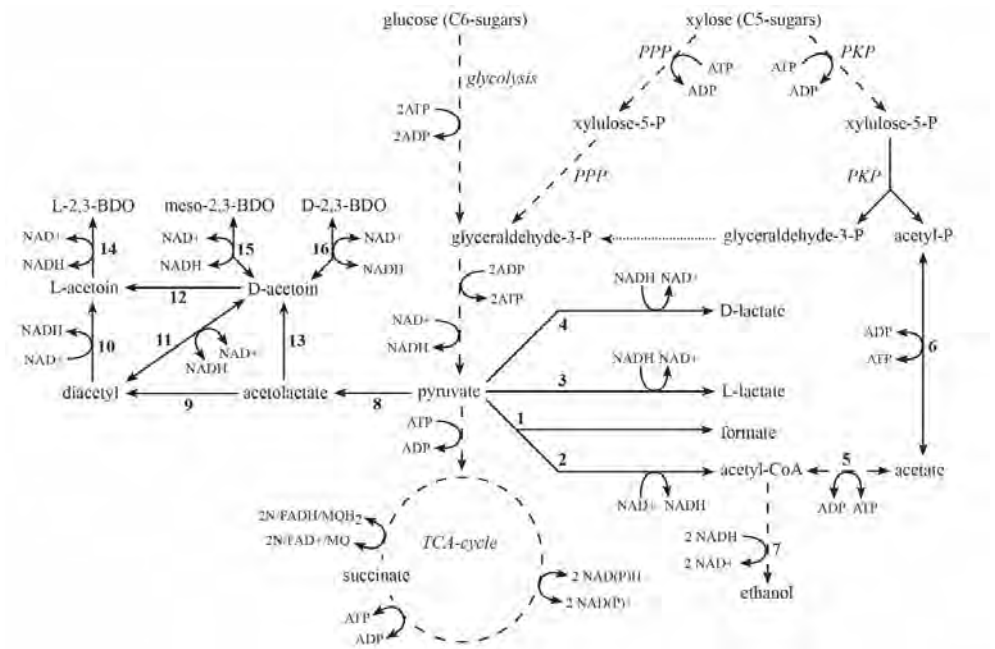

Figure 1. Metabolic pathways discussed in text. Dashed arrows indicate pathways in which intermediate molecules are formed but not depicted. Dotted line indicates that the same molecule is used in different pathways. PPP: pentose phosphate pathway; PKP: phosphoketolase pathway; BDO: butanediol. Enzymes (indicated by numbers) and relevant abbreviations: 1: pyruvate formate lyase (PFL); 2: pyrvate dehydrogenase (PDH) 3: L-lactate dehydrogenase (L-LDH); 4: D-lactate dehydrogenase (D-LDH); 5: phosphotransacetylase (PTA) and acetate kinase (ACK); 6: acetate kinase; 7: acetaldehyde dehydrogenase and ethanol dehydrogenase; 8: acetolactate synthase (ALS); 9: spontaneous decarboxylation; 10: diacetyl reductase; 11: acetylbutanediol hydrolase; 12: racemase; 13: acetolactate decarboxylase; 14: L-2,3-butanediol dehydrogenase; 15: meso-2,3-butanediol dehydrogenase; 16: D-2,3-butanediol dehydrogenase. References used: (Beauprez et al., 2010; Celińska & Grajek, 2009; Hofvendahl & Hahn–Hägerdal, 2000; Otagiri et al., 2010; Tanaka et al., 2002; Thauer, 1988; Wang et al., 2012b).

used substrates glucose and sucrose, they are not well capable of converting C_5-sugars (Abdel-Rahman et al., 2011). Besides, LAB require high amounts of complex nitrogen sources, which add significantly to the costs (Abdel-Rahman et al., 2011; Jang et al., 2012). For example, an osmotic-tolerant *Lactobacillus casei* obtained via random mutagenesis and screening produced 198 g/L lactate at an average productivity of 5.5 g/L/h on 210 g/L glucose in an aerated reactor supplemented with biotin and 20 g/L yeast extract (Ge et al., 2011). Recently, several examples (Table 1) have shown that lactic acid-producing thermophiles might have substantial advantages over the traditional LAB. Not only do these organisms have the general advantages of fermentation at elevated temperatures as mentioned before, but many thermophiles are also able to utilize inexpensive carbon sources, require less complex nitrogen sources and in most cases no aeration is required. In this section, we will discuss several of the examples shown in Table 1 in more detail.

Due to the use of different fermentation conditions in different studies (Table 1), it is hard to compare all strains with each other. Still, some interesting observations can be made. So far, all studies on lactate production by extremophiles are on thermophilic or alkaliphilic bacilli. This group of organisms can naturally utilize both C_5 and C_6 sugars, requires relatively little nutrients, and consists of facultative anaerobes, making them both easy to handle in the laboratory and enabling fermentation without expensive aeration. Also, several of the strains have been shown to perform well at pH values as low as 5.0, which matches the optimum pH of many commercial cellulases used in SSF (Ou et al., 2009; Patel et al., 2005; Wang et al., 2012a). As mentioned earlier, SSF at 50°C and pH 5.0 reduces enzyme costs and improves productivity (Ou et al., 2009). Furthermore, in several of the listed studies (Table 1), the advantage of using thermophiles to prevent contaminations is shown by the use of non-sterile fermentation conditions, in which neither medium nor reactor are sterilized prior to fermentation (Qin et al., 2009; Ye et al., 2013; Zhao et al., 2010; Zhou et al., 2013b). This not only reduces costs by omitting an energy-demanding sterilization step, but also prevents sugar loss during sterilization (Zhou et al., 2013b). Furthermore, thermophilic *B. coagulans* is able to utilize both C_5- and C_6-sugars efficiently and simultaneously (Ou et al., 2011). In 2006, Patel *et al.* selected three *B. coagulans* strains from a collection of 380 isolates from 77 environmental samples for the production of L-lactic acid at 50°C and pH 5.0 from xylose in a minimal medium (Patel et al., 2006). Isolation on minimal medium is important in order to find industrially relevant strains that do not require expensive nutrients such as yeast extract. It was shown that all three isolated strains fermented both xylose and glucose to lactate at high yields, and were also able to efficiently ferment cellulose in an SSF setting. In later studies, two of the strains (17C5 and 36D1) were shown to efficiently produce lactate from sugar cane bagasse and crystalline cellulose in SSF processes (Table 1) and the third strain, P4-102B, was used to develop genetic tools, which will be discussed later (Patel et al., 2006; Rhee et al., 2007; Wang et al., 2011a).

Table 1. Overview of extremophiles used for the production of bio-based building block chemicals.

Prod.[1]	Species and strain	Geno-type[2]	T °C	pH	pH con-trol	Mode[3]	Medium addi-tions (/L)[4]	Substrate type[5]	[Sub-strate] g/L[6]	Titre g/L	Time (h)	Max. prod[7] g/L/h	Av. prod[8] g/L/h	Yield g/g	Reference
D-lac	*B. coagulans* QZ19	Δ*ldhL* Δ*alsS* ev.	50	5.0	KOH	b, ma	10g P, 5g YE, $CaCO_3$	glucose	108	99.8	72	nd	1.39	0.96	(Wang et al., 2011a)
	B. coagulans QZ19	Δ*ldhL* Δ*alsS* ev.	50	5.0	KOH	b, ma	10g P, 5g YE, $CaCO_3$	xylose	80	72.1	94	nd	0.77	0.90	(Wang et al., 2011a)
L-lac	*B. coagulans* XZL4	WT	50	5.3	$CaCO_3$ in medium	fb, ma	10g CSL, 60g $CaCO_3$	Jerusalem artichoke hyd.	140^	134	54	nd	2.5	0.96	(Wang et al., 2012a)
	B. coagulans XZL4	WT	50	5.3	$CaCO_3$ in medium	b, ma	10g CSL, 60g $CaCO_3$	Jerusalem artichoke hyd.	72^	70	47	nd	1.5	0.97	(Wang et al., 2012a)
	Bacillus sp. XZL9	WT	50	5.8	$CaCO_3$ in medium	b, ma	10g YE, 90g $CaCO_3$	corncob molasses	91.4^	70	144	nd	0.49	*0.77*	(Wang et al., 2010)
	B. coagulans	WT	50	6.0	NaOH	b, ma	10g YE	corn fibre hydrolysate	50^	40.2	72-106	2.7	*0.50*	*0.80*	(Bischoff et al., 2010)
	B. coagulans MXL-9	WT	50	6.5	KOH	b, an	10g T, 5g YE	hyd. larch extract	45^	33	58	nd	*0.57*	0.75	(Walton et al., 2010)
	B. coagulans MXL-9	WT	50	6.5	KOH	b, an	10g T, 5g YE	hyd. hard-wood extract	21.4^	20.8	24	nd	*0.87*	0.94	(Walton et al., 2010)
	Bacillus sp. 17C5	WT	50	5.0	KOH	b, ma	5g CSL	sugar cane bagasse hyd.	60^	55.6	192	0.8	0.29	0.89	(Patel et al., 2004)
	B. coagulans DSM 2314	WT	50	6.0	alk.sub.+ $Ca(OH)_2$	fb, ns, SSF	7.5g YE	lime-treated wheat straw	135.5	40.7	55	2.96	0.74	0.81	(Maas et al., 2008)
	B. coagulans 36D1	WT	55	5.0	KOH	b, SSF, ma	2.5g CSL	sugar cane hyd. + SF-cel-lulose	81.3+20	36	144	0.6	0.30	*0.36*	(Patel et al., 2005)
	B. coagulans P4-102B	WT	50	5.0	none	sc, SSF, an	10g P, 5g YE, 4.4g $CaCO_3$	paper sludge (cellulose)	94	90	400	nd	*0.23*	0.83	(Budhavaram & Fan, 2009)
	B. coagulans 36D1	WT	50	5.0	KOH	fb, SSF, ma	2.5g CSL, 75g $CaCO_3$	SF-cellulose	96	80	264	nd	*0.3*	0.88	(Ou et al., 2011)
	B. coagulans 36D1	WT	50	5.0	KOH	b, SSF, ma	5mL CSL	SF-cellulose	40	27	96	nd	*0.28*	0.89	(Ou et al., 2009)E
	B. licheniformis BL3	ev.	50	7.0	KOH	b, ma	MS	cellobiose	15	13.8	48	0.8	*0.29*	0.91	(Wang et al., 2011b)
	B. coagulans WCP10-4	WT	50	6.0	$Ca(OH)_2$	b, ns, SSF, ma	20g YE	corn starch	200	202	38	nd	5.4	0.98	(Zhou et al., 2013b)

Table 1 continued.

	B. coagulans WCP10-4	WT	50	6.0	$Ca(OH)_2$	b, ns, ma	20g YE	glucose	240	210	60	nd	3.5	0.96	(Zhou et al., 2013b)
	B. coagulans 36D1	WT	50	6.0	KOH	fb, ma	10g P, 5g YE	glucose	200	182.2	216	nd	*0.84*	0.92	(Ou et al., 2011)
	Bacillus sp.	WT	45	9.0	NaOH	fb, dp	20g PM	glucose	205	180	112	nd	1.61	0.98	(Meng et al., 2012)
	B. coagulans 2-6	WT	55	5.6	$CaCO_3$	fb, ns, ma	12.6g YE, 1.2g SP, 4g CP	glucose	178	173	70	2.88	*2.47*	0.96	(Qin et al., 2009)
	B. coagulans 2-6	WT	50	5.6	$CaCO_3$	b, ns, ma	12.6g YE, 1.2g SP, 3g CP	glucose	121	118	30	4.37	*3.93*	0.97	(Qin et al., 2009)
	B. coagulans 2-6 Na-2	ev.	50	6.0	NaOH	b, dp	10g YE	glucose	140	106	30	nd	3.53	0.94	(Qin et al., 2010)
	B. coagulans 2-6	WT	50	6.5	NaOH	b**, ns, ma	20g YE, 3g $CaCO_3$	glucose	130	107	37	nd	2.9	0.95	(Zhao et al., 2010)
	B. licheniformis BL3	ev.	50	7.0	KOH	b, ma	MS	glucose	30	25.3	48	1.9	*0.53*	0.82	(Wang et al., 2011b)
	B. coagulans 36D1	WT	50	6.0	KOH	fb, ma	10g P, 5g YE	xylose	200	163	216	nd	*0.75*	0.87	(Ou et al., 2011)
	B. coagulans C106	WT	50	6.0	$Ca(OH)_2$	b, ns, ma	10g YE	xylose	154	141	29	nd	4.8	0.92	(Ye et al., 2013)
	B. coagulans C106	WT	50	6.0	NaOH	b, ns, ma	10g YE, 30g $CaCO_3$	xylose	85	83.6	12	7.5	6.97	0.98	(Ye et al., 2013)
	B. licheniformis BL3	ev.	50	7.0	KOH	b, ma	MS	xylose	30	24.8	48	1.2	*0.52*	0.82	(Wang et al., 2011b)
2,3but	*B. licheniformis* BL8	Δ*ldhL* ev.	50	5.0	KOH	b, ma/ox	10g P, 5g YE	xylose	30	13.8	14	2.64	*0.99*	0.46	(Wang et al., 2012b)
	B. licheniformis BL5	Δ*ldhL*	50	5.0	KOH	b, ma/ox	10g P, 5g YE	glucose	30	13	12	2.35	*1.08*	0.43	(Wang et al., 2012b)
2,3but/ acetoin	*Geobacillus* sp. XT15	WT	55	8.0	none	b, ox	60g CSL, 10g YE	glucose	220	14.5 + 7.7	48	nd	*0.46**	0.35*	(Xiao et al., 2012)
acetoin	*Geobacillus* sp. XT15	WT	55	8.0	none	b, ox	20g P, 10g YE	glucose	20	8	10	nd	*0.8*	0.40	(Xiao et al., 2012)
suc	*C. thermosuccino-genes* DSM 5809	WT	58	7.0	none	b, an	0.025g YE, 0.025g aa+vit	inulin	1.5^	0.7	nd	nd	nd	0.47	(Drent et al., 1991)
	C. thermosuccinogenes DSM 5809	WT	58	7.3	Na_2CO_3	b, an	1g YE, 0.03g aa+vit	glucose	5	nd	nd	nd	nd	0.54	(Sridhar & Eiteman, 2001)

Legend see next page.

Legend Table 1.
The table is sorted by product and subsequently by titre per substrate. nd: no data/not determined. Numbers in italics have been calculated by the authors of this review, numbers in standard font are obtained directly from the original article. Yield (g/g) = titre (g/L)/consumed substrate (g/L). Average productivity (g/L/h) = titre (g/L) / time (h).
[1] Prod: product. D-lac: D-lactate; L-lac: L-lactate; 2,3but: 2,3-butanediol; suc: succinate.
[2] ev.: evolved via laboratory evolution.
[3] sc: semi-continuous; b: batch; fb: fed-batch; dp: dual phase (aerated phase followed by non-aerated phase); ns: not sterile; ma: micro-aerobic (no gas added to the fermentation); an: anaerobic (CO_2 or N_2 added to the fermentation); ox: aerobic (O_2 or air added to the fermentation). ** repeated batches, batch nr 8 is shown.
[4] YE: yeast extract; SP: soy peptide; P: peptone; CP: cottonseed protein; T: tryptone; CSL: corn steep liquor; ME: meat extract; PM: peanut meal; MS: mineral salts; aa: amino acids; vit: vitamins.
[5] hyd: hydrolysed/hydrolysate; SF-cellulose: Solka Floc cellulose.
[6] ^sugar amount is the total added monomeric contents in g/L, as reported by the authors of the original article.
[7] Max. prod: maximum productivity. In the case of fed-batch, this is mostly the maximum average productivity during the course of the fed-batch.
[8] Av. prod: average productivity (g/L product/hours of fermentation time).
* These data reflect the combined 2,3-butanediol and acetoin productivity and yield.

Traditional LAB are either unable to utilize pentoses, or they convert pentoses to lactate and acetate in an equimolar amount by using the phosphoketolase pathway (PKP) (Abdel-Rahman et al., 2011), whereas the *B. coagulans* isolates of Patel *et al.* were shown to use the homolactic pentose phosphate pathway (PPP), which maximises the yield of lactate per pentose (Patel et al., 2006) (Figure 1). As an exception to most LAB, *Lactococcus lactis* IO-1 was shown to use a combination of PKP and PPP, thereby slightly increasing the yield on xylose to 0.67 g/g (Tanaka et al., 2002). *Lactobacillus plantarum* has been engineered to metabolize pentoses via the PPP and this resulted in a yield of 89% of the theoretical maximum (Okano et al., 2009). The yield of *B. coagulans* C106 on xylose is 0.98 g/g or 98% of the theoretical maximum, making it superior to LAB in C_5-utilization (Ye et al., 2013).

Besides *B. coagulans*, also another *Bacillus* sp. has recently been described for L-lactate production. This is a thermophilic *B. licheniformis*, and its use constitutes the only process so far that uses a completely minimal medium (Wang et al., 2011b). In most cases *B. coagulans* requires less complex nitrogen sources than traditional LAB and in some cases only 2.5 g/L of the relatively cheap corn steep liquor (CSL) is used. However, still complex nitrogen sources are used in almost all fermentation studies published to date. In the case of *B. licheniformis*, the strain was evolved to be able to perform well in minimal medium (Wang et al., 2011b). The use of such laboratory evolution as a tool to optimize production strains in combination with metabolic engineering has more often been shown to be a powerful tool and has recently been reviewed (Portnoy et al., 2011). Another *B. licheniformis* strain has been shown to be able to convert saccharified liquid of municipal food waste ('model kitchen refuse' or MKR) efficiently to lactic acid (Sakai & Yamanami, 2006). Several other thermophilic bacilli utilizing non-food or waste substrates have been described, either in efficient SSF processes such as *B. coagulans* DSM 2314 on wheat straw (Maas et al., 2008) or after pre-hydrolysis of the substrate such as *B. coagulans* MXL-9 on hydrolysed hard-wood extract (Walton et al., 2010) (Table 1).

Since the pre-treatment and hydrolysis still remain costly steps in lignocellulose-derived fuel and chemical production (Barakat et al., 2013; Hasunuma & Kondo, 2012), most industrial processes still rely on edible crops, which are mostly sucrose and corn starch or the glucose derived from that (Chen & Patel, 2012; Mojović et al., 2012). Therefore, some authors anticipate that in the near future substrates will likely remain to be mainly starch-like materials and thus focus on improving fermentation capabilities on starch (Zhou et al., 2013b). An important factor is the substrate tolerance of the strain, since often high substrate concentrations have an inhibitory effect. In the case of substrate inhibition, fed-batch strategies can be used. However, the process is simplified by adding high amounts of substrate at once, which stresses the importance of finding strains more tolerant to high substrate concentrations. Therefore, a *B. coagulans* strain was isolated that showed the highest glucose tolerance and utilization reported so far, namely from 240 g/L glucose it

produced 210 g/L L-lactate with a productivity of 3.5 g/L/h. Also, this strain was able to efficiently utilize starch in a non-sterile SSF fermentation at 50°C and pH 6.0 (Zhou et al., 2013b). In SSF and other hydrolysate fermentations, substrate inhibition might be less of a problem, depending on the efficiency of the used saccharification enzymes. However, in these processes the tolerance to other inhibitory products such as furans and acetate becomes important. Walton *et al.* used a *B. coagulans* strain that was found to not only be tolerant to inhibitors, but even metabolize them (Walton et al., 2010).

In many industrial lactic acid fermentations, the pH is controlled by adding $Ca(OH)_2$ or $CaCO_3$, creating calcium-lactate in the fermentation broth. The advantage of trapping lactate as calcium-lactate is that product inhibition is reduced (Ou et al., 2011). This was clearly shown by Ye *et al.*, who showed an almost 2-fold higher productivity at a 2-fold increased sugar concentration when the pH was regulated by $Ca(OH)_2$ instead of NaOH (Ye et al., 2013). However, during lactic acid purification in downstream processing, H_2SO_4 needs to be added to obtain lactic acid out of the calcium-lactate, thereby forming insoluble $CaSO_4$ (gypsum). In this way, approximately one ton of crude gypsum is produced for every ton of lactic acid (Datta & Henry, 2006). A solution for this might be to control the pH with NaOH instead of $CaCO_3$, which does not yield gypsum and allows lactic acid to be obtained via alternative downstream processing methods (Meng et al., 2012; Qin et al., 2010). However, most organisms are not tolerant to the high concentrations of Na^+ established in this way, thereby reducing the final titres that can be reached. Qin *et al.* used random mutagenesis to obtain a sodium-lactate tolerant strain (strain Na-2) from parent strain *B. coagulans* 2-6. The mutant strain reached an approximately two times higher cell dry weight than its parent strain and produced 10.5% more lactic acid after 38h of fermentation (Qin et al., 2010). In this study, also the effect of aeration and agitation was tested. In accordance with other publications, it was found that aeration had a positive effect on growth but a negative effect on lactate production since more by-products were formed. Since a higher cell density would make the culture more resistant to the sodium lactate, the authors tested a dual-phase fermentation in which the culture was aerated during the first 12h, after which aeration was switched off. In this way, 106 g/L lactate was produced at 3.53 g/L/h from 140 g/L glucose at pH 6.0 and 50°C, using NaOH as a neutralizing agent (Qin et al., 2010). Although these are interesting results, a drawback of a dual phase system is that it requires aerated fermenters, which substantially adds to the costs on an industrial scale because reactors should be equipped with gas supply. The same set-up is used by Meng *et al.*, who used an alkaliphilic, thermotolerant *Bacillus* sp. to produce lactate with NaOH as neutralizing agent. Alkaliphiles are organisms growing optimally at a pH above 9.0 and are generally also resistant to monovalent ions such as sodium ions (Meng et al., 2012). Another interesting feature of this strain is its nutrient requirements: its optimum nitrogen source was 20 g/L peanut meal, which is a cheap by-product of oil extraction from peanut seeds in China. Using

a single-pulse fed-batch fermentation in which the pH was controlled at 9.0, this strain has been reported to produce 180 g/L lactate at 1.61 g/L/h at 98.6% yield in 112h from in total 205 g/L glucose (Meng et al., 2012). An overview of other biotechnological applications of alkaliphiles can be found in (Sarethy et al., 2011), but most of these are not yet relevant for green chemical production.

Patel *et al.* found only one of their 380 *B. coagulans* isolates to be genetically accessible (Patel et al., 2006), indicating that the majority of the *B. coagulans* strains are not genetically accessible and are therefore mainly suitable for L-lactate production, which many of them produce in a homo-fermentative way (Patel et al., 2006). To date, two strains have been found to be transformable by exogenous DNA. Strains DSM 1 and QZ19 (originally strain P4-102B from Patel *et al.*) were engineered to efficiently produce optically pure D-lactate instead of L-lactate (van Kranenburg et al., 2007; Wang et al., 2011a) (Table 1). A different metabolic engineering strategy was applied to the two strains, resulting in different metabolic patterns. In DSM 1, the native *ldhL* gene encoding L-lactate dehydrogenase was exchanged with the *L. delbrueckii ldhA* gene, encoding D-lactate dehydrogenase, yielding a homolactic strain in which by-products were below the detection limit. This strain produced 28 g/L pure D-lactic acid from 30 g/L glucose (van Kranenburg et al., 2007). Strain DSM 1 was further used for establishing a genetic toolbox using the Cre-*lox* system (Kovacs et al., 2010), which will be discussed in the section on genetic tools. In strain QZ19, first an *ldhL* deletion mutant was made, which showed decreased growth and sugar consumption, an increased production of ethanol, acetoin and D-2,3-butanediol (2,3-BDO) and only a small increase in D-lactate production. Therefore, the *alsS* gene (Figure 1) was deleted in order to prevent acetoin and 2,3-BDO production. At pH 7.0 this double knockout strain was able to grow, but it produced only minor amounts of D-lactate and in total only 5% of the amount of acids produced by the wild type, and it did not grow at pH 5.0. After in total 278 days of metabolic evolution by serial transfers of the double knockout strain, the strain was able to produce 99.8 g/L D-lactate from 108 g/L glucose in 72h at 50°C and pH 5.0. Also, the evolved strain was able to efficiently function in an SSF process on cellulose at 50°C and pH 5.0. It was shown that in the evolved strain, a glycerol dehydrogenase had acquired D-LDH activity through two point mutations, thus improving D-lactate production (Wang et al., 2011a). This is another example of the power of laboratory evolution for strain improvement.

Other building block chemicals: 2,3-butanediol, acetoin and succinic acid

As already mentioned above, the use of extremophiles for the production of chemicals is a relatively new field and has so far mainly focussed on lactic acid. Partly hampered by the lack of genetic tools to create platform organisms, there are not that many examples

yet of the use of these organisms for other products, although recently two examples have been published.

The same parent strain of *B. licheniformis* that was used to develop a strain that produces L-lactate in minimal medium from glucose or xylose (Wang et al., 2011b), has been used to create an *ldhL* knockout strain that produces chirally pure D-2,3-BDO. The latter product can potentially be applied as a liquid fuel as well as as a platform chemical for the pharmaceutical, food, agro or fine chemical industries. 2,3-BDO has 3 isoforms: dextro, levo and meso. In chemical synthesis the chiral purification is a costly step, but like with lactic acid, the pure forms can be produced by microbes (Figure 1), which makes microbial fermentation for 2,3-BDO economically very interesting (Zeng & Sabra, 2011). The best current producer of meso-2,3-BDO is a randomly mutated *Klebsiella pneumonia* SDM strain which in a glucose fed-batch fermentation supplemented with 8.27 g/L CSL produced 150 g/L meso-2,3-BDO in 38h with a productivity of 4.21 g/L/h and a yield of 0.48 g/g (Ma et al., 2009), which is close to the theoretical maximum of 0.5 g 2,3-BDO per g glucose. Most organisms so far produce meso- or L-2,3-BDO, whereas *B. licheniformis* BL5 produces pure D-2,3-BDO, which has an important application as antifreeze for which the meso- and L-form cannot be used. The strain produced 13.8 g/L pure D-2,3-BDO from 27 g glucose in 12h (average productivity 1.08 g/L/h), at a yield of 0.43 g/g with a maximum productivity of 2.35 g/L/h (Wang et al., 2012b). For comparison, the known mesophilic 2,3-BDO producer *Klebsiella oxytoca* produces 90.8 g/L 2,3-BDO from 187.8 g glucose in 60h with an average productivity of 1.51 g/L/h and a yield of 0.48 g/g (Ji et al., 2010). *B. licheniformis* is a class I organism, making it preferred over the mesophilic *Klebsiella* spp., which are class II pathogens. Moreover, because of its ability to ferment at pH 5.0 and 50°C, *B. licheniformis* BL5 is well suited for use in an economically viable SSF setting (Wang et al., 2012b). Although several by-products were still present and further metabolic engineering could optimize 2,3-BDO titres and yields, this is the first report on a thermophilic 2,3-BDO producer, illustrating the importance of genetic tools to develop these bacteria into platform organisms.

Recently, a newly isolated *Geobacillus* sp. was reported to be the first thermophile to naturally produce 2,3-butanediol and acetoin as major products and also showed the ability to utilize galactose, mannitol, arabinose and xylose, creating potential for lignocellulose utilization in an SSF process (Xiao et al., 2012) (Table 1). A mixture of acetoin and 2,3-BDO is easy to separate due to their different boiling points (Xiao et al., 2012). Acetoin excretion is unusual in the genus *Geobacillus*, but with the newly isolated strain a maximum acetoin yield of 0.40 g/g could be achieved on 20 g/L glucose, which is 82% of the theoretical maximum (Xiao et al., 2012). At 20 g/L glucose no 2,3-BDO was observed, but when 220 g/L glucose was used, also 2,3-BDO was produced and 155 g/L glucose remained unutilized, resulting in a yield of 69%, which is significantly lower than the initial 82%. The authors indicate that bacterial growth was not influenced by substrate inhibition, but that acetoin

Table 2. Overview of completed genome sequences of extremophiles relevant for the production of green chemicals.

Organism	Gen. acc.	Current product	Size (Mbp)	CDR (nr.)	G+C %	Pl.	Reference
Bacillus coagulans 36D1	nd	Lactate	3.55	3449	47	-	(Rhee et al., 2011)
Bacillus coagulans 2-6	nd	Lactate	3.07	3052	47	+	(Su et al., 2011a)
Bacillus coagulans DSM1 ^{T}	+	Lactate	3.02	3437	47	nd	(Su et al., 2012)
Bacillus licheniformis 10-1	nd	2,3-BDO	4.32	4650	46	-	(Li et al., 2012)
Bacillus licheniformis 5-2	nd	2,3-BDO	4.16	4452	46	-	(Li et al., 2012)
Geobacillus thermoglucosidans TNO-09.020	nd	n.a.	3.75	4400	44	nd	(Zhao et al., 2012b)
Geobacillus thermoglucosidans C56-YS93	nd	n.a.	3.99	4135	44	+	GOLD; Mead *et al*., 2011
Thermoanaerobacter pseudethanolicus 39E	rel +	Ethanol	2.36	2243	35	-	(Hemme et al., 2011)
Thermoanaerobacter sp. X514	rel +	Ethanol	2.46	2349	35	-	(Hemme et al., 2011)
Thermoanaerobacter mathranii DSM 11426 ^{T}	rel +	Ethanol	2.31	2361	34	-	GOLD; Hemme *et al*., 2010
Thermoanaerobacter ethanolicus JW200	+	Ethanol	3.09	3449	34	-	GOLD; Hemme *et al*., 2011
Thermoanaerobacterium thermosaccharolyticum M0795	rel +	Ethanol	2.78	2921	34	+	GOLD; Shaw *et al*., 2012
Thermoanaerobacterium saccharolyticum JW/SL-YS485	+	Ethanol	2.72	2931	35	+	GOLD; Herring *et al*., 2012
Thermoanaerobacterium thermosaccharolyticum DSM 571^{T}	rel +	Ethanol	2.79	2830	34	-	GOLD; Hemme *et al*., 2010
Clostridium thermocellum DSM 1313	+	Ethanol	3.56	3102	39	-	(Feinberg et al., 2011)
Clostridium thermocellum ATCC 27405 ^{T}	nd	n.a.	3.84	3335	39	-	GOLD; Wu *et al*., 2007
Caldicellulosiruptor bescii DSM 6725 ^{T}	+	Ethanol	2.91	2662	35	+	(Kataeva et al., 2009)
Several *Caldicellulosiruptor* spp.	nd	Ethanol/H_2	2.40-2.97	2322-2707	35-36	+/-	(Blumer-Schuette et al., 2011)
Caldicellulosiruptor saccharolyticus	nd	H_2	2.97	2679	35	-	(van de Werken et al., 2008)
Caldicellulosiruptor obsidiansis OB47 ^{T}	nd	n.a.	2.53	2389	35	-	(Elkins et al., 2010)

List includes strains that are used for chemical production, strains that are relevant for fuel production and have been shown to be genetically accessible, and strains of which several related strains of the same species are genetically accessible (indicated as 'rel +'). Abbreviations: n.a.: not applicable; Gen. acc.: genetic accessibility; CDR: number of predicted coding regions; Pl.: plasmid; GOLD: Genome Online Database; nd: not determined/no data. Note that "*G. thermoglucosidans*" was previously named "*G. thermoglucosidasius*" (Coorevits et al., 2012).

and 2,3-BDO production is a growth-associated phenomenon (Xiao et al., 2012). This is a potentially interesting process and organism and there is some room for improvement by process engineering, e.g. by using a fed-batch strategy as was successfully applied by Ma et al. for *Klebsiella* sp., in which the glucose concentration was kept at 20-30 g/L maximally (Ma et al., 2009). At about the same time as the publication of the 2,3-BDO-producing *Geobacillus* sp., the genomes of two newly isolated thermophilic *B. licheniformis* strains were published (Li et al., 2012) (Table 2). In the genome announcement it is mentioned that these two organisms naturally produce 2,3-BDO at yields of 96%, but these results have not yet been published (Li et al., 2012).

Another very interesting building block chemical is succinic acid, which was listed by the DOE as one of the most high-potential building blocks (Bozell & Petersen, 2010; Werpy et al., 2004). Several mesophiles have been reported to naturally produce high amounts of succinic acid (Guettler et al., 1999; Kuhnert et al., 2010; Lee et al., 2002; Nghiem et al., 1997), but so far only one thermophile is known that naturally produces high amounts of succinate (Table 1). *Clostridium thermosuccinogenes* is able to grow on several different carbon sources in a medium containing relatively low amounts of complex nutrients (Sridhar & Eiteman, 2001). Its succinate yield on glucose is comparable to that of mesophilic natural succinate producers (Beauprez et al., 2010; Sridhar & Eiteman, 2001). However, little literature is available on this strain and fermentation data on substrate concentrations above 5 g/L have not been reported. Hence, it is hard to estimate the potential of this organism.

EXAMPLES OF POLYMER PRODUCTION BY EXTREMOPHILES

Polyhydroxyalkanoates (PHAs) are polymers that many micro-organisms accumulate as intracellular energy storage compounds when there is excess carbon but little other nutrients are available (Quillaguamán et al., 2010). PHAs are biodegradable and produced from renewable resources by microbes, which makes them well suited as a sustainable alternative for traditional petrochemical-based polymers (Quillaguamán et al., 2010). Industrial production is currently achieved with *E. coli* (Quillaguamán et al., 2010), but in order to expand the markets PHAs should become more price-competitive with traditional oil-based polymers. Several thermophiles and halophiles have been shown to accumulate PHAs and they might be interesting alternative production hosts.

Poly(3-hydroxybutyrate) (P(3HB)) is the most common PHA-form synthesized by microbes (Quillaguamán et al., 2010). Recently, it was shown that when grown in a mineral salts medium on glucose, moderately thermophilic *Chelatococcus* sp. MW10 produced up to 55 wt% at a cell density of 12.7 g/L in a 42 L batch culture with a P(3HB) yield of 0.14 g/g at

50°C (Ibrahim & Steinbüchel, 2010). The only disadvantage of this strain is that it is not able to utilize xylose and hardly able to utilize glycerol (Ibrahim et al., 2010), which makes it rather limited in carbon utilization abilities. Ibrahim *et al.* isolated P(3HB)-accumulating thermophilic bacteria using minimal medium containing different carbon sources and Nile red to stain for hydrophobic cellular inclusions. This resulted in several PHA-producing isolates at 50°C, whereas at 60°C only low cell densities were observed and no PHA production was found (Ibrahim et al., 2010).

Like thermophiles offer the advantage of lowering cooling costs and every type of extremophile offers the advantage of lowering contamination risks, extremophiles other than thermophiles might offer different advantages for particular processes. For example, the use of halophiles could make downstream processing of intracellular compounds such as P(3HB) easier since the organisms lyse in the absence of salts, which simplifies the extraction of the compounds (Don et al., 2006; Oren, 2010). Moreover, the use of halophiles can reduce medium cost if they can grow in non-sterilized sea water. This possibility was shown using *Halomonas* sp. SK5 for P(3HB) production (Rathi et al., 2013). Using non-sterilized sea water with oil palm trunk sap (OPTS), a substrate obtained from felled oil palms, this organism produced P(3HB) up to 24% of its cell weight at a cell density of 2.4 g/L in 48h. In artificial sea water this was higher with 44 wt% at a biomass of 2.8 g/L (Rathi et al., 2013), but nevertheless the use of sea water is an interesting perspective. Another halophilic PHA-production organism is the archaeon *Haloferax mediterranei*, which has little nutritional requirements and utilizes several carbon sources (Rodriguez-Valera et al., 1983). This organism was used to produce a co-polymer of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (P((3HB)-co-3HV)), a stronger and more thermo-resistant PHA-variant. Most strains producing this compound require co-feeding of propionate, but *Haloferax mediterranei* could produce P((3HB)-co-3HV) on pure glucose (Don et al., 2006). The halophilic γ-proteobacterium *Halomonas boliviensis* might also be a promising candidate since it can accumulate PHA up to 88% of its cell dry weight (Quillaguamán et al., 2006).

Although the above mentioned organisms show potential for the production of biodegradable plastics, none of them is currently used on a larger scale. Other biotechnological applications of halophiles have been reviewed elsewhere (Oren, 2010) and the use of halophiles and archaea for PHA production has also been reviewed more extensively elsewhere (Poli et al., 2011; Quillaguamán et al., 2010).

POLYSACCHARIDE-UTILIZING EXTREMOPHILES

Isolation and screening studies are very important to explore the potential of naturally occurring organisms that have characteristics such as a broad substrate range, that have the ability to form certain products, or that are able to perform well under extreme conditions. In the previous sections, several isolation studies were described that were aimed at finding extremophiles suitable for the efficient production of specific products. Next to screening for products, screening for the utilization of complex substrates is very important to develop sustainable fermentation technologies, especially for CBP-setups. Here, we will discuss several examples of the isolation of polysaccharide-utilizing extremophiles. The focus is on isolation because most of the polysaccharide-utilizing isolates are used for bio-fuel production (reviewed in (Olson et al., 2012)), but isolations might as well be targeted at the conversion of polymers to chemicals. Also, the ethanol-producing, polymer-utilizing extremophiles might be used for green chemical production if genetic tools are developed, which will be discussed in the next section. Most thermophilic strains isolated on polymeric substrates are heterofermentative Firmicutes, to which also the above-described bacilli belong.

Several anaerobic, cellulolytic, thermophilic bacteria were isolated by Ozkan *et al.* on minimal medium, after which their fermentation products and cellulolytic capacities at 55°C were determined (Ozkan et al., 2001). Most of the strains isolated by Ozkan *et al.*, of which the majority was shown to be *Clostridium thermocellum*, produced ethanol and acetate as their major fermentation products, but several strains also produced lactate and one strain produced lactate as its major product. All isolates were also tested for restriction endonuclease activity to estimate their potential genetic accessibility (Ozkan et al., 2001), which was examined subsequently (Tyurin et al., 2004). Nowadays, *C. thermocellum* is one of the best-studied cellulolytic thermophiles with potential use in CBP-processes, for which several genetic tools have been developed, as will be discussed later in this review. In another study, thermophilic microbial communities were enriched at 55°C from self-heated compost and the microbial diversity of this stable cellulose-enrichment was examined. In 3 different enrichments, 3-18 different strains were found and identified to all belong to the family *Clostridiaceae* (Izquierdo et al., 2010). In a subsequent study, nine cellulolytic isolates were obtained from the enrichments (Sizova et al., 2011). These were shown to be different from *C. thermocellum*, but displayed similar cellulolytic activity and two isolates also showed hemicellulolytic activity and utilized xylan and cellulose simultaneously (Sizova et al., 2011). *C. thermocellum* is not capable of utilizing the hemicellulose fraction and is therefore often used in very efficient co-cultures with hemicellulolytic species (Demain et al., 2005). Methods to isolate cellulose-utilizing organisms, including an overview of different substrates and enzymes have been extensively reviewed by McDonald *et al.* (McDonald et al., 2012).

Close relatives of *Clostridia* are *Thermoanaerobacter* species, which are strictly anaerobic, thermophilic Gram positives that have potential for lignocellulose utilization. Balk *et al.* isolated a novel *Thermoanaerobacter thermohydrosulfuricus* subsp. *carboxydovorans*, strain TLO, from a geothermal spring in Turkey (Balk et al., 2009). It was proposed to be a novel subspecies because of differences in lipid composition compared to close relatives and also because of its ability to use CO as an electron donor, which is converted to H_2 and CO_2. The strain showed a wide growth range of 40-80°C and pH 4.5-9.0, with its optimum at 70°C and pH 6.5 and it was able to utilize a wide range of substrates, including xylan, pectin, starch, cellobiose and to a lesser extent also carboxymethyl cellulose (CMC). The major end product of sugar fermentation of the new isolate was lactate, whereas many other *Thermoanaerobacter* spp. produce mainly ethanol and hydrogen. However, for several strains of this species genetic tools have been developed, so they also have potential for the production of chemicals. Interestingly, the product formation of strain TLO was pH-dependent. At pH 4.5-5.0, lactate was produced at 1.7 mol/mol glucose with ethanol as a by-product at 0.1 mol/mol glucose, whereas at pH 9.0, acetate was produced at 1.6 mol/mol glucose and lactate only at 0.4 mol/mol glucose. The homolactic fermentation at low pH combined with the ability to utilize a wide range of substrates makes strain TLO an interesting candidate for lactic acid production. Furthermore, its ability to utilize CO makes it a candidate for cheap production of H_2 (Balk et al., 2009).

The genus *Caldicellulosiruptor* is another example of complex biomass-degrading thermophiles belonging to the class of *Clostridia*. Members of the anaerobic Gram-positive bacterial genus *Caldicellulosiruptor* are the most extremely thermophilic cellulolytic organisms known with an optimum growth temperature of 70°C and they contain a large array of genes involved in polysaccharide utilization (van de Werken et al., 2008). *Caldicellulosiruptor* spp. produce ethanol or hydrogen, with lactate and acetate as major by-products (Bielen et al., 2013; Yee et al., 2012) and are also able to grow on a chemically defined medium (Willquist & van Niel, 2012). Several (hemi)cellulolytic *Caldicellulosiruptor* strains, as well as several strains of the less well-known clostridial *Caloramator* species, were isolated by a high-throughput method using flow cytometry to isolate single cells (Hamilton-Brehm et al., 2012). This method had previously been used for isolating marine bacteria, but had to be optimized for isolation of thermophilic anaerobes by sparging the flow cytometer with N_2 and subsequent plating in 96-wells plates in an anaerobic chamber (Hamilton-Brehm et al., 2012). In another study, several *Caldicellulosiruptor* and *Thermoanaerobacter* strains were isolated on filter paper, as well as on unwashed, pre-treated poplar wood in order to isolate strains resistant to inhibitors released during pre-treatment (Svetlitchnyi et al., 2013). Both species were shown to efficiently utilize several lignocellulosic substrates directly to ethanol in mono-culture as well as in mixed cultures (Svetlitchnyi et al., 2013).

A less well-known process for which extremophiles have been isolated is the sustai-

nable production of hemp fibres. These fibres are produced by bast fibre plants on which they are embedded in a pectic polysaccharide network. To be able to use these fibres in the textile and paper industry they need to be separated from the woody core. Nowadays, this is performed by processes that can damage the fibres or by processes that are highly environmentally unfriendly. An alternative process is enzymatic treatment, but no suitable process has been developed so far. Valladares Juárez *et al.* (Valladares Juárez et al., 2009) searched for a whole-organism solution to sustainably and efficiently utilize the pectin surroundings of the hemp fibres while keeping the fibres intact. The organism should be thermophilic to prevent contamination of the process and alkaliphilic because an alkaline pH helps to remove the initial breakdown products of pectin. A *Geobacillus thermoglucosidans* strain PB94A was isolated at 65°C which could grow on hemp pectin and did not have cellulose activity, which is important in order not to degrade the fibres. Using this organism, finer fibres of higher quality could be produced than is possible with conventional processes (Valladares Juárez et al., 2009). This is an elegant example of the possibilities offered by extremophiles for creating more sustainable and efficient processes.

DEVELOPMENT OF GENETIC TOOLS FOR EXTREMOPHILES

Several extremophiles naturally produce building block chemicals or polymers in an enantiomeric pure form. Some organisms do this in a homofermentative way, meaning that they produce one product with hardly any by-products. Heterofermentatives on the other hand produce a mixture of several organic acids and/or alcohols. To make microbial production processes economically viable, it is necessary to optimize titres and yields by metabolic engineering. In the case of heterofermentatives, metabolic engineering can reduce by-product formation, thereby lowering downstream processing costs and increasing product yields. For homofermentatives, mainly lactate producers, fluxes can be increased to achieve higher titres, the enantio-specificity can be changed, or fluxes can be redirected to alternative products. Altogether, the development of genetic tools is a crucial step towards creating a versatile extremophilic platform organism for the production of green chemicals. Genetically accessible extremophilic organisms that have potential to serve as such platform organisms (Table 3) are mainly low G+C, thermophilic, Gram positive (facultative) anaerobes belonging to the phylum *Firmicutes* which were described in the previous sections. The development of competency methods, gene transfer systems and genetic tools for thermophilic ethanol-producing bacteria has recently been reviewed by Taylor *et al.* (Taylor et al., 2011). Many of the organisms described by Taylor *et al.* are polysaccharide-utilizing organisms mentioned also in the previous section of this review. These organisms can potentially be used in CBP-processes, making it very interesting to modify these organisms

to produce chemicals in a CBP-process. Here, we will focus on the most recent examples and also include chemical-producing organisms which we described in previous sections and that are mainly suitable in SSF-processes. Since genome sequencing is an important tool to establish a rational engineering strategy, to enable identification of mutations after laboratory evolution via re-sequencing and to simplify cloning genomic regions of interest, an overview of all currently sequenced extremophilic organisms relevant for the production of green chemicals or fuels is shown in Table 2.

Transformation methods

To start metabolic engineering, a transformation system is required, for which there are many options such as electroporation, natural transformation and protoplast transformation. The first issue hampering genetic tool development is that genetic accessibility appears to be not only species-specific, but sometimes even strain-specific. This often relates to the presence of strain-specific defence systems that prevent the entry of foreign DNA such as restriction modification (RM)-systems or CRISPR-Cas systems (Tock & Dryden, 2005; Westra et al., 2012). Strain-specificity makes it hard to copy genetic tools from a strain that is shown to be genetically accessible to a strain that might be a better production host. Although defence systems may hamper genetic tool development, they might on the other hand be important for acquiring phage resistance in an industrial setting. An example of the successful circumvention of RM-systems in bacilli is the protoplast transformation with *in vivo* methylated plasmid DNA in *Bacillus* sp. N16-5, which is a facultative alkaliphile and secretes organic acids. It was found to possess a Type II restriction endonuclease (REase) system. When a plasmid containing the Type II methyl-transferase (MTase) of *Bacillus* sp. N16-5 was cotransformed to *E. coli* together with the expression vector with the gene of interest, the transformation efficiency increased with a factor ten. Together with an optimized protoplast transformation protocol, the maximum efficiency was 1.1×10^5 transformed colonies per µg plasmid DNA (CFU) (Gao et al., 2011). For comparison, electrotransformation of *E. coli* can yield a CFU up till 5×10^{10} (Hanahan et al., 1991; Xue et al., 1999). RM was also found to constitute a barrier to transformation in *Caldicellulosiruptor bescii*, an organism that might be used in CBP-processes. *In vitro* methylation of DNA prior to transformation to *E. coli* with a unique α-class N4-Cytosine methyltransferase from *C. bescii* was required for transformation (Chung et al., 2012). In contrast, in *B. coagulans* it was shown that *in vitro* methylation or changing the source-host of DNA did not increase transformation efficiencies (Rhee et al., 2007; van Kranenburg et al., 2007). In *C. thermocellum* effects of factors affecting transformation efficiency such as methylation and cell-wall-weakening agents were strain-specific (Tyurin et al., 2004). For this species, an electroporation method was developed using a custom-built electroporator. For three different strains, efficiencies of up to 5×10^4 transformants per µg DNA were achieved (Tyurin et al., 2004).

Table 3. Genetically accessible extremophiles relevant for the production of green chemicals

Organism	Available genotypes	Transformation method
B. coagulans P4-102B	Δ*ldhL* Δ*alsS*	Electroporation
B. coagulans DSM 1^{T}	Δ*sigF*	Electroporation with Ts plasmid
B. licheniformis BL5/8	Δ*ldhL*	Protoplast transformation with Ts plasmid
Bacillus sp. N16-5	EV only	Protoplast transformation + *in vivo* methylation
M. thermoacetica ATCC 39073	*ldh**	Electroporation
G. thermoglucosidasius NCIMB 11955 & DL33	Δ*ldhL* Δ*pflB pdh**	Electroporation
C. bescii DSM 6725^{T}	Δ*pyrF*	Electroporation + *in vitro* methylation
C. thermocellum DSM 1313	Δ*pta*	Electroporation, custom-built
C. thermocellum DSM 1313	Δ*pta* Δ*ldhL*	Electroporation, custom-built
T. mathranii BG1	Δ*ldhL gldA**, Δ*ldhL adhE**	Electroporation, custom-built
Thermoanaerobacter & *Thermoanaerobacterium spp.*	several	Natural competence
T. aotearoense SCUT27	Δ*ldhL*	Electroporation of protoplasts
T. saccharolyticum JW/SL-YS485	Δ*ldhL* Δ*ackA* Δ*pta*	Electroporation
T. tengongensis CGMCC1.2430	Δ*hisG*	Natural transformation with Ts plasmid
T. ethanolicus JW200	EV only	Electroporation of protoplasts

n.a.: not applicable; EV: empty vector; *: overexpression; Ts: thermosensitive.

and fuels.

Integrant selection method	Current product	Ref.
Screening	D-lactate	(Wang et al., 2011a)
Cre-*lox* system	L-lactate	(Kovacs et al., 2010)
Screening	D-2,3-butanediol	(Wang et al., 2012b)
n.a.	Organic acids	(Gao et al., 2011)
pyrF-based counter-selection	Acetate/ L-lactate	(Kita et al., 2012)
Screening	Ethanol	(Cripps et al., 2009)
pyrF-based counter-selection	Ethanol	(Chung et al., 2012)
pyrF-based counter-selection + antibiotic cassette insertion	Ethanol	(Tripathi et al., 2010)
hpt and *thyA/tdk* – based counter-selection	Ethanol	(Argyros et al., 2011)
Linear (non-replicating) DNA + antibiotic cassette insertion	Ethanol	(Yao & Mikkelsen, 2010b), (Yao & Mikkelsen, 2010a)
	Ethanol	(Shaw et al., 2010)
Screening (antibiotic cassette insertion)	Ethanol	(Cai et al., 2011)
Screening (antibiotic cassette insertion)	Ethanol	(Shaw et al., 2008)
Antibiotic cassette insertion, non-replicating plasmid (only *E. coli* ori)	Ethanol	(Liu et al., 2012)
n.a.	Ethanol	(Peng et al., 2006)

Genetic accessibility might be taken into account when isolating novel green chemical producers, even though this is a highly laborious procedure. An example of screening for genetic accessibility after isolation of novel production organisms has been shown by Patel *et al.* who isolated 380 *B. coagulans* strains for lactate production and found only one of their isolates being transformable with plasmid DNA by electroporation (Patel et al., 2006). An *E. coli/Bacillus* shuttle vector and an electroporation method were developed for the genetically accessible *B. coagulans* P4-102B strain isolated by Patel *et al.*, yielding a maximum of 1×10^4 colonies per µg DNA (Rhee et al., 2007). Optimal electroporation settings appeared to be plasmids-specific, while natural and chemical competence did not yield any transformants (Rhee et al., 2007). Subsequently, this strain was engineered to produce D-lactate instead of L-lactate, which was achieved by a combination of rational metabolic engineering and adaptive evolution as described earlier (Wang et al., 2011a). The development of shuttle vectors is often required since most extremophiles show relatively low transformation efficiencies, which disables the use of PCR or ligation products for direct transformation and therefore the constructs should first be transformed into *E. coli*. An extensive overview of requirements for thermophilic vectors has recently been published (Taylor et al., 2011). In our own work, we are evaluating the genetic accessibility of thermophilic *Bacillus* and *Geobacillus* species using shuttle vectors and electroporation. In line with the findings described above, we observe that genetic accessibility is highly strain-specific. Most of the tested strains are not genetically accessible using the protocols available in literature, although a slight change in electroporation settings can make a large difference and this also appears to be strain-specific, which makes it hard to screen many isolates (unpublished data). A shuttle vector that appears to be suitable for many strains is pNW33n, which is also used for electroporation of for example *C. thermocellum* (Olson & Lynd, 2012) and *G. thermoglucosidans* (Taylor et al., 2008).

Wang *et al.* tested genetic accessibility for thermophilic, lactate-producing *B. licheniformis* BL3 and found the electroporation method of Rhee *et al.* for *B. coagulans* and several other competency methods not to yield any transformants. When using protoplast transformation they initially found one or two colonies per µg DNA. Optimization experiments showed that the best choice of regeneration medium was different for different plasmids. Finally, they achieved 25 transformants per µg DNA using an *E. coli/Bacillus* shuttle vector (Wang et al., 2012b).

A number of *B. licheniformis* strains are known to be naturally competent (Gwinn & Thorne, 1964), and lately, two strains were shown to possess genes to develop competence (Kovács et al., 2009), but natural competence has not been shown for strain BL3. Recently, several examples of induced natural competence in mesophilic *Bacillus* species have been published (Hoffmann et al., 2010; Mirończuk et al., 2008). In mesophilic *B. licheniformis* DSM 13, *comP* was found to be inactivated by an insertion element. Highly

competent strains could be obtained when the coding region of *comP* was replaced with a functional copy in combination with inducible expression of *comK*, the key regulator for the transcription of competence genes (Hoffmann et al., 2010). *B. cereus* was shown to become competent when the *B. subtilis comK* gene was heterologously expressed (Mirończuk et al., 2008). Natural competence in *B. cereus* had not been shown before, even though most genes required for competence could be identified (Mirończuk et al., 2008). Also in thermophilic *B. coagulans* DSM 1 natural competence genes were identified (Kovács et al., 2009), but even though its *comK* was shown to recognize several *comK*-responsive elements, overexpression of *comK* did not induce reproducible DNA-uptake (Kovács et al., 2013). On the other hand, natural competence has been shown to be an efficient way of DNA-uptake in thermophilic *Thermoanaerobacterium* and *Thermoanaerobacter* species (Shaw et al., 2010), which are mainly hydrogen and ethanol producers but also produce significant amounts of organic acids. Natural competence in *Thermoanaerobacterium* and *Thermoanaerobacter* was discovered because a non-pulsed control during an electroporation experiment yielded more transformants than the pulsed ones. It was then found that the only step required for transformation is a cell outgrowth period, which was confirmed using a natural transformation protocol in which DNA was added during different growth stages, after which cells were allowed to grow for two more hours before plating on selective medium. Cells were found most competent during exponential growth and did not require any external factors to develop competency. Importantly, using this protocol transformation with linear DNA containing regions homologous to the genome was also possible with an efficiency of 5.7×10^3 in *T. saccharolyticum* JW/SL-YS485, which makes integration into the genome significantly easier (Shaw et al., 2010). By now, a number of different transformation protocols and genetic tools have been developed for several *Thermoanaerobacter* and *Thermoanaerobacterium* species (Cai et al., 2011; Desai et al., 2004; Kita et al., 2012; Liu et al., 2012; Mai et al., 1997; Peng et al., 2006; Shaw et al., 2011; Shaw et al., 2010; Shaw et al., 2008; Yao & Mikkelsen, 2010b) (Table 2).

Integration methods

Integration of DNA into the genome is required for stable gene insertion or deletion. Whether integrants can easily be detected is determined by the combination of transformation and recombination efficiencies. The transformation efficiency determines the number of cells that have taken up DNA, whereas the recombination efficiency determines the number of integration events into the genome. The use of non-replicative plasmids, linear DNA fragments or ligation products allows for selection of direct integration into the genome and simplifies screening for integrants. However, this is generally only possible when transformation efficiencies are high (Biswas et al., 1993). Since extremophiles mostly show relatively low transformation efficiencies, only replicating plasmids can be used, which makes

screening for integration events laborious if recombination efficiencies are not exceptionally high. A method to overcome the low transformation/recombination efficiency barrier while still enabling the use of a non-replicating plasmid is using an integration vector with a conditional replicon, such as a thermosensitive plasmid. The replicon of such a plasmid is not stable at high temperatures, thereby disabling plasmid replication at elevated temperatures (Olson & Lynd, 2012). This can be a powerful tool especially for thermophiles since they usually have a large temperature range. For example, a thermosensitive integration plasmid is transformed to *B. coagulans* and plated below 50°C in order to enable screening for transformants (Kovacs et al., 2010; van Kranenburg et al., 2007; Wang et al., 2011a). Subsequently, the temperature is increased to 50°C or higher, disabling plasmid replication and enabling selection for integration events into the genome via Campbell recombination over the homologous regions cloned into the integration plasmid (Kovacs et al., 2010; van Kranenburg et al., 2007; Wang et al., 2011a). Also Wang *et al.* used a thermosensitive plasmid for constructing their D-2,3-BDO-producing *B. licheniformis*. With only 25 transformants per µg DNA this strain showed very low efficiencies, but using a thermosensitive plasmid nevertheless enabled gene deletions (Wang et al., 2012b). Recently, a computer algorithm was used to predict temperature-sensitive variants of the often-used *E. coli/Geobacillus* shuttle vector pNW33n, so it can be used for easier integration in *C. thermocellum*, which grows at 45-62°C (Olson & Lynd, 2012). The development of such an algorithm might be extended to plasmids for other extremophilic species as well.

In some thermophiles, the combined transformation and recombination efficiencies are high enough to be able to transform the strains with non-replicating DNA, simplifying integration. In *C. bescii* it was shown to be possible to integrate non-replicating plasmids into the genome despite a low transformation efficiency of 50 transformants per µg DNA, although the authors indicate this might be an underestimation due to low plating efficiencies of the strain (Chung et al., 2012). *Thermoanaerobacter mathranii* BG1 can be transformed with linearized, non-replicating plasmid DNA (Yao & Mikkelsen, 2010a; Yao & Mikkelsen, 2010b). This is also possible in *Thermoanaerobacterium aotearoense*, which was transformed with linearized, non-replicating plasmid DNA via electroporation of protoplasts (Cai et al., 2011; Mai et al., 1997). In all these three cases, an antibiotic resistance cassette flanked by homologous regions of the gene of interest was cloned into the plasmid/fragment. Because the plasmid was unable to replicate, double crossover knockout mutants could readily be obtained due to insertion of the antibiotic cassette in the target gene. The drawback of these mutants is that the antibiotic cassette is still inserted, requiring further development of a markerless gene deletion method. In many studies, the gene of interest is replaced by an antibiotic cassette. Although in the case of transformation with linear DNA this method simplifies screening for double-crossover mutants, it disables the re-use of antibiotic markers for further research and the acquired antibiotic resistance is highly undesired in an

industrial organism. Since the number of markers that can be used at high temperatures is limited (Taylor et al., 2011), methods enabling the re-use of markers are especially important for use in thermophiles.

Although it is possible to create mutants by homologous recombination without replacing the gene of interest with an antibiotic resistance cassette, this may require more screening. Also, when transformation with linear DNA is not possible and a (thermosensitive) plasmid is used, screening for double crossovers might be very laborious due to the lack of pressure for double crossover events. An example is the creation of a Δ*ldhL* Δ*alsS* mutant of *B. coagulans* QZ19 (originally P4-102B) for D-lactate production (Wang et al., 2011a). After obtaining single crossover integration of a thermosensitive plasmid conveying antibiotic resistance, further culturing without antibiotic pressure generally yields both wild type revertants and the desired double crossover knockout mutants. However, since there is no positive selection for the double crossover and plasmid removal, and the wild-type may have a growth advantage over the knockout, obtaining such a knockout may be very laborious (Wang et al., 2011a). Another *Bacillus*-related organism for which similar genetic tools are now firmly established, is the facultatively anaerobic thermophile *Geobacillus thermoglucosidans*. This organism is capable of converting a wide range of substrates into ethanol, formate, lactate and acetate. Taylor *et al.* set up a shuttle vector-based electroporation method for this species (Taylor et al., 2008), which has been successfully used for metabolic engineering of two strains of this species by Cripps *et al.* to improve ethanol production (Cripps et al., 2009). By using a thermosensitive plasmid and homologous recombination without leaving antibiotic markers the *ldhL* and *pflB* genes were inactivated. Also, the same integration system was used to up-regulate the expression of pyruvate dehydrogenase by replacing the promoter (Cripps et al., 2009).

Counter-selection systems

As indicated above, methods are required to easily obtain the second recombination and cure the plasmid after obtaining a single crossover with a (thermosensitive) plasmid. Such methods are known as counter-selection tools, in which there are positive selection methods for both insertion of the plasmid via a single crossover and its subsequent removal by double crossover. For current platform organisms like *E. coli*, such systems are well established, but for extremophiles they are just emerging.

An example of a counter-selection system is the *pyrF*-system (Figure 2A). The *pyrF* gene encodes the enzyme orotidine 5'-phosphate decarboxylase, which is required for a step in pyrimidine biosynthesis, *i.e.* the conversion of orotidine 5'-phosphate to UMP and CO_2. In addition, PyrF catalyses the conversion of the unnatural 5-fluoroorotic acid (5-FOA) to 5-fluorouridine 5'-monophosphate, which is further metabolized to toxic 5-fluorodeoxyuridine 5'-monophosphate (Suzuki et al., 2012). In the presence of uracil, the *pyrF* gene can

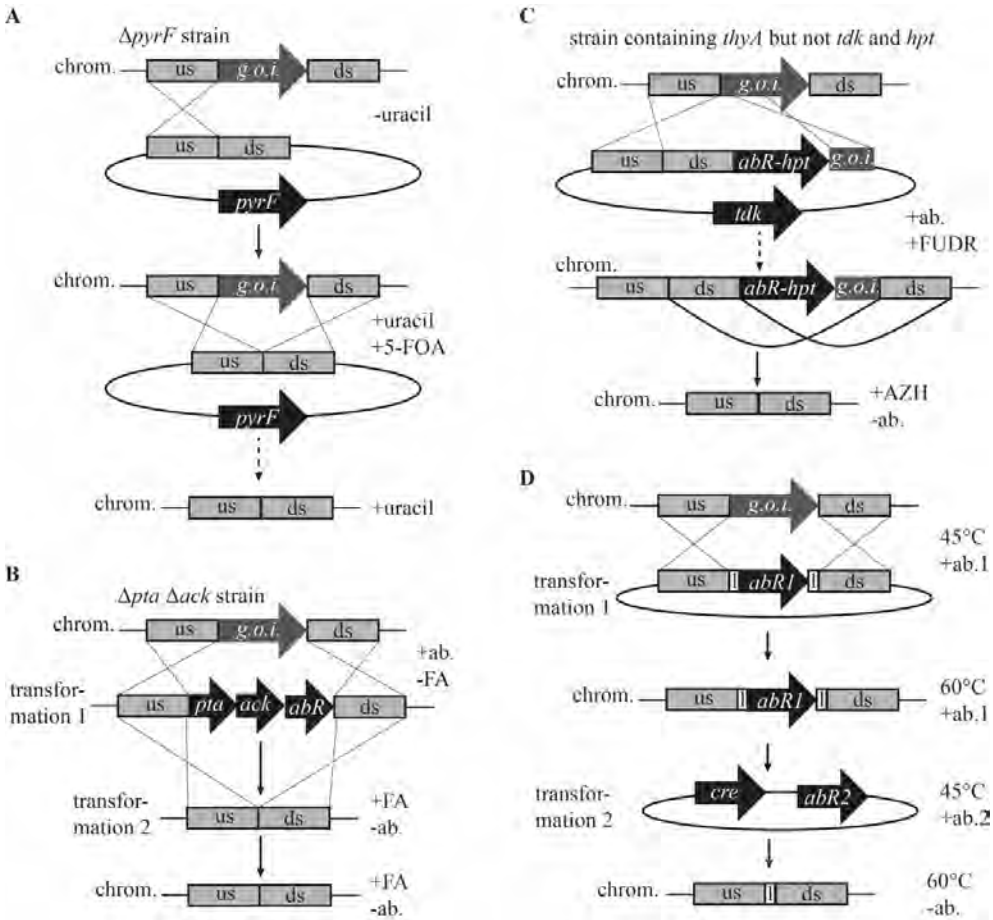

Figure 2. Overview of counter-selection and antibiotic-cassette removal systems used in thermophiles for targeted double crossovers. Abbreviations: chrom.: chromosome; us: upstream flanking region; ds: downstream flanking region; g.o.i.: gene of interest; ab.: antibiotic; abR: antibiotic-resistance gene. **A:** *pyrF*-system. After insertion of the *pyrF*-containing plasmid into a Δ*pyrF* strain, single crossovers can be obtained by selecting for restoration of the uracil auxotrophy. Double crossovers are obtained by adding 5-fluoroorotic acid (5-FOA), which will be toxic for strains still harbouring the *pyrF*-plasmid, and by adding uracil to select for strains being uracil auxotrophs due to lack of *pyrF*. Based on (Kita et al., 2012; Suzuki et al., 2012) **B:** *pta-ack*-system. In this case, linear DNA fragments were used for transformation, which allows simple and direct screening for double crossovers, but the system can also be used with plasmid DNA. First, the *pta* and *ack* genes for acetate formation need to be deleted, resulting in a strain resistant against haloacetate compounds such as fluoroacetate (FA). This strain is then transformed with DNA containing *pta*, *ack* and an antibiotic-resistance gene, which replace the gene of interest via the flanking regions. Successful double crossover and thus insertion of the three genes is detected by antibiotic resistance. In a second transformation, the strain is transformed with fused flanking regions of the gene of interest and cultured on FA, thereby deleting the *pta*, *ack* and an antibiotic-resistance gene since FA is toxic when *pta* and *ack* are present. Based on (Shaw et al., 2011). **C:** *tdk-hpt* system. After selecting transformants based on antibiotic-resistance, double crossovers can be detected by growing the culture on antibiotics and fluoro-deoxyuracil (FUDR), to which the strains will be resistant due to loss of *tdk* (which together with the native *thyA* makes the strain sensitive to FUDR) after they have recombined and cured the plasmid. Subsequently, clean knockouts created by a second crossover event can be detected by culturing on 8-azahypoxannthine (AZH), to which strains will be resistant when they have recombined and lost the *hpt* and antibiotic-resistance gene. Based on (Argyros et al., 2011). **D:** Cre-*lox* system. Using a thermosensitive plasmid, in a first step the gene of interest is replaced by an antibiotic resistance gene flanked by lox sites, enabling selection for antibiotic resistance. In a second transformation, a second thermosensitive plasmid coding for the Cre-recombinase and conveying a different antibiotic resistance is inserted. The Cre-recombinase recombines the *lox66* and *lox71* sites to a *lox72* site, thereby cleaving out the first antibiotic-resistance gene. Based on (Kovacs et al., 2010).

easily be inactivated by growing the target strain on toxic concentrations of 5-FOA, thereby selecting for spontaneous or targeted *pyrF*-mutants which will be resistant to 5-FOA and auxotrophic for uracil. Expression of the *pyrF* gene on a plasmid in the absence of uracil can then be used as a selection marker for plasmid uptake or insertion based on restoration of uracil prototrophy, and as a subsequent counter-selection for double crossover and plasmid loss based on 5-FOA resistance (in the presence of uracil) (Tripathi et al., 2010). The *pyrF*-system has been used in many organisms, including the thermophilic bacteria *Moorella thermoacetica*, *T. saccharolyticum*, *Geobacillus kaustophilus* and *C. thermocellum* (Kita et al., 2012; Shaw et al., 2011; Suzuki et al., 2012; Tripathi et al., 2010).

Although the *pyrF*-system simplifies the detection of double crossovers due to the counter-selection of plasmid loss, the ratio between wild type and mutant double crossovers might still be an issue in the case of essential genes that are hard to delete due to evolutionary pressure. In those cases, screening systems might be needed to discriminate mutants from wild type. An example of such a screening system is the culturing of potential *ldhL*-mutants on $CaCO_3$-containing plates, on which mutants that produce less acids due to *ldhL*-disruption will form a smaller clearing zone around the colony because less $CaCO_3$ is converted to calcium-lactate (Wang et al., 2011a). However, such a system cannot be used in most cases and therefore, many studies so far still use insertion of an antibiotic cassette to simplify discrimination between wild type and mutant double crossovers. Another elegant way of overcoming this problem was shown by Shaw *et al.* (Shaw et al., 2011) (Figure 2B). Natural transformation with linear DNA fragments was used for transformation, enabling direct selection for double crossovers. Instead of the *pyrF* gene deletion combined with 5-FOA resistance, the *pta* and *ack* genes were deleted, resulting in strains that are resistant against haloacetate compounds such as fluoroacetate (FA) and that do no longer produce acetate. Subsequently, the Δ*pta-ack* strain was transformed with DNA containing the *pta* and *ack* genes together with a kanamycin resistance cassette that is flanked by fragments homologous to the *ldhL* gene. Due to the use of linear DNA fragments, double recombinants can directly be identified by kanamycin resistance and FA sensitivity. Next, these kanamycin-resistant mutants were transformed with linear fragments containing fused flanking regions of the *ldhL*-gene, thereby removing the kanamycin resistance and *pta* and *ack* genes, yielding FA-resistant strains (Shaw et al., 2011). Like the *pyrF*-system, this method is based on the selection of strains resistant to other compounds than antibiotics, using double selections. However, the reason why the *pta-ack* system yields final mutants without antibiotic markers is that two rounds of recombination around the same gene are used. After the first round positive selection for kanamycin resistance is performed and in the second round, when antibiotic resistance is no longer present, positive selection is based on FA-resistance due to the removal of *pta* and *ack* together with the kanamycin cassette. The same result could be obtained with the *pyrF*-system if the same two-round recombination

would be used. The initial knockout of *pta* and *ack* does not use any antibiotic resistance but is purely based on FA-resistance of *pta-ack* knockouts. In the case of the work of Shaw *et al.* this was enabled by the possibility to transform with linear DNA, but in other cases screening for random mutations could yield the same knockouts.

The *pyrF*-counter-selection system creates strains that are uracil auxotroph and thus essentially have a growth-defect which needs to be restored, creating an extra step in the strain development. The same holds true for the *pta-ack* system, unless this is used in a strain that should be deficient in acetate production. Therefore, efforts were successfully made to establish a markerless gene deletion method for *C. thermocellum* and *T. saccharolyticum* that did not require any growth-affecting deletions such as *pyrF* (Figure 2C). Instead of *pyrF*, *hpt* was deleted, which is a gene coding for a purine-reassimilating protein that can lead to cellular toxicity in the presence of the purine analogue 8-azahypoxanthine (AZH) (Argyros et al., 2011). If *hpt* is absent, this does not lead to a growth defect but eliminates AZH-toxicity, thereby creating a counter-selection tool. The second part of the selection tool is also based on nucleic acid metabolism by using the toxicity of fluoro-deoxyuracil (FUDR) in the presence of the enzymes thymidine kinase (encoded by *tdk*) and thymidylate synthetase (encoded by *thyA*). Tdk converts FUDR to fluoro-dUMP, which inhibits ThyA (Argyros et al., 2011). A counter-selection tool was established by combining *C. thermocellum* native *thyA* with *T. saccharolyticum tdk* expression, which together make the strain sensitive to fluoro-deoxyuracil (FUDR). To make mutants, transformants were obtained by screening for antibiotic resistance encoded on the plasmid. Subsequently, the screen for integration and plasmid curing was based on FUDR-resistance since *tdk* is encoded by the plasmid and is lost after the first crossover events. In a final step, AZH-resistant colonies were selected that lost the antibiotic-*hpt* cassette and thus showed a clean deletion genotype. This resulted in a very easy selection of mutants (Argyros et al., 2011). By selecting for two double crossover events, this system eliminates the possibility of obtaining wild type revertants. Furthermore, this system has the advantage of not leaving any antibiotic resistance.

Well-known systems to remove inserted antibiotic cassettes are for example the Cre-*lox* and FLP/*FRT*-system. Recently, a Cre-*lox* system was developed for thermophilic *B. coagulans* DSM 1 (Figure 2D). In this case, the system was used to delete the *sigF* gene, creating a sporulation-deficient strain, which can be desirable for safety reasons when genetically modified organisms are used (Kovacs et al., 2010). Using a thermosensitive plasmid, in a first step the *sigF* gene was replaced by a *lox66-cat-lox71* cassette via homologous recombination over the flanking regions of *sigF* which were cloned next to the *lox66-cat-lox71* cassette, conveying chloramphenicol resistance via the *cat* gene. In a second step, another plasmid coding for the Cre-recombinase and harbouring tetracycline resistance was transformed to a colony with a double crossover replacement of *sigF* by *lox66-cat-lox71*.

The Cre-recombinase recombines the *lox66* and *lox71* sites to a *lox72* site, which is not recognized anymore by Cre-recombinase, thereby removing the *cat* gene and enabling the re-use of the same marker. Subsequently, the same system with the same markers was used to delete the *lacZ* gene from the *B. coagulans* DSM 1 Δ*sigF* genome in order to use this gene as a reporter for promoter activities (Kovacs et al., 2010). The Cre-*lox* system is a highly useful tool since it leaves no antibiotic resistance markers. It is not entirely markerless since it leaves a small *lox72* site, but since this site is different from the starting *lox66* and *lox71* and is no longer recognized by Cre, this does not interfere with subsequent integration events using the same system. Only prerequisites are that at least two different antibiotic resistance markers should be available for the target host, that the plasmids used can easily be cured, and that the Cre-recombinase should be functionally produced.

Our own experience with creating double crossovers in thermophilic *Bacillus* species is that recombination is generally fast and efficient (unpublished data). However, a major problem is that often single colonies contain mixed genotypes of single crossovers over both flanking regions, wild type and knockout. This emphasises the importance of an efficient counter selection system in order to cure the plasmid and eliminate single crossovers. Methods that can also separate knockout and wild type genotypes are even more preferred, such as the *hpt-tdk*-system described above.

CONCLUSIONS

Extremophiles, especially thermophiles, are an extremely interesting and promising group of microbes to be used as production hosts for green bulk chemicals. Their versatile metabolic capacities to utilize a wide range of substrates makes them very suitable for next generation processes in which non-food biomass is used. Furthermore, their ability to thrive under extreme conditions offers advantages such as a lowered contamination risk and lower production costs due to eliminating the need for sterilization. Thermophiles offer also other advantages such as lower enzyme costs in SSF settings and more efficient hydrolysis. Although many of the extremophiles require less complex nitrogen sources than traditionally used mesophiles, adaptive evolution experiments might further decrease the requirement for complex nitrogen sources to make the processes more economically viable. Isolation and screening studies can yield novel interesting extremophilic organisms. The exploration of the potential of extremophiles for chemical production has only just emerged, and more studies are needed to achieve higher productivities, titres and yields. An important requirement to make such improvements is the development of genetic tools. For several extremophiles such tools are already available, but these are mostly used for optimization of ethanol production. Engineering these organisms for the production of chemicals would be an important next step.

CONFLICT OF INTEREST

The authors declare to have no conflict of interest. RvK is employed by the commercial company Corbion (Gorinchem, The Netherlands).

ACKNOWLEDGEMENTS

The work of EFB is supported by Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

Chapter 3

Isolation and screening of thermophilic bacilli from compost for electrotransformation and fermentation: Characterization of *Bacillus smithii* ET 138 as a new biocatalyst

This chapter has been published as:
Bosma, E.F., van de Weijer, A.H.P., Daas, M.J.A., van der Oost, J., de Vos, W.M., van Kranenburg, R. 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), 1874-1883.

ABSTRACT

Thermophilic bacteria are regarded as attractive production organisms for cost-efficient conversion of renewable resources to green chemicals, but their genetic accessibility is a major bottleneck in developing them into versatile platform organisms. In this study, we aimed to isolate thermophilic, facultatively anaerobic bacilli that are genetically accessible and have potential as platform organisms. From compost, we isolated 267 strains that produced acids from C_5 and C_6 sugars at temperatures of 55°C or 65°C. Subsequently, 44 strains that showed the highest production of acids were screened for genetic accessibility by electroporation. Two *Geobacillus thermodenitrificans* isolates and one *Bacillus smithii* isolate were found to be transformable with plasmid pNW33n. Of these, *B. smithii* ET 138 was the best-performing strain in lab-scale fermentations and was capable of producing organic acids from glucose as well as from xylose. It is an acidotolerant strain able to produce organic acids until a lower limit of approximately pH 4.5. As genetic accessibility of *B. smithii* had not been described previously, six other *B. smithii* strains from the DSMZ culture collection were tested for electroporation efficiencies and we found the type strain DSM 4216^T and strain DSM 460 to be transformable. The transformation protocol for *B. smithii* isolate ET 138 was optimized to obtain approximately $5x10^3$ colonies per µg plasmid pNW33n. Genetic accessibility combined with robust acid production capacities on C_5 and C_6 sugars at a relatively broad pH range make *B. smithii* ET 138 an attractive biocatalyst for production of lactic acid and, potentially, other green chemicals.

KEYWORDS

thermophilic bacilli, electroporation, transformation, isolation, strain selection, biotechnology, green chemicals, *Bacillus smithii*, genetic accessibility, compost, screening, organic acids

INTRODUCTION

Green chemicals are sustainable bio-based alternatives for chemicals based on fossil resources. Biomass is considered an attractive renewable resource for the production of such green chemicals. Major challenges for using biomass are to develop cost-effective production processes and to convert substrates that go beyond first-generation pure sugars. In this perspective, the use of microbial fermentation processes that convert lignocellulosic sugars is gaining considerable attention. For particular products natural producers can be used, but often *E. coli* and *S. cerevisiae* are used as platform organisms, because of their well-known physiology and ease of engineering for the production of many different products (Chen et al., 2013; Hong & Nielsen, 2012).

In general, most organisms used for the production of green chemicals and fuels are mesophiles and current processes are still mainly based on first generation feedstocks, *i.e.* sucrose from sugar beet or sugarcane or glucose derived from starches from corn or tapioca. Although engineering of platform organisms broadens the spectrum of possible substrates (Buschke et al., 2012), the efficiency of the process could be increased by the use of organisms naturally capable of degrading lignocellulose-derived sugars (Blumer-Schuette et al., 2013). To further increase the efficiency and reduce the costs of the microbial production of green chemicals, moderately thermophilic hosts offer several advantages as compared to mesophilic hosts, including i) reduced cooling costs, ii) lower contamination risk, iii) increased substrate and product solubility and iv) temperature optima of these bacteria matching those of enzymes used for simultaneous saccharification and fermentation (SSF), lowering the enzyme load and costs (Ou et al., 2009). The use of anaerobic or facultatively anaerobic bacteria eliminates the need for expensive aerated industrial reactors. For SSF, facultative anaerobes have the advantage over strict anaerobes that reduction of the medium is not necessary, as reduced medium has been shown to inhibit saccharolytic enzymes (Podkaminer et al., 2012).

Altogether, organisms of particular interest for the production of green chemicals from lignocellulose in an SSF setting are (facultatively) anaerobic thermophiles that can grow in minimal medium and ferment a wide range of substrates including C_5 and C_6 sugars. They can potentially be used as platform organisms if they are genetically accessible. A number of facultatively anaerobic thermophilic hosts has been studied for green chemical and fuel production (reviewed in (Bosma et al., 2013; Taylor et al., 2009)), such as *Bacillus coagulans* for lactic acid (Ma et al., 2014; Wang et al., 2011a), *Bacillus licheniformis* for 2,3-butanediol (Li et al., 2013; Wang et al., 2012b) and *Geobacillus thermoglucosidans* for ethanol production (Cripps et al., 2009). Examples of strictly anaerobic thermophiles that have been studied for biofuel production are *Clostridium thermocellum* (Shao et al., 2011; Veen et al., 2013), several *Caldicellulosiruptor* spp. (Chung et al., 2014; Svetlitchnyi et al., 2013; Wil-

lquist & van Niel, 2012) as well as *Thermoanaerobacterium* spp. and *Thermoanaerobacter* spp. (Bhandiwad et al., 2014; Saripan & Reungsang, 2013; Svetlitchnyi et al., 2013; Yao & Mikkelsen, 2010b). For each of these species, one or more strains have been shown to be genetically accessible and engineering tools have been developed (Bosma et al., 2013; Taylor et al., 2011). Still, it is desirable to further develop more efficient production organisms and improved genetic tools. On the one hand this will enable a better understanding and exploitation of the diversity in metabolism found in thermophilic organisms and on the other hand this will expand the genetic toolbox for thermophiles, which at the moment is still rather limited compared to those available for mesophilic platform organisms.

In this study we aimed to extend the collection of potential industrial platform organisms by isolating thermophilic, facultatively anaerobic bacilli and selecting a suitable production host by screening for C_5 and C_6 sugar utilization, organic acid production, fermentation performance and genetic accessibility in order to select a new biocatalyst that has the potential to be developed into a platform organism.

RESULTS

Isolation, initial selection and identification

Thermophiles able to grow on minimal medium (TMM) at 55°C or 65°C were obtained from compost (Table 1). Single colonies (842) from the dilution series of compost samples were transferred to plates containing different carbon sources (glucose, sucrose, xylose, or arabinose). All colonies grew on glucose, and 584 colonies also grew on one or both of the C_5 sugars xylose and arabinose at 55°C or 65°C. These 584 colonies were then inoculated into liquid TMM supplemented with a pH indicator under microaerobic conditions. Glycerol stocks were made of 267 strains that showed both growth and acidification in liquid culture on glucose (Table 1). To select the best acid-producing strains, all 267 isolates were inoculated in duplicate in TMM liquid medium supplemented with glucose, $CaCO_3$ and 0.5 g/L yeast extract and grown for 48 h under microaerobic conditions, after which HPLC analysis was performed. For each isolation condition, strains were found that either produced relatively high quantities of total products, or showed products that were rarely found among strains isolated under the same conditions. A total of 35 strains was selected from the five different isolation conditions (Table 1 and S1) to make pure cultures to screen for genetic accessibility. For ten isolates (ET 011, ET 072, ET 129, ET 130, ET 131, ET 144, ET 230, ET 239, ET 261, ET 267), purification revealed distinct morphologies and resulted in more than one pure culture from the same original colony (*e.g.* 144-1 and 144-2). The 44 pure cultures were identified by 16S rRNA gene sequence analysis (Figure 1). The majority of strains was

Table 1. Overview of the number of isolates per step for each isolation condition.

Isolation condition[a]	Strain (ET) nrs	Nr of colonies selected from dilution series and plated on 4 sugars	Nr of strains growing on at least 3 sugars[b] and inoculated into liquid medium	Nr of strains acidified and stocked	Nr of strains selected based on HPLC data and purified[c]
1: TMMscr 65°C	001-054	257	217	53	5
2: TMMscr+$CaCO_3$ 65°C	055-131	260	151	77	5
3: TMMscr+$CaCO_3$ 55°C	132-157	65	45	26	7
4: TMMglc+$CaCO_3$ 65°C	158-213	130	77	56	5
5: TMMglc+$CaCO_3$ 55°C	214-267	130	94	55	13
Total		842	584	267	35

[a] Abbreviations: scr: sucrose; glc: glucose.
[b] Tested sugars were D-glucose, D-xylose, L-arabinose and sucrose. To be selected, strains should grow on at least one of the C_5 sugars.
[c] The final 35 selected strains were purified, resulting in 44 pure cultures, which were subsequently tested for genetic accessibility.

identified as *Geobacillus thermodenitrificans*, followed by *Bacillus thermocopriae*, *Geobacillus caldoxylosilyticus*, *Bacillus coagulans* and *Aeribacillus pallidus*. One strain of *Bacillus smithii* was obtained. The highest species diversity was obtained with isolation condition 5, whereas this was the lowest in conditions 1 and 4, in which mainly *G. thermodenitrificans* were found. As expected, at 65°C mainly *Geobacillus* species were obtained, whereas at 55°C also *Bacillus* species were found and the species diversity was larger.

HPLC data for the tube test for the selected strains are shown in Table S1. Independent of the species, all strains produced mainly lactate, with acetate as the major by-product and succinate as a minor by-product. Gas production was not analyzed. All strains, except for *Bacillus smithii*, also produced small amounts of one or more other minor products such as ethanol, 2,3-butanediol (2,3-BDO), formate, propionate or malate. By-products varied not only between species, but also between strains of the same species. For example, *Geobacillus thermodenitrificans* ET 236 and ET 251 produced malate, ethanol, formate and 2,3-BDO, whereas *G. thermodenitrificans* ET 136 did not produce detectable amounts of malate and 2,3-BDO, but higher amounts of ethanol, and *G. thermodenitrificans* ET 241 did not produce any ethanol. Most *B. thermocopriae* strains produced propionate. The addition of $CaCO_3$ in the tubes might have enhanced production of C4-dicarboxylic acids, but these were never observed as major products. The highest titer of these products was reached by strain *G. caldoxylosilyticus* ET 208, with 0.58 g/L succinate and 0.18 g/L malate, together constituting 15% of its total production, with its main products being lactate and acetate. The total amount of products varied strongly from 1.27 g/L to 7.61 g/L. On average, strains

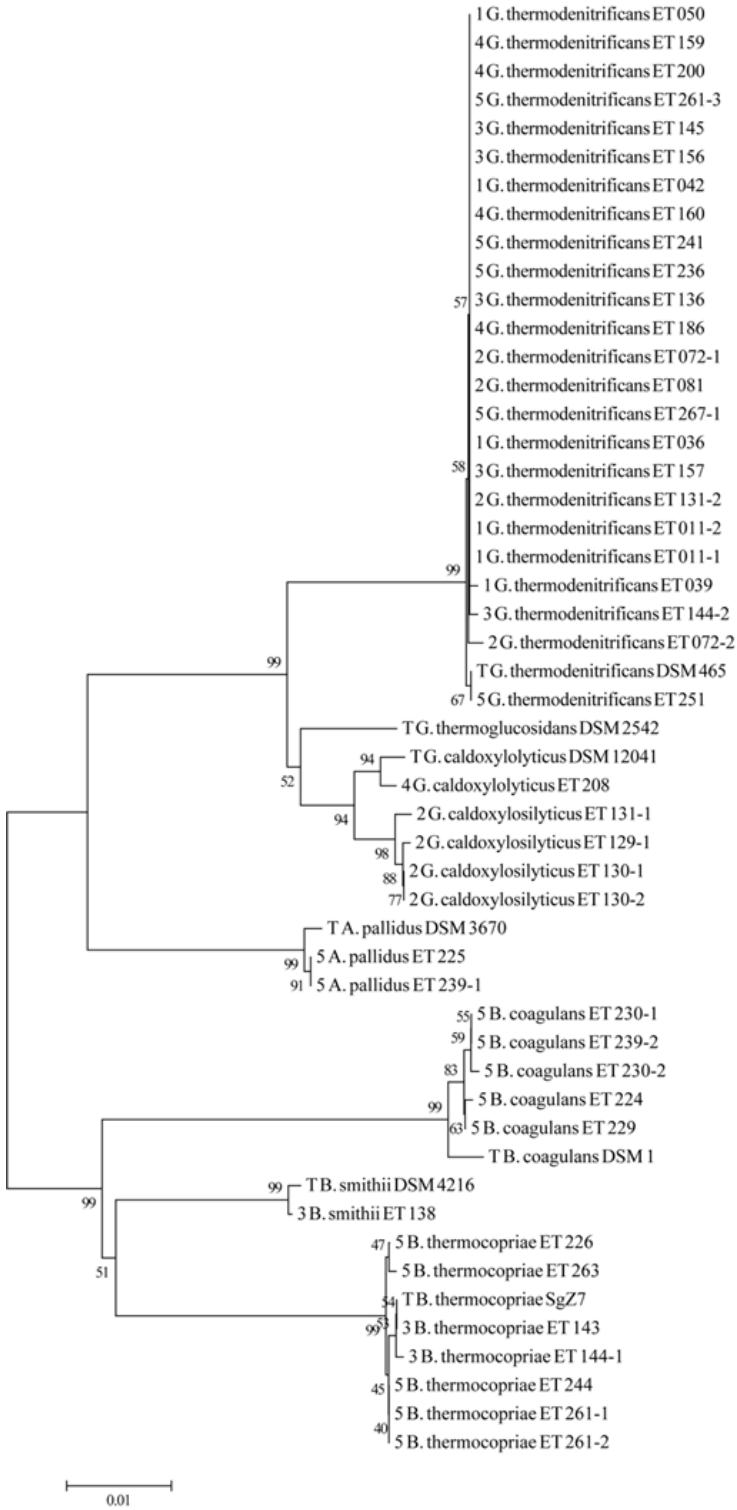

Figure 1. Neighbour-joining tree of 16S rRNA gene sequences of selected isolates and type strains. The first number indicates the isolation round, whereas 'T' indicates a type strain. Numbers at the nodes represent bootstrap values out of 1000 replicates.

isolated in condition 5 (glucose 55°C with $CaCO_3$) produced most acids, while strains from condition 2 (sucrose 65°C) produced the least and strains isolated at 55°C on average produced more acid compared to the 65°C-isolates. No difference was observed in average production between strains isolated on sucrose and glucose.

Screening for genetic accessibility by electrotransformation

The 44 selected pure cultures were tested for their transformation efficiency by electroporation. Depending on the species, slightly modified versions of either a *Bacillus* (Rhee et al., 2007) or a *Geobacillus* (Taylor et al., 2008) protocol were used to prepare competent cells. Subsequently, the obtained cell suspensions were electroporated using *E. coli* – (*Geo*) *bacillus* shuttle vector pNW33n. Strains that were successfully transformed with pNW33n were *B. smithii* strain ET 138 and *G. thermodenitrificans* strains ET 144-2 and ET 251. Transformation was confirmed by colony PCR and by isolating plasmid DNA from the transformed strains, which was subsequently confirmed to be pNW33n by restriction analysis of the isolated plasmid DNA from the transformants (*G. thermodenitrificans* strains ET 144-2 and ET 251) or from retransformed *E. coli* (*B. smithii* ET 138) (Figure S1). Transformation of *G. thermodenitrificans* strains ET 144-2 and ET 251 according to the *Geobacillus* protocol (with settings 2.0 kV, 25 µF and 200 Ω in a 2 mm cuvette) resulted in 6 and 4 colonies per µg DNA, respectively. *B. smithii* strain ET 138 made competent according to the *Bacillus* protocol and electroporation settings (Rhee et al., 2007) yielded 3 colonies per µg DNA, whereas the *Geobacillus* electroporation settings yielded 33 colonies per µg DNA for this strain. Subsequent retesting of other *Bacillus* isolates with the *Geobacillus* settings did not result in any transformants. For strains ET 144-2 and ET 251, transformation was reproducible but not always successful, whereas transformation for ET 138 was highly reproducible despite the low efficiency. For fourteen strains, occasionally colonies also appeared when no DNA was added during electroporation, while the strains did not grow on chloramphenicol prior to transformation (Table S1). None of the colonies formed by these strains tested positive in the PCR on the pNW33n *repB-cat*-gene combination. As transformation is crucial for the development of an organism into an efficient platform organism, only the three transformable strains were selected for further studies.

Fermentation and growth characteristics of transformable strains

Transformable isolates *B. smithii* ET 138 and *G. thermodenitrificans* ET 144-2 and ET 251 were tested in 1 L pH-controlled fermentations on glucose. Similar to the results of the acidifying tube tests, for all three strains, lactate was the major product and acetate the second most abundant product, with succinate as minor by-product (Table 2). For all three strains, ethanol production was observed in one of the two duplicate fermentations, while

this was not observed during the tube experiments for *B. smithii* ET 138 (Table S1). *G. thermodenitrificans* ET 251 showed very low titers and productivity in the reactor, namely a total of 6.58 ±1.35 g/L with 0.05 ±0.01 g/L/h (Table 2), even though it performed well in tubes (Table S1). *G. thermodenitrificans* ET 144-2 produced a total of 14.80 ±4.03 g/L products, but showed a rather low productivity of 0.08 ±0.01 g/L/h (Table 2). In contrast, *B. smithii* ET 138 outperformed the other strains both in product titer and productivity with 19.35 ±2.34 g/L at 0.14 ±0.01 g/L/h (Table 2). As it was also outperforming the other two isolates in transformation efficiency, *B. smithii* ET 138 was selected for further studies.

Table 2. Fermentation performance of transformable isolates.

Strain	Time (h)	Products (g/L)[a]						Av. prod. (g/L/h)[b]
		Lac	Ace	Suc	Mal	Eth	Total	
B. smithii ET 138	135 *±8*	18.15 *±3.03*	0.95 *±0.45*	0.09 *±0.00*	nd	0.16 *±0.22*	19.35 *±2.34*	0.14 *±0.01*
G. thermodenitrificans ET 144-2	190 *±70*	13.22 *±3.26*	1.01 *±0.16*	0.29 *±0.22*	0.06 ±0.08	0.21 *±0.30*	14.80 *±4.03*	0.08 *±0.01*
G. thermodenitrificans ET 251	132 *±11*	5.70 *±1.38*	0.68 *±0.05*	0.15 *±0.04*	nd	0.04 *±0.06*	6.58 *±1.35*	0.05 *±0.01*

The results shown are the averages of duplicates; standard deviations are indicated in italics after the '±'.
[a] Abbreviations: Lac: lactate, Ace: acetate, Suc: succinate, Mal: malate, Eth: ethanol, nd: not detected.
[b] Av. prod.: average productivity. Calculated by dividing total products by fermentation time.

Table 3. Fermentation of *B. smithii* ET 138 at different pHs.

pH	Time (h)	Products (g/L)[b]				Av. prod. (g/L/h)[c]
		Lac	Ace	Suc	Total	
4.5	43[a]	3.3	0.3	0.0	3.6	0.08
5.5	43	9.9	0.3	0.1	10.3	0.24
	141	14.0	0.6	0.1	14.7	0.10
6.5	43	10.5	0.3	0.0	10.8	0.25
	141	21.1	0.6	0.1	21.8	0.15

[a] At pH 4.5, no more acids were produced after 43 h so no further time points are shown.
[b] Abbreviations: Lac: lactate, Ace: acetate, Suc: succinate.
[c] Av. prod.: average productivity. Calculated by dividing total products by fermentation time.

B. smithii ET 138 showed growth at the tested temperatures between 37 and 63°C, with optimal growth between 51 and 57°C and no growth at 66°C or higher. The optimum pH was found to be 6.5, with little difference between pH values of 6.0-7.0 (data not shown). In further experiments, growth conditions were kept at 55°C and pH 6.5. The lowest initial pH value still supporting growth was 4.5. To further evaluate the ability of *B. smithii* ET 138 to ferment at low pH, fermentations in pH-controlled reactors were performed at pH 4.5, 5.5 and 6.5 (Table 3). During the first 43 h of fermentation, product titers and productivities at pH 5.5 and 6.5 were almost identical with 10.3 g/L at 0.24 g/L/h and 10.8 g/L at 0.25 g/L/h, respectively. After 141 h, both titer and productivity were 1.5-fold lower at pH 5.5 compared to pH 6.5 (Table 3). At pH 4.5 the strain was still able to grow and produce for 43 h, but titer and productivity were reduced to 3.6 g/L and 0.08 g/L/h. Hardly any acid was produced after 43 h, indicating severe acid stress (Table 3). At pH 5.5 and 6.5 the relative product distributions were equal with 96.5% lactate, 3% acetate and 0.5% succinate. At pH 4.5, this changed slightly to 91% lactate, 8% acetate and 1% succinate.

Next, we evaluated the ability of *B. smithii* ET 138 to ferment xylose (Figure 2 and S2), which is the most abundant sugar in lignocellulose next to glucose. The medium was changed from TMMY to TVMY and lower substrate concentrations of 20 g/L were used to decrease browning of the medium with xylose, which was observed when using 30 or 25 g/L xylose, probably due to a Maillard reaction. The growth curve on both substrates was highly similar (Figure 2 and S2). On both glucose and xylose, growth stopped after approximately 24 h while the acid production continued with only a small decrease in productivity, indicating that production is uncoupled from growth. The product profile on both substrates was found to be similar with lactate being the major product, *i.e.* 91% (13.9 g/L) on xylose and 92% (17.46 g/L) on glucose, with an optical purity of 99% L-lactate. On both sugars, acetate was the major by-product, followed by minor amounts of succinate and malate (Figure 2). In the xylose fermentation, two unknown peaks were observed in the HPLC spectrum, which were analyzed by NMR. One peak was identified as xylitol and was produced in amounts of approximately 1.5 g/L. The other peak was likely to be an acetylated form of xylose but could not be further identified with certainty. Whereas glucose is completely consumed at the end of fermentation, xylose is not. This might be related to browning of the medium after approximately 70 h, which was also observed at 25 g/L xylose concentrations (Figure S2) and was followed by a slight pyruvate accumulation.

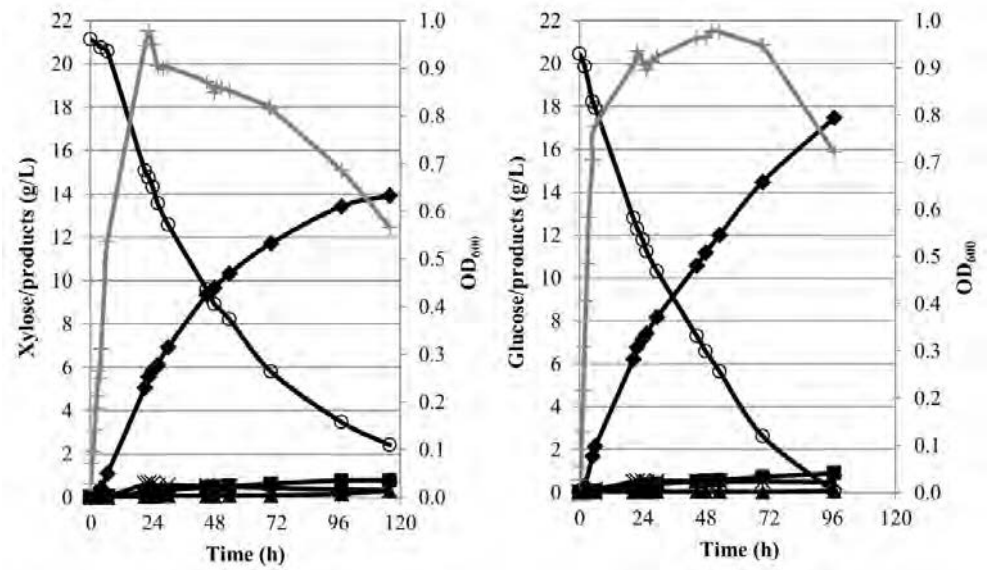

Figure 2. Fermentation of *B. smithii* ET 138 on xylose and glucose. Fermentation was carried out in 1 L TVMY supplemented with 20 g/L xylose or glucose, at 55°C, pH 6.5, 150 rpm and without any gas addition. Grey line with plus-symbol: OD_{600}; open circles: xylose or glucose; filled diamonds: lactate; closed squares: acetate; crosses: malate and succinate; closed triangles: pyruvate. During the xylose fermentation, browning of the medium was observed after approximately 70 h.

Optimization of transformation for *B. smithii* ET 138 and other *B. smithii* strains

The electroporation protocol for *B. smithii* ET 138 was optimized for increased transformation efficiencies. Out of 5 different tested electroporation settings (Table S2), the best was found to be 2.0 kV, 25 µF, 400 Ω in a 2 mm cuvette, resulting in 160 colonies per µg DNA (Table 4). Induction of antibiotic resistance by addition of a sub-lethal (1000x diluted) concentration of antibiotics after 2 h of recovery as described for *B. coagulans* (van Kranenburg et al., 2007) had no effect (Table S2). In contrast, the recovery medium had a large impact. Changing the original RG2 recovery medium (LB2 with 121 g/L sucrose and 10 g/L glucose) to LB2 resulted in an increase in transformation frequencies to ~1000 colonies per µg DNA. Transformation efficiency could be further increased to ~2000 by growing the cells in larger flasks prior to making them competent to allow for more aeration and faster growth (Table 4). Substituting the SG-buffer for the *Geobacillus*-electroporation buffer strongly reduced efficiencies to 40 colonies per µg DNA (Table S2). Lastly, we tested smaller amounts of added plasmid DNA. In all initial experiments, 1.0-2.5 µg DNA was added. Using only 20 ng DNA for transformation of *B. smithii* ET 138 resulted in a maximum efficiency of ~5000 colonies per µg DNA (Table 4). Storage of competent cells at -80°C was tested for periods up to one year and did not affect the transformation efficiency.

We also tested *B. smithii* strains from the DSMZ culture collection for transformation to evaluate their genetic accessibility. Using the optimized protocol for ET 138, a transformation efficiency of 10-100 colonies per µg plasmid DNA was obtained for type strain DSM 4216^T, which is approximately 20-200 times lower than that for strain ET 138 with the same conditions. Also for DSM 4216^T, different electroporation conditions and recovery media were tested (data not shown). For DSM 4216^T, the difference between recovery medium LB2 and RG2 was less pronounced than for ET 138. The best settings for DSM 4216^T were 1.5 kV, 25 µF, 600 Ω in a 1 mm cuvette with recovery for 3 h in LB2, yielding 10-200 colonies per µg plasmid DNA. Under the tested conditions, transformation for the type strain was found to be less reproducible than for isolate ET 138. The optimal settings for type strain DSM 4216^T and isolate ET 138 were used to evaluate transformation of 5 more publically available *B. smithii* strains. Of those strains, only DSM 460 generated transformants (25 colonies per µg DNA) but no attempts for further optimization were made. The best and most reproducibly transformable *B. smithii* strain so far is our isolate ET 138 with a maximum of ~5000 colonies per µg DNA.

Similar attempts to increase transformation efficiencies by changing electroporation settings and buffers for ET 144-2 were unsuccessful (data not shown). For strain ET 251, no transformation optimization was performed because this strain showed poor fermentation performance (Table 2).

Table 4. Optimization of electrotransformation for *B. smithii* ET 138.

Parameter changed	Final OD[a]	Hrs of growth[b]	Electroporation settings			Cuvette mm	µg DNA	Recovery medium[c]	CFU[d]
			kV	µF	Ω				
Settings[e]	0.483	1.75	1.5	25	600	1	1	RG2	3
Settings[f]	0.483	1.75	2.0	25	200	2	1	RG2	33
Settings	0.424	2.9	2.0	25	400	2	0.7	RG2	157
Recovery medium	0.519	1.75	2.0	25	400	2	1	LB2	960
Recovery medium	0.519	1.75	2.0	25	400	2	1	RG2	5
Fast growth[g]	0.539	1	2.0	25	400	2	2.5	LB2	1900
µg DNA added[g]	0.617	1.3	2.0	25	400	2	0.2	LB2	2409
µg DNA added[g]	0.617	1.3	2.0	25	400	2	0.02	LB2	5118

[a] Final OD_{600} after the indicated number of hours ([b]) when growing cells prior to making them competent.
[b] Number of hours cells had grown before making them competent.
[c] RG2 is LB2 with 121 g/L sucrose and 10 g/L glucose (Rhee et al., 2007).
[d] CFU: colony forming units per µg DNA. Using fresh cells or cells stored at -80°C did not change CFUs.
[e] Settings from (Rhee et al., 2007).
[f] Settings based on (Narumi et al., 1992), same settings as used for screening of *Geobacillus* strains.
[g] In these experiments, after overnight growth, cells were transferred to 500 mL Erlenmeyer flasks or 1 L bottles (having a similar bottom surface) instead of to a 250 mL Erlenmeyer to allow for more aeration.

DISCUSSION

In this study we aimed at isolating thermophilic bacilli that are genetically accessible and our isolation conditions were aimed at finding industrially relevant strains: minimal medium was used, temperature was set to moderately thermophilic temperatures, both micro-aerobic and anaerobic conditions were applied during the isolation procedure to find facultative anaerobes, and utilization of both C_5 and C_6 sugars was evaluated. Using 5 different isolation conditions, we isolated 267 strains from compost, of which 44 were identified by 16S rRNA gene sequencing that all belonged to the family *Bacillaceae*. The observed abundance of the genera *Bacillus* and *Geobacillus* (notably *G. thermodenitrificans)* among the isolates from the thermophilic compost is in line with a recent report describing the microbial succession during thermophilic composting (Li et al., 2014). In contrast to Li *et al*., we did not find any *Ureibacilli* and *B. licheniformis* but isolated several *B. coagulans* strains and one *B. smithii* (Li et al., 2014). These differences might be explained by the fact that for isolation we used a minimal medium and compost that consisted solely of plant material, whereas the other study made use of rich LB medium and compost that contained both plant material and manure (Li et al., 2014; Wang et al., 2011b).

Product formation on glucose was evaluated for all our isolates. In general, the product profiles are in line with other reports on thermophilic bacilli (Fong et al., 2006; Patel et al., 2006; Wang et al., 2011b). Whereas these studies mainly focused on a particular species or strain, our data provide a comparison of several thermophilic *Bacillus* species and strains grown under the same conditions. This revealed homogeneity in the main products lactate and acetate, whereas a large diversity was observed in the formation of by-products between species as well as between strains of the same species. This diversity creates opportunities for the use of these organisms for the production of different chemicals and their use as platform organisms or source of enzymes and pathways. *G. thermoglucosidans* DSM 2542^T and *B. coagulans* DSM 1^T were included as benchmarks because these two organisms have been studied for production of ethanol (Atkinson et al., 2008) and lactate (van Kranenburg et al., 2007), respectively, and both are genetically accessible (Kovacs et al., 2010; Taylor et al., 2008; van Kranenburg et al., 2007). *B. smithii* ET 138 was the only strain from our study that had genetic accessibility and product titers comparable to the benchmark organisms.

To the best of our knowledge, this is the first large-scale screen of genetic accessibility among different species of thermophilic bacilli. Although several thermophilic *Bacillus* species have been described to be genetically accessible (Blanchard et al., 2014; Gao et al., 2011; Rhee et al., 2007; Suzuki & Yoshida, 2012; Taylor et al., 2008; van Kranenburg et al., 2007; Wang et al., 2012b), often laborious protocols such as protoplast transformation are required. Moreover, for each species highly variable transformation efficiencies have

been reported for a limited number of strains, suggesting that genetic accessibility is very strain-specific. Strain-specificity is also implicated by Patel *et al.* who reported only one of their *B. coagulans* isolates to be genetically accessible (Patel et al., 2006). This is supported by our findings: in the first isolation round, only 2 out of 25 tested *G. thermodenitrificans* isolates were genetically accessible and no *G. stearothermophilus* or *B. coagulans* strains were transformable under the tested conditions. After testing the electrotransformation for all selected isolates, 3 strains were transformable: *B. smithii* ET 138 and *G. thermodenitrificans* ET 144-2 and ET 251, of which ET 138 was the most reproducible. A second isolation round, in which we enriched strains in reactors, resulted in four *G. thermoglucosidans* strains that could be reproducibly transformed and one *G. toebii* strain that was not reproducibly transformable (data not shown). Further optimization of the protocols per strain or the use of other transformation methods such as protoplast transformation or conjugation might increase the amount of transformable strains, but for high throughput engineering and use as platform organism, electroporation is a fast and convenient method. As *B. smithii* ET 138 outperformed the other strains in both fermentation and transformation via simple electroporation, we focused optimization experiments on this strain. The optimized electroporation efficiency of *B. smithii* ET 138 of $5x10^3$ using pNW33n is close to that found in *B. coagulans* P4-102B with $1x10^4$ using pNW33n (Rhee et al., 2007) and *G. thermoglucosidans* with $1x10^4$ using *E. coli-Geobacillus* shuttle vector pUCG18 (Taylor et al., 2008).

B. smithii was originally re-classified from *B. coagulans* (Nakamura et al., 1988), but contrary to *B. coagulans*, genetic accessibility has not been reported previously. *B. smithii* has been described for the production of several biotechnologically interesting enzymes such as nitrile hydratase (Takashima et al., 2000), endoinulinase (Gao et al., 2009) and alkaline lipase (Lailaja & Chandrasekaran, 2013). Its biotechnological potential was further reported in an isolation study targeted at xylose-utilizing, ethanol-tolerant ethanol-producers (Qi et al., 2011), as well as during lignocellulose-based composting, where *B. smithii* was among three bacterial compost isolates showing ligninolytic activity (Jurado et al., 2014). Several *B. smithii* strains were isolated from a sugar beet factory and shown to harbour glycosylated S-layer proteins and to secrete lactic acid as their main product (Messner et al., 1997). The sugar beet factory isolates showed temperature curves ranging from 37°C to 65°C, although not all strains grew at 37°C and the type strain did not grow at 65°C. The sugar beet extraction plant that was used as isolation source was around pH 4.5, indicating acid tolerance of the strains (Messner et al., 1997). In line with these findings, our isolated strain ET 138 grows between 37°C and 65°C, with highly similar growth rates between 51 and 57°C and it is still well able to grow and produce acids at pH 5.5 and even to some extent at pH 4.5. Such robustness against temperature and pH fluctuations is an important advantage in industrial applications to increase process stability. Another useful industrial property is its capacity to uncouple growth and acid production. Its robustness, combined

with the ability to efficiently utilize C_5 and C_6 sugars present in lignocellulosic substrates and being genetic accessible, makes *B. smithii* makes *B. smithii* an attractive biocatalyst for lactic acid production, having the potential to be developed into a novel platform organism. Genome sequencing is currently on-going, as well as the development of genetic tools to allow its further exploitation by metabolic engineering.

CONCLUSIONS

Thermophilic bacilli able to ferment glucose and xylose can readily be isolated from compost. All produce lactic acid and acetic acid under micro-aerobic conditions, while production of other products such as succinate, malate, ethanol, formate, 2,3-butanediol, or propionate is variable. Genetic accessibility via electroporation is scarce and strain-dependent. We have isolated *B. smithii* ET 138 and established an efficient transformation protocol. The combination of fermentation robustness and genetic accessibility makes this strain an attractive biocatalyst and candidate for further development into a platform organism for production of green chemicals from renewable resources.

MATERIALS AND METHODS

MEDIA, CULTIVATION METHODS AND STRAINS

Thermophile Minimal Medium (TMM, adjusted from (Fong et al., 2006)) contained per L: 8.37 g MOPS and 100 mL 10x concentrated Salt Solution (TSS-10x, containing per L: 2.3 g K_2HPO_4; 5.1 g NH_4Cl; 50 g NaCl; 14.7 g Na_2SO_4; 0.8 g $NaHCO_3$; 2.5 g KCl; 18.7 g $MgCl_2.6H_2O$; 4.1 g $CaCl_2.2H_2O$; 0.08 g $SrCl_2.6H_2O$; 0.08 g H_3BO_3; 9.0 g $NaNO_3$). The pH was set at 6.94 at room temperature unless indicated otherwise, the medium was autoclaved for 20 min at 121°C and cooled to max. 80°C , after which substrate and 10 mL per L of 1 mM $FeSO_4.7H_2O$ in 0.4 M tricine was added.

TMMY is TMM supplemented with 0.5 g/L yeast extract (Roth, Germany).

Thermophile Vitamin Medium with Yeast extract (TVMY) resembles TMMY, with the exception that $SrCl_2.6H_2O$, H_3BO_3 and $NaNO_3$ is omitted and that metal and vitamin mixtures are added after sterilization. Metal mix (1000x concentrated) contained per L: 16.0 g $MnCl_2.6H_2O$; 1.0 g $ZnSO_4$; 2.0 g H_3BO_3; 0.1 g $CuSO_4.5H_2O$; 0.1 g $Na_2MoO_4.2H_20$; 1.0 g $CoCl_2.6H_2O$; 7.0 g $FeSO_4.7H_2O$. Vitamin mix (1000x concentrated) contained per L: 0.1 g thiamine; 0.1 g riboflavin; 0.5 g nicotinic acid; 0.1 g panthothenic 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 L: 10 g tryptone (Oxoid, United Kingdom), 5 g yeast extract (Roth), 100 mL TSS-10x. pH was set to 6.94 – 7.00 at room temperature.

TGP2 (adjusted from (Taylor et al., 2008)) contained per L: 17 g tryptone (Oxoid), 3 g neutralized soy peptone (Oxoid), 100 mL TSS-10x. pH was set to 6.94 – 7.00 at room temperature. After autoclaving, 4 mL/L sterile 60% glycerol and 4 g/L filter sterilized 200 g/L sodium pyruvate were added.

For all tube and plate cultures, carbon substrates were used at a concentration of 10 g/L. For cultivation in reactors, carbon substrates were used in a concentration of 30 g/L unless indicated otherwise. Substrates were added separately as 50% sterile solutions after autoclaving of the medium; xylose and arabinose were filter sterilized and glucose and sucrose were autoclaved. For plates, 5 g/L gelrite (Roth) was added. The acid indicator bromocresol purple was used to monitor acidification in plates and tubes. Anaerobic cultivation of plates was performed in an anaerobic jar (Oxoid) containing a GasPak (Oxoid). All strains were routinely grown at their isolation temperature (55°C or 65°C), unless stated otherwise. All strains obtained from DSMZ were routinely grown at 55°C.

To evaluate the optimum temperature of strain ET 138, cells were grown overnight from glycerol stock in 10 mL LB2 at 55°C and next morning diluted to an OD_{600} of 0.05 in 20 mL TMMY supplemented with 10 g/L glucose in 50 mL Greiner tubes and incubated in water baths at different temperatures without shaking. A second experiment was performed in the same way, but with the overnight cultures grown at the same temperatures as to which they were transferred next morning. Tested temperatures were 37, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 57, 60, 63, 66 and 69°C. All these experiments were performed in triplicate. The same experimental set-up was used to evaluate the optimum pH of ET 138, but then the temperature was set at 55°C and the initial pH of the medium was varied from 3.0 to 7.0 with intervals of 0.5. For pH values from 3.0-5.5, MES buffer was used instead of MOPS.

The following strains were obtained from DSMZ (Germany): *G. thermoglucosidans* DSM 2542^T, *B. coagulans* DSM 1^T, *B. smithii* DSM 4216^T, *B. smithii* DSM 459, *B. smithii* DSM 460, *B. smithii* DSM 2319, *B. smithii* DSM 2320 and *B. smithii* DSM 2321.

Sampling, isolation and initial selection

Fresh compost samples were collected from two compost heaps at Recom Ede (NL) on August 5, 2010, both consisting solely of plant material. One heap was sieved through a 2-3 cm mesh and the other through a 1-2 cm mesh. The heaps are mixed every two weeks, resulting in an environment suitable for isolating facultatively anaerobic organisms. The temperature of the compost was 75°C at ±30 cm depth and sampling material was taken from the surface until ± 30 cm depth. Samples were taken by scraping compost into a plastic jar, transported to the laboratory (± 20 min) at ambient temperature to the lab and

stored at 4°C until use (1-4 weeks).

Five different conditions were applied during the isolation procedure, namely: 1: sucrose at 55°C, 2: sucrose at 65°C supplemented with 5 g/L $CaCO_3$, 3: sucrose at 55°C supplemented with 5 g/L $CaCO_3$, 4: glucose at 65°C supplemented with 5 g/L $CaCO_3$, 5: glucose at 55°C supplemented with 5 g/L $CaCO_3$. In all conditions, TMM was used containing the indicated carbon source and were kept the same throughout the whole procedure for each condition. The isolation procedure was as follows: for each of the two compost types, 5 g compost was added to a 250 mL Erlenmeyer flask containing 50 mL TMM medium supplemented with the indicated carbon source and shaken for 6 h at 120 rpm at either 55°C or 65°C. After 6 h, dilution series were plated on the same medium containing 5 g/L GelRite (Roth) and grown anaerobically for 24-48 h at either 55°C or 65°C. Subsequently, single colonies were picked and each colony was streaked to 4 plates, containing TMM with either glucose, xylose, arabinose or sucrose and incubated anaerobically for 24-48 h at the same temperature. Strains grown on 3 or more sugars, of which at least one C_5 and one C_6 sugar, were inoculated into 8 mL of their isolation medium containing bromocresol purple in 15 mL screw-capped tubes and incubated without shaking for 24-48 h at 55 or 65°C. For the conditions containing $CaCO_3$, this was added to the tubes as 500 µL autoclaved solution containing 80% (v/v) $CaCO_3$ grains. For strains that grew and acidified the medium, glycerol stocks were made by addition of 60% (v/v) glycerol to make 15% (v/v) and stored at -80°C.

In order to perform a production-based screening, all isolated strains were inoculated in duplicate from glycerol stocks into 8 mL TMMY supplemented with 10 g/L glucose and 500 µL of 80% $CaCO_3$ in 15 mL tubes, placed horizontally while shaking at 40 rpm and incubated for 48 h at their isolation temperature (55°C or 65°C), after which OD_{600} and pH were measured and HPLC analysis was performed. Based on HPLC results, pure cultures were made for selected strains by transferring single colonies twice on TMMY plates supplemented with 10 g/L glucose before finally transferring them to the same medium in liquid form to make stocks after 24-30h of growth. Pure cultures were subjected to 16S identification and used for further studies.

16S rRNA IDENTIFICATION AND PHYLOGENETIC ANALYSIS

A little material from single colonies was transferred from plates to a PCR tube using a sterile toothpick and subsequently incubated in a microwave for 1.5 min at 800W. PCR mix was added containing 2.5 U DreamTaq DNA polymerase (Fermentas, UK), DreamTaq buffer (Fermentas), 1 mM dNTPs (Fermentas), 0.2 µM primers GM3 (AGAGTTTGATCATGGC) and GM4 (TACCTTGTTACGACTT) and MilliQ water to a total volume of 50 µL. PCR products were checked on agarose gel. Products were purified either by GeneJet PCR purification kit (Fermentas), GeneJet Gel Extraction kit (Fermentas) or by a sequencing company, being either BaseClear (NL) or GATC (DE). GM3 and GM4 sequences were

assembled to one sequence using CloneManager and manually curated and trimmed until 1358 bp, after which BLASTn was used for identification against the 16S database. Mega6 software (Tamura et al., 2013) was used for creating the alignment using Clustal (Thompson et al., 1994) after which the Neighbour Joining method (Saitou & Nei, 1987) was used to create the phylogram and bootstrap analysis (Felsenstein, 1985) was performed using 1000 replicate analyses.

Testing genetic accessibility

Depending on their 16S rRNA gene-based genus identification, strains selected in the initial selection were subjected to a transformation protocol for either *Geobacillus* (Taylor et al., 2008) or *Bacillus* (Rhee et al., 2007) with slight modifications as follows: strains were grown overnight at their isolation temperature in 10 mL medium in a 50 mL Greiner tube and next morning, cultures were diluted to an OD_{600} of 0.08 in 100 mL medium for *Bacilli* and 50 mL medium for *Geobacilli*. For strains belonging to the genus *Bacillus*, LB2 medium was used and for strains belonging to the *Geobacillus* genus, TGP2 medium was used, with the exception of *Geobacillus* strains that grew better on LB2 (isolates ET 036, ET 050, ET 130, ET 144-2 and ET 261-3). Cultures were grown until they reached an OD_{600} of 0.45-0.65 for *Bacilli*, and around 1.00 for *Geobacilli.* After washing 3x with SG buffer for *Bacilli* (per L: 171.2 g sucrose, 0.2 g $MgCl_2$, 50 mL glycerol) or 4x with electroporation buffer for *Geobacilli* (per L: 91 g sorbitol, 91 g mannitol, 100 mL glycerol), final pellets were resuspended in 240 µL SG buffer for *Bacilli*, and 900 µL electroporation buffer for *Geobacilli*, after which cells were either stored at -80°C or electroporation was performed directly. Electroporation settings were 2.0 kV, 25 µF, 200 Ω in a 2 mm cuvette (based on (Narumi et al., 1992)) for all strains. Additional settings were tested as follows: 1.5 kV, 25 µF, 600 Ω in a 1 mm cuvette (Rhee et al., 2007) for *Bacilli* strains ET 138, ET 143, ET 224, ET 226, ET 229, ET 230-1, ET 239-2, ET 244, ET 261-1 and ET 263; 2.5 kV, 25 µF, 600 Ω in a 1 mm cuvette (Taylor et al., 2008) for *Geobacilli* strains ET 006, ET 011-1, ET 036, ET 042, ET 050 and ET 208; 2.5 kV, 25 µF, 600 Ω in a 2 mm cuvette (based on (Taylor et al., 2008)) for ET 006, ET 036, ET 042, ET 050, ET 072, ET 129 and ET 208. Recovery was performed at 3°C below the isolation temperatures for 3 h in RG2 medium (LB2 with 121 g/L sucrose and 10 g/L glucose) for *Bacilli* and for 2 h in LB2 or TGP2 supplemented with 0.5% (w/v) glucose for *Geobacilli*. *G. thermoglucosidans* DSM 2542^T and *B. coagulans* DSM 1^T were used as positive controls. All strains were tested with plasmid pNW33n (Bacillus Genetic Stock Centre) and 7 µg/mL chloramphenicol was added to the plates after transformation. Negative controls were performed by electroporating without adding any plasmid DNA.

Colonies appearing after transformation were streaked to a new plate and grown overnight, after which grown streaks were subjected to colony PCR as described above, using primers BG3464 (AACTCTCCGTCGCTATTGTAACCA) and BG3465 (TATGCGTGCAACG-

GAAGTGAC). To confirm transformation, positive colonies were inoculated into 10 mL LB2 or TGP2 supplemented with 7 µg/mL chloramphenicol in a 50 mL tube and grown overnight at their isolation temperature. From this culture, plasmid was extracted using the GeneJET Plasmid Miniprep Kit (Fermentas) with 4 mg/mL lysozyme (Sigma, USA) added to the resuspension buffer, in which resuspended pellets were incubated for 30 min at 37°C. Extracted plasmid was used for PCR as described above and restriction analysis using 550-650 ng plasmid DNA with Fermentas enzymes StuI and HindIII for 2 h at 37°C. Plasmid extraction on non-transformed strains was used as negative control. In case of inconclusive band patterns, the isolated plasmid DNA was transformed to *E. coli* DH5α and subsequently re-isolated and digested.

Fermentations

Pre-cultures of transformable isolates were grown overnight from glycerol stocks in 10 ml LB2 at 55°C in a 50 mL Greiner tube without shaking. Next morning, 10 mL culture was transferred to 40 mL TMMY medium supplemented with 30 g/L glucose in a 250 mL Erlenmeyer flask and incubated at 55°C without shaking. When the cultures reached an OD_{600} of 0.3-0.7 (exponential growth phase), 20 mL was used to inoculatea reactor containing 1 L TMMY medium without MOPS and supplemented with 30 g/L glucose. Glass reactors of 2 L working volume were used (Applikon, The Netherlands) under control of an ADI 1010 Bio-controller (Applikon) with an ADI 1025 Bio-console. Temperature was controlled at 55°C, stirring speed was 150 rpm and pH was maintained at 6.5 unless stated otherwise by addition of 3 M KOH. Growth was monitored off-line by absorbance at 600 nm (OD_{600}) and sugar and fermentation products were measured by high-pressure liquid chromatography (HPLC).

Fermentations of ET 138 at different pHs were performed in the same way as described above, but overnight cultures were transferred to TMMY medium with the same pH as the reactor (either 4.5, 5.5 or 6.5), and pH of the reactors were controlled at pH 4.5, 5.5 or 6.5.

For the comparison of ET 138 on glucose and xylose, pre-cultures were inoculated from glycerol stock and grown overnight at 55°C in 10 mL TVMY medium supplemented with either 10 g/L glucose or 10 g/L xylose in a 50 mL tube at 150 rpm. Next morning, 1 mL culture was transferred to 50 mL of the same medium in a 250 mL Erlenmeyer flask at 55°C and 150 rpm. When this culture reached an OD_{600} of 0.5-0.7, 20 mL was used to inoculate a reactor. In the reactor, 1 L TVMY medium without MOPS was used, supplemented with 20 or 25 g/L xylose or glucose, as indicated.

Analytical techniques

Sugars and fermentation products were quantified using an HPLC system from Thermo containing a P2000 pump, an AS3000 autosampler, a UV/VIS1000 detector and an RI-150 refraction index detector. A Shodex RSpak KC-811 cation-exchange column was used with a mobile phase of 3 mM H_2SO_4 and operated at 0.8 mL/min and 80°C. All samples were diluted 1 on 1 with 10 mM DMSO in 0.04 N H_2SO_4. D-lactate and L-lactate were distinguished via enzymatic assay kits (MegaZyme D-LATE and K-DATE).

Unknown peaks observed in the HPLC spectrum in the xylose fermentation were collected from HPLC, freeze-dried and analyzed by NMR by Biqualys (NL). NMR-identification was confirmed by running the corresponding standards on HPLC.

GenBank accession numbers

All the 16S rRNA gene sequences generated in this study have been submitted to the GenBank database with accession numbers KP010222-KP010265.

Type strains used for the phylogenetic tree were trimmed to the same 1358 bp as the isolates and were derived from the following GenBank accession numbers: NR_043021.1 (*G. thermodenitrificans* DSM 465), NR_036987.1 (*B. smithii* DSM 4216), NR_109664.1 (*B. thermocopriae* SgZ7), NR_115727.1 (*B. coagulans* DSM 1), NR_026515.1 (*A. pallidus* DSM 3670), NR_043022.1 (*G. thermoglucosidans* DSM 2542), NR_028708.1 (*G. caldoxylosilyticus* DSM 12041).

CONFLICT OF INTEREST

The authors declare to have no conflict of interest. RvK is employed by the commercial company Corbion (Gorinchem, The Netherlands).

ACKNOWLEDGEMENTS

We thank Willem van Harten from Recom Ede for allowing us to sample their compost. We are grateful to Dr. Jacques Vervoort of Biqualis for discussions and NMR analyses and to Dr. Ron Winkler from Dutch Technology Foundation STW for his involvement in an earlier phase of this work. This work was financially supported by Corbion.

SUPPLEMENTARY DATA

Table S1. Growth, end product formation and genetic accessibility of selected isolates.

Strain	Identification[1]	Isol.[2]	Sugars supporting growth[3]				CFU[5]
			G	S	X	A	
ET 138	***B. smithii***	3	+	+	+	+/-	3
DSM 1	*B. coagulans*	T	+[7]	-[7]	-[7]	-[7]	20[6]
ET 236	*G. thermodenitrificans*	5	+/-	+	+	+	0
ET 239	*1: A. pallidus* *2: B. coagulans*	5	+/-	+/-	++	+/-	1: 0 2: bg
ET 224	*B. coagulans*	5	+/-	+/-	+	+/-	0
ET 251	***G. thermodenitrificans***	**5**	**+**	+/-	+/-	+/-	4
ET 131	*1: G. caldoxylosilyticus* *2: G. thermodenitrificans*	2	++	+/-	++	+/-	1/2: 0
DSM 2542	*G. thermoglucosidans*	T	+[8]	+[8]	+[8]	-[8]	2300
ET 130	*G. caldoxylosilyticus*	2	++	++	+	+/-	1: bg 2: 0
ET 186	*G. thermodenitrificans*	4	+/-	+	+/-	+/-	0
ET 144	*1: B. thermocopriae* ***2: G. thermodenitrificans***	3	+/-	+/-	+	+/-	1: 0 **2: 6**
ET 241	*G. thermodenitrificans*	5	+/-	+/-	+	+/-	0
ET 050	*G. thermodenitrificans*	1	+/-	+	++	-	bg
ET 208	*G. caldoxylosilyticus*	4	+/-	+/-	-	+/-	bg
ET 129	*1: G. caldoxylosilyticus* *2: G. thermodenitrificans*	2	++	+++	+	-	1: 0 2: bg
ET 143	*B. thermocopriae*	3	+	+	+	+/-	res
ET 072	*G. thermodenitrificans*	2	+	+	++	-	1/2: bg
ET 157	*G. thermodenitrificans*	3	+	+/-	+/-	-	0
ET 136	*G. thermodenitrificans*	3	+	++	+	+	0
ET 244	*B. thermocopriae*	5	+	+/-	+/-	+/-	0
ET 226	*B. thermocopriae*	5	+/-	+/-	+++	+/-	0

Products (g/L)[4]									
Lac	**Ace**	**Suc**	**Mal**	**Eth**	**For**	**2,3-BDO**	**Prop**	**Total**	**OD_{600}**
6.74 ±*0.18*	**0.81 ±*0.02***	**0.07 ±*0.00***	**nd**	**nd**	**nd**	**nd**	**nd**	**7.61 ±*0.16***	**1.425 ±*0.058***
6.37 ±*0.76*	0.36 ±*0.05*	0.13 ±*0.05*	0.18 ±*0.30*	0.04 ±*0.07*	nd	0.10 ±*0.02*	nd	7.17 ±*1.16*	1.467 ±*0.147*
5.37 ±*0.60*	1.15 ±*0.07*	0.13 ±*0.01*	0.17 ±*0.11*	0.04 ±*0.02*	0.07 ±*0.04*	0.05 ±*0.01*	nd	6.97 ±*0.85*	0.833 ±*0.006*
5.03 ±*0.22*	0.87 ±*0.19*	0.15 ±*0.04*	nd[1]	0.07 ±*0.00*	nd	0.04 ±*0.03*	nd	6.68 ±*0.27*	0.928 ±*0.041*
4.47 ±*0.12*	0.72 ±*0.02*	0.20 ±*0.00*	nd[1]	nd	nd	0.06 ±*0.01*	nd	6.20 ±*0.03*	0.852 ±*0.011*
4.89 ±*0.03*	**0.85 ±*0.02***	**0.13 ±*0.00***	**0.11 ±*0.01***	**0.13 ±*0.02***	**0.00 ±*0.00***	**0.07 ±*0.01***	**nd**	**6.18 ±*0.03***	**0.876 ±*0.002***
4.71 ±*0.37*	0.91 ±*0.00*	0.14 ±*0.04*	0.12 ±*0.02*	nd	0.02 ±*0.01*	nd	nd	5.91 ±*0.39*	0.538 ±*0.076*
4.39 ±*0.19*	0.92 ±*0.03*	0.21 ±*0.05*	0.03 ±*0.02*	0.09 ±*0.13*	0.14 ±*0.15*	nd	0.02 ±*0.01*	5.81 ±*0.26*	0.695 ±*0.195*
4.59 ±*0.37*	0.76 ±*0.04*	0.18 ±*0.01*	0.08 ±*0.01*	nd	0.04 ±*0.00*	nd	nd	5.64 ±*0.40*	0.578 ±*0.048*
4.12 ±*0.39*	0.92 ±*0.01*	0.21 ±*0.07*	0.07 ±*0.03*	nd	0.02 ±*0.00*	0.04 ±*0.06*	nd	5.37 ±*0.49*	0.803 ±*0.168*
3.63 ±*0.34*	**1.02 ±*0.03***	**0.15 ±*0.02***	**0.12 ±*0.10***	**0.21 ±*0.18***	**0.11 ±*0.02***	**0.08 ±*0.01***	**nd**	**5.32 ±*0.06***	**0.817 ±*0.101***
3.99 ±*0.89*	0.98 ±*0.07*	0.16 ±*0.05*	0.01 ±*0.01*	nd	0.06 ±*0.05*	0.03 ±*0.05*	nd	5.23 ±*1.00*	0.662 ±*0.018*
3.72 ±*0.49*	1.01 ±*0.24*	0.12 ±*0.02*	0.09 ±*0.04*	nd	0.02 ±*0.00*	0.05 ±*0.07*	nd	5.00 ±*0.85*	0.611 ±*0.125*
3.14 ±*0.80*	1.06 ±*0.14*	0.57 ±*0.20*	0.16 ±*0.03*	nd	0.04 ±*0.01*	nd	nd	4.98 ±*1.18*	0.881 ±*0.081*
3.56 ±*0.48*	0.99 ±*0.06*	0.16 ±*0.04*	0.08 ±*0.02*	nd	0.02 ±*0.01*	nd	nd	4.81 ±*0.36*	0.579 ±*0.082*
2.81 ±*0.33*	1.07 ±*0.17*	0.15 ±*0.04*	nd	0.26 ±*0.15*	0.20 ±*0.12*	0.01 ±*0.01*	0.09 ±*0.08*	4.60 ±*0.90*	0.719 ±*0.095*
3.48 ±*0.41*	0.81 ±*0.18*	0.10 ±*0.01*	nd	nd	0.01 ±*0.00*	0.11 ±*0.15*	nd	4.52 ±*0.45*	0.356 ±*0.020*
3.45 ±*0.03*	0.94 ±*0.06*	0.07 ±*0.01*	nd	nd	0.01 ±*0.00*	0.03 ±*0.02*	nd	4.49 ±*0.03*	0.553 ±*0.028*
2.46 ±*0.07*	1.14 ±*0.01*	0.16 ±*0.03*	nd	0.50 ±*0.42*	0.16 ±*0.04*	nd	0.03 ±*0.02*	4.45 ±*0.58*	0.564 ±*0.090*
3.46 ±*0.38*	0.36 ±*0.02*	0.01 ±*0.00*	nd[1]	nd	nd	0.04 ±*0.00*	0.02 ±*0.00*	3.89 ±*0.41*	0.403 ±*0.054*
3.41 ±*0.18*	0.38 ±*0.01*	0.02 ±*0.01*	nd	nd	nd	0.01 ±*0.01*	0.02 ±*0.00*	3.84 ±*0.21*	0.663 ±*0.020*

Table S1 continued.

Strain	Identification[1]	Isol.[2]	Sugars supporting growth[3]				CFU[5]
			G	S	X	A	
ET 225	*A. pallidus*	5	+/-	+/-	+	+/-	bg
ET 039	*G. thermodenitrificans*	1	+	+	+	+	bg
ET 261	*1+2: B. thermocopriae* *3: G. thermodenitrificans*	5	+	+/-	+	+	1: bg 2/3: 0
ET 145	*G. thermodenitrificans*	3	+	+/-	+	+/-	0
ET 200	*G. thermodenitrificans*	4	+	+/-	+	+/-	0
ET 036	*G. thermodenitrificans*	1	+	+	+	+	bg
ET 156	*G. thermodenitrificans*	3	++	++	+	+/-	0
ET 042	*G. thermodenitrificans*	1	+	+	+	+	0
ET 229	B. coagulans	5	+/-	+/-	+	+/-	0
ET 267	*1+2: G. thermodenitrificans*	5	+	+/-	+	+/-	1/2: 0
ET 011	*1+2: G. thermodenitrificans*	1	+	+	+	+	1: 0 2: bg
ET 081	G. thermodenitrificans	2	+	+	++	-	bg
ET 230	*1+2: B. coagulans*	5	+/-	+/-	++	+++	1/2: bg
ET 263	*B. thermocopriae*	5	+	+/-	+	+/-	0
ET 159	*G. thermodenitrificans*	4	+/-	+/-	+/-	+/-	0
ET 160	*G. thermodenitrificans*	4	+/-	+/-	-	+/-	bg

[1] The selection based on products was made prior to making pure cultures – the identification shown is that after making pure cultures; in cases where the pure culture was split in multiple species both are shown because in this selection round it was still a mixed culture. In the CFU-column, the identification for the subcultures is provided after the corresponding strain sub-number (e.g. 1, 2, or 3) as this was performed after making pure cultures.
[2] Isol: isolation condition (see Table 1).
[3] Sugars used in the selection step in the isolation procedure (see Table 1), prior to making pure cultures. Abbreviations: G: glucose; S: sucrose; X: xylose; A: arabinose; na: not applicable.
[4] Cultures were analyzed by HPLC after 48 h of growth at their isolation temperature (55 or 65°C) in 15 mL screw-capped tubes with 8 mL TMM + 0.5 g/L yeast extract + 10 g/L glucose + $CaCO_3$. Transformable strains are shown in bold. Two genetically accessible *Bacillus* type strains known from literature for green chemical or fuel production were taken along as reference strains, e.g. *G. thermoglucosidans* DSM 2542^T (Cripps et al., 2009) and *B. coagulans* DSM 1^T (Kovacs et al., 2010; van Kranenburg et al., 2007). Data are the average of duplicates; standard deviations are

Products (g/L)[4]									
Lac	**Ace**	**Suc**	**Mal**	**Eth**	**For**	**2,3-BDO**	**Prop**	**Total**	**OD_{600}**
3.07 ±*0.87*	0.63 ±*0.00*	0.07 ±*0.03*	nd	nd	nd	nd	0.03 ±*0.01*	3.79 ±*0.89*	0.558 ±*0.033*
2.57 ±*0.08*	0.92 ±*0.07*	0.07 ±*0.03*	0.04 ±*0.01*	nd	0.02 ±*0.00*	nd	nd	3.62 ±*0.19*	0.481 ±*0.010*
1.96 ±*0.17*	1.14 ±*0.10*	0.06 ±*0.02*	nd	0.23 ±*0.01*	0.20 ±*0.09*	nd	0.03 ±*0.00*	3.62 ±*0.02*	0.613 ±*0.014*
1.91 ±*0.03*	1.18 ±*0.09*	0.11 ±*0.06*	nd	0.24 ±*0.05*	0.10 ±*0.02*	0.01 ±*0.01*	nd	3.54 ±*0.14*	0.505 ±*0.029*
2.65 ±*0.01*	0.64 ±*0.25*	0.16 ±*0.04*	0.07 ±*0.05*	nd	nd	nd	nd	3.53 ±*0.34*	0.729 ±*0.055*
2.50 ±*1.41*	0.80 ±*0.31*	0.08 ±*0.07*	0.12 ±*0.11*	nd	0.02 ±*0.01*	nd	nd	3.52 ±*1.69*	0.456 ±*0.052*
2.34 ±*1.02*	0.90 ±*0.20*	0.16 ±*0.11*	0.02 ±*0.02*	nd	0.01 ±*0.01*	0.04 ±*0.05*	nd	3.46 ±*1.36*	0.448 ±*0.083*
2.36 ±*0.21*	0.89 ±*0.07*	0.03 ±*0.01*	0.07 ±*0.03*	nd	0.02 ±*0.01*	nd	nd	3.37 ±*0.27*	0.627 ±*0.106*
2.83 ±*1.61*	0.47 ±*0.03*	0.03 ±*0.02*	nd	nd	nd	nd	0.02 ±*0.00*	3.35 ±*1.67*	0.373 ±*0.045*
2.19 ±*0.90*	0.83 ±*0.00*	0.05 ±*0.03*	0.11 ±*0.15*	0.03 ±*0.04*	0.02 ±*0.00*	nd	nd	3.22 ±*1.13*	0.432 ±*0.048*
2.00 ±*0.91*	0.79 ±*0.22*	0.08 ±*0.03*	0.08 ±*0.01*	0.14 ±*0.20*	0.02 ±*0.00*	nd	nd	3.11 ±*1.35*	0.539 ±*0.117*
2.03 ±*0.20*	0.93 ±*0.11*	0.07 ±*0.00*	0.06 ±*0.02*	nd	0.01 ±*0.00*	nd	nd	3.10 ±*0.08*	0.740 ±*0.091*
2.34 ±*1.00*	0.66 ±*0.00*	0.04 ±*0.01*	nd	nd	nd	nd	0.02 ±*0.00*	3.06 ±*1.01*	0.419 ±*0.072*
1.72 ±*0.54*	0.85 ±*0.16*	0.03 ±*0.01*	nd	0.24 ±*0.10*	0.05 ±*0.00*	nd	nd	2.89 ±*0.80*	0.600 ±*0.098*
0.74 ±*0.02*	0.67 ±*0.06*	nd	nd	nd	0.05 ±*0.01*	nd	0.01 ±*0.00*	1.46 ±*0.06*	0.253 ±*0.001*
0.69 ±*0.03*	0.53 ±*0.03*	nd	nd	nd	0.04 ±*0.06*	nd	0.00 ±*0.01*	1.27 ±*0.11*	0.315 ±*0.103*

indicated in italics after the '±'. nd = not detected; nd[1] = a very high malate peak with a large shoulder was observed, indicating the presence of another product besides or instead of malate but this was not further evaluated. Abbreviations: Lac: lactate; Ace: acetate; Suc: succinate; Mal: malate; Eth: ethanol; For: formate; 2,3-BDO: 2,3-butanediol; Prop: propionate.

[5] CFU = colony forming units per µg DNA after electrotransformation with pNW33n. Transformation was confirmed by isolating plasmid material and subsequent PCR and restriction analysis (Figure S1). Res = naturally resistant to chloramphenicol; nr = not reproducible, bg = background colonies after electroporation, also when no DNA was added, while these strains did not grow on chloramphenicol prior to transformation – colonies by these strains did not test positive in PCR. Results are shown for pure cultures, e.g. 239-1 and 239-2 are '1:0 2:bg' meaning 239-1 showed 0 colonies and 239-2 showed background colonies.

[6] As reported by (van Kranenburg et al., 2007).

[7] Acid production from these sugars as determined by (Jung et al., 2009).

[8] Acid production from these sugars as determined by (Suzuki et al., 1983).

Table S2. Optimization of electrotransformation for *B. smithii* ET 138.

Nr	Parameter changed	Buffer[1]	Final OD[2]	Hrs of growth[3]	kV	µF	Ω	µg DNA	Cuvette mm	Rec. medium[4]	Cm rec[5]	CFU fresh[6]	CFU -80[6]
1	Cm rec+sett.	SG-5	0.483	1.75	1.5	25	600	1	1	RG	N	3	nd
2	Cm rec+sett.	SG-5	0.500	1.5	1.5	25	600	1	1	RG	Y	8	nd
3	Cm rec+sett.	SG-5	0.483	1.75	2.0	25	200	1	2	RG	N	33	nd
4	Cm rec+sett.	SG-5	0.500	1.5	2.0	25	200	5	2	RG	Y	3	nd
5	Settings	SG-5	0.424	2.9	2.0	25	200	1	2	RG	N	12	nd
6	Settings	SG-5	0.424	2.9	1.5	25	200	0.7	1	RG	N	nd	113
7	Settings	SG-5	0.424	2.9	2.0	25	200	0.7	2	RG	N	nd	7
8	Settings	SG-5	0.424	2.9	2.0	25	400	0.7	2	RG	N	nd	157
9	OD	SG-5	0.452	1.4	2.0	25	400	1	2	RG	N	70	nd
10	OD	SG-5	0.678	1.4	2.0	25	400	1	2	RG	N	11	nd
11	Rec. medium	SG-5	0.574	2.1	2.0	25	400	1	2	LB2	N	712	1149
12	Rec. medium	SG-5	0.574	2.1	2.0	25	400	1	2	RG	N	0	nd
13	Rec. medium	SG-5	0.514	2.1	2.0	25	400	1	2	LB2	N	406	443
14	Rec. medium	SG-5	0.514	2.1	2.0	25	400	1	2	RG	N	1	nd
15	Glycerol, rec.	SG-10	0.446	2.5	2.0	25	400	1	2	LB2	N	796	476
16	Glycerol, rec.	SG-10	0.446	2.5	2.0	25	400	1	2	RG	N	36	nd
17	Rec. medium	SG-5	0.522	2.1	2.0	25	400	1	2	LB2	N	590	nd
18	Rec. medium	SG-5	0.522	2.1	2.0	25	400	1	2	RG	N	0	nd
19	Rec. medium	SG-5	0.519	1.75	2.0	25	400	1	2	LB2	N	960	nd
20	Rec. medium	SG-5	0.519	1.75	2.0	25	400	1	2	RG	N	5	nd
21	Buffer	SM-10	0.454	2.5	2.0	25	400	1	2	LB2	N	40	0
22	Buffer	SM-10	0.454	2.5	2.0	25	400	1	2	RG	N	0	nd
23	Fast growth[7]	SG-5	0.539	1	2.0	25	400	2.5	2	LB2	N	1900	nd
24	Fast growth[7]	SG-5	0.512	1.25	2.0	25	400	1	2	LB2	N	nd	1122
25	µg DNA[7]	SG-5	0.617	1.3	2.0	25	400	0.2	2	LB2	N	2409	nd
26	µg DNA[7]	SG-5	0.617	1.3	2.0	25	400	0.02	2	LB2	N	5118	nd

Lines indicate different experiments. Nr. 1 is the original protocol as described by Rhee *et al*.(Rhee et al., 2007). Nr. 3 are settings based on based on (Narumi et al., 1992), which are the settings as used for screening of *Geobacillus* strains.

[1] 5 or 10 indicates the % of glycerol. SG = sucrose glycerol buffer from *Bacillus* protocol (Rhee et al., 2007); SM = sorbitol mannitol buffer from *Geobacillus* protocol (Taylor et al., 2008).

[2] Final OD_{600} after the indicated number of hours ([3]) when growing cells prior to making them competent.

[3] Number of hours cells had grown before making them competent.

[4] RG is LB with 121 g/L sucrose and 10 g/L glucose (Rhee et al., 2007).

[5] Whether or not 1/1000 diluted chloramphenicol was added after 2h of recovery. Y = yes; N = no.

[6] CFU: colony forming units per µg DNA. Cells were either electroporated directly after making them competent ('fresh') or after storage in -80°C ('-80'). nd = not determined.

[7] In these experiments, after overnight growth, cells were transferred to 500 mL Erlenmeyer flasks or 1 L bottles (having a similar bottom surface) instead of to a 250 mL Erlenmeyer to allow for more aeration.

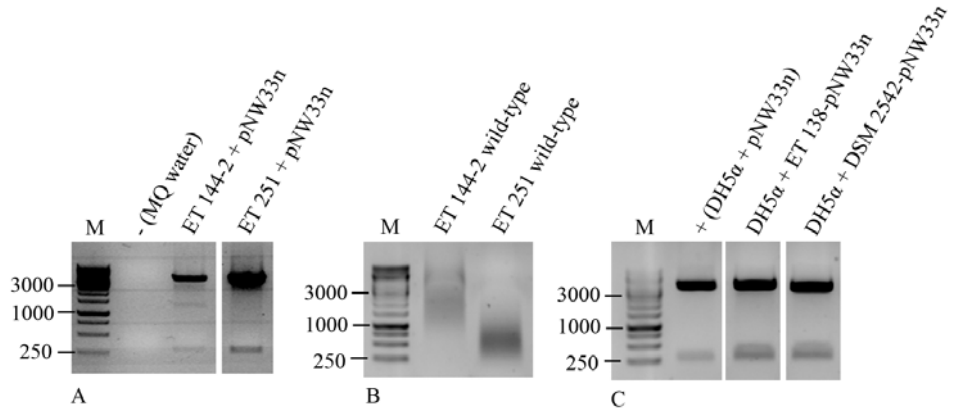

Figure S1. Restriction analysis of plasmid DNA extracted from isolates and control strain transformed with pNW33n. Restriction was performed with StuI and HindIII; pNW33n digested with StuI and HindIII yields bands of 282 and 3955 bp. M: Fermentas 1 kb DNA ladder. For each digestion reaction, 550-650 ng isolated plasmid DNA was used. **A:** Plasmids isolated from *G. thermodenitrificans* strains ET 144-2 and ET 251 transformed with pNW33n. **B:** Plasmids isolated from *G. thermodenitrificans* strains ET 144-2 and ET 251 prior to transformation with pNW33n (negative control) **C:** Plasmids isolated from *E. coli* DH5α transformed pNW33n (original source) and after retransformation with plasmids isolated from *B. smithii* ET 138 transformants and *G. thermoglucosidans* DSM 2542 transformants.

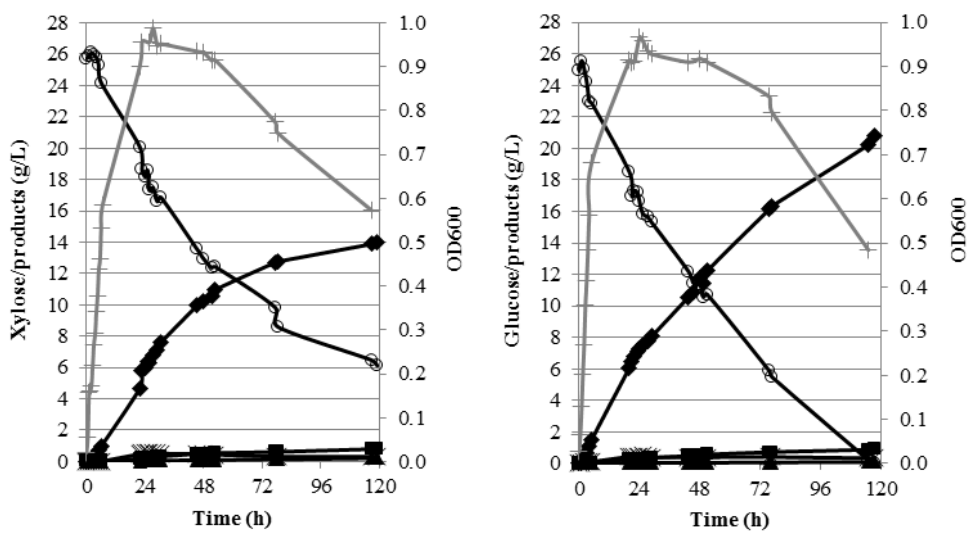

Figure S2. Fermentation of *B. smithii* ET 138 on 25 g/L xylose and glucose. Fermentation was carried out in 1 L TVMY supplemented with 25 g/L xylose or glucose, at 55°C, pH 6.5, 150 rpm and without any gas addition. Grey line with plus-sign: OD600; open circles: xylose or glucose; filled diamonds: lactate; closed squares: acetate; crosses: malate and succinate; closed triangles: pyruvate. During the xylose fermentation, severe browning of the medium was observed after approximately 70h.

CHAPTER 4

COMPLETE GENOME SEQUENCE OF THERMOPHILIC *BACILLUS SMITHII* TYPE STRAIN DSM 4216^T

Submitted for publication.

Bosma, E.F., Koehorst, J.J., van Hijum, S.A.F.T., Renckens, B. Vriesendorp, B., van de Weijer, A.H.P., Schaap, P.J., de Vos, W.M., van der Oost, J., van Kranenburg, R.

ABSTRACT

Bacillus smithii is a facultatively anaerobic, thermophilic bacterium able to use a variety of sugars that can be derived from lignocellulosic feedstocks. Combined with its genetic accessibility it is a potential new host for biotechnological production of green chemicals from renewable resources. We determined the complete genomic sequence of the *B. smithii* type strain DSM 4216^T, which consists of a 3,368,778 bp chromosome (GenBank accession number CP012024) and a 12,514 bp plasmid (GenBank accession number CP012025), together encoding 3880 genes. Genome annotation via RAST was complemented by a protein domain analysis. Some unique features of *B. smithii* central metabolism in comparison to related organisms included the lack of a standard acetate production pathway with no apparent pyruvate formate lyase, phosphotransacetylase, and acetate kinase genes, while acetate was the second fermentation product.

KEYWORDS

Bacillus smithii, genome sequence, lactic acid, thermophile, thermophilic bacillus, biotechnology

INTRODUCTION

Bacillus smithii is a facultatively anaerobic, thermophilic Gram positive bacterium, originally identified as *Bacillus coagulans* (Hammer, 1915; Nakamura et al., 1988). Similar to its close relative *B. coagulans*, *B. smithii* has biotechnological potential as it is able to ferment a range of carbon sources (Nakamura et al., 1988) into lactate and other green building block chemicals (Bosma et al., 2015a; Bosma et al., 2015b). The production of such green chemicals at elevated temperatures from lignocellulosic biomass has the potential to lower production costs of these chemicals. Compared to currently used mesophilic production hosts, such as Lactic Acid Bacteria (LAB) or *Escherichia coli*, the amount of enzymes needed for hydrolysis of lignocellulose is ~3-fold lower around 50-60°C, which is the temperature of moderately thermophilic temperatures (Ou et al., 2009). Furthermore, fermentation at higher temperatures decreases contamination risks and cooling costs and increases product and substrate solubility (Bosma et al., 2013; Taylor et al., 2009). In order to enable the development of *B. smithii* as a platform organism, genetic tools were recently developed for it (Bosma et al., 2015a; Bosma et al., 2015b). To fully exploit the biotechnological potential of this species and to gain insight into its metabolic pathways, we sequenced the genome of the *B. smithii* type strain. Reconstruction of the central metabolic pathways based on the genome reveals some remarkable differences with its close relative *B. coagulans.*

ORGANISM INFORMATION

Classification and features

B. smithii DSM 4216^T is a motile, spore-forming, rod-shaped (0.8-1.0 by 5.0-6.0 µm (Nakamura et al., 1988)/0.5-1.0 by 2.0-6.0 µm, Figure 1), facultatively anaerobic, facultatively thermophilic bacterium with wide ranges of both temperature (25-65°C) and pH (5.5-7.0) (Nakamura et al., 1988). An electron micrograph of *B. smithii* DSM 4216^T is shown in Figure 1. Based on existing literature (Nakamura et al., 1988), HPLC analysis (Bosma et al., 2015a; Bosma et al., 2015b) and API-tests, it is concluded that the species is able to ferment a range of carbon sources into mainly lactate, with acetate as the major by-product and minor amounts of succinate and malate (Table 1).

In order to compare the *B. smithii* DSM 4216^T genome to other fully sequenced *Bacillus* genomes, a phylogenetic tree was constructed based on 16S rRNA and the analysis of protein domains of *B. smithii* DSM 4216^T and other currently available *Bacillus* genomes (Figure 2, and Figure 2 Chapter 1) (Worm et al., 2014). These analyses indicated that *B. smithii* is most closely related to *B. coagulans* and *Bacillus thermocopriae*, which are both

Table 1. Classification and general features of *B. smithii* DSM 4216^T according to MIGS standards.

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain Bacteria	TAS (Woese et al., 1990)
		Phylum Firmicutes	TAS (Garrity & Holt, 2001; Gibbons & Murray, 1978; Murray, 1984)
		Class bacilli	TAS (Euzéby, 2010; Ludwig et al., 2009)
		Order *Bacillales*	TAS (Prevot, 1953; Skerman et al., 1980)
		Family *Bacillaceae*	TAS (Fischer, 1894; Skerman et al., 1980)
		Genus *Bacillus*	TAS (Cohn, 1872; Fischer, 1894; Skerman et al., 1980)
		Species *Bacillus smithii*	TAS (Nakamura et al., 1988)
		Type strain: DSM 4216^T	
	Gram stain	Positive*	TAS (Nakamura et al., 1988)
	Cell shape	Rod	IDA (Fig. 1), TAS (Nakamura et al., 1988)
	Motility	Motile	TAS (Nakamura et al., 1988)
	Sporulation	Terminal or sub terminal, oval or cylindrical endospores, non-swollen to slightly swollen sporangia	IDA (Fig. 1), TAS (Nakamura et al., 1988)
	Temperature range	25-65°C	TAS (Nakamura et al., 1988)
	Optimum temperature	55°C	IDA
	pH range; Optimum	5.5-6.8; 6.5	TAS (Nakamura et al., 1988), IDA

Table 1 continued.

	Carbon source	D-glucose, D-xylose, L-xylose, L-arabinose, D-ribose, glycerol, D-adonitol, D-fructose, L-sorbose, D-galactose, L-rhamnose, inositol, D-mannitol, sucrose, D-trehalose, xylitol, Methyl-α-D-glucopyranoside, esculin, salicin, D-maltose, D-turanose, D-lyxose, D-tagatose, D-arabitol, K-gluconate, K-5-ketogluconate	IDA(API), TAS (Nakamura et al., 1988)
MIGS-6	Habitat	Type strain: cheese. Other strains: evaporated milk, canned food, compost, hot spring soil, sugar beet juice from extraction installations.	TAS (Jurado et al., 2014; Messner et al., 1997; Nakamura et al., 1988; Qi et al., 2011)
MIGS-6.3	Salinity	Not in 3% NaCl (w/v)	TAS (Nakamura et al., 1988)
MIGS-22	Oxygen requirement	Facultative anaerobe	TAS (Nakamura et al., 1988)
MIGS-15	Biotic relationship	Free-living	TAS (Nakamura et al., 1988)
MIGS-14	Pathogenicity	Non-pathogen	TAS (Lücking et al., 2013; Suitso et al., 2014)
MIGS-4	Geographic location	USA	TAS (Gordon & Smith, 1949; Nakamura et al., 1988)
MIGS-5	Sample collection	~1946	TAS (Gordon & Smith, 1949; Nakamura et al., 1988)
MIGS-4.1	Latitude	Unknown	
MIGS-4.2	Longitude	Unknown	
MIGS-4.4	Altitude	Unknown	

a Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project.
* As described in the species description by Nakamura et al.: “Young cells of both groups were Gram positive. With increasing age the cells became Gram variable and finally Gram negative. The KOH and aminopeptidase tests were negative, as is typical for gram-positive organisms.”

moderately thermophilic or thermotolerant species (Han et al., 2013; Nakamura et al., 1988).

The *B. smithii* type strain DSM 4216^T was isolated from cheese (Hammer, 1915; Nakamura et al., 1988), but other *B. smithii* strains have been isolated from compost (Bosma et al., 2015a; Jurado et al., 2014), hot spring soil (Qi et al., 2011), and a sugar beet factory (Messner et al., 1997). It is a free-living organism that was shown to be non-cytotoxic (Lücking et al., 2013). In addition, the safety of the probiotic *B. smithii* TMBI 12 was recently reported in piglets studies (Suitso et al., 2014). Basic morphological and physiological features have been described by Nakamura *et al.*(Nakamura et al., 1988). Genetic accessibility, a wide temperature and pH range and the ability to utilize a wide range of carbon sources in a relatively minimal medium make *B. smithii* an interesting new host for biotechnological applications (Bosma et al., 2015a; Bosma et al., 2015b).

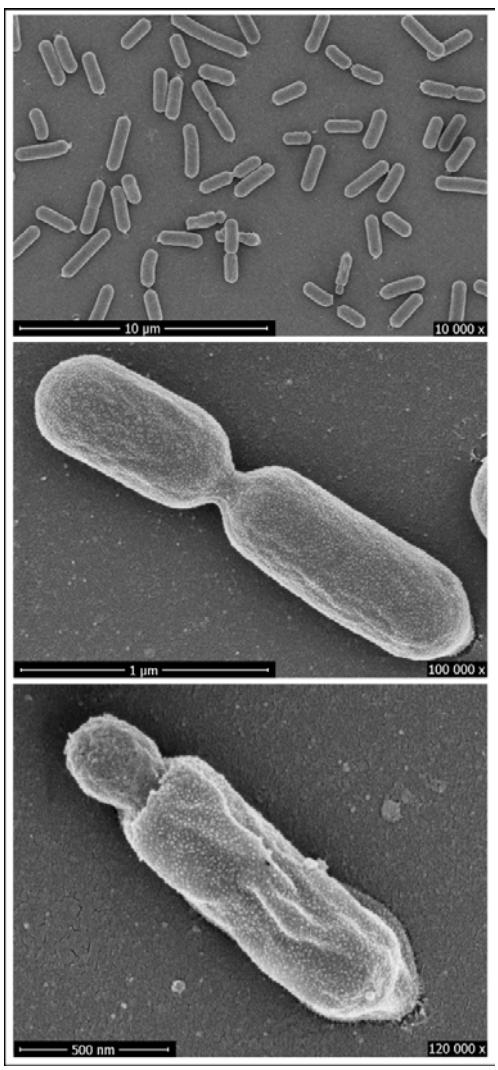

Figure 1. Scanning electron micrographs of *B. smithii* DSM 4216^T.

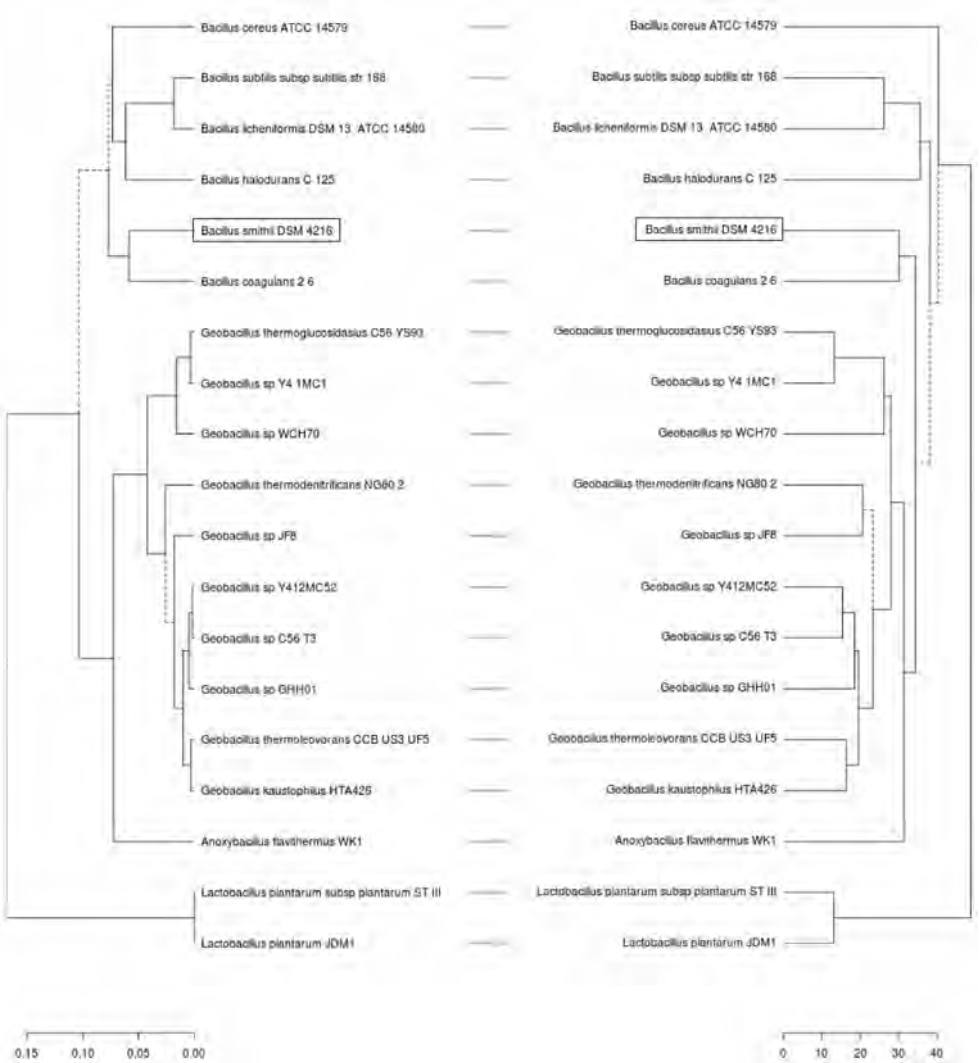

Figure 2. Phylogenetic tree based on 16S rRNA gene sequences (left) and protein domains (right). A comparison is included (horizontal lines) between the two trees, showing the position of *Bacillus smithii* DSM 4216^T relative to other *Bacillus* strains, as well as several industrially important Lactic Acid Bacterium strains. Only strains were used for which a complete genome sequence is available (as on 18 September 2014) in order to be able to perform the domain-based analysis. The 16S sequences were aligned using DECIPHER (R) (Wright, 2013) and the distance analysis was performed using a Jukes-Cantor correction. Phylogenetic analysis of all domains was performed by re-annotation of all proteins from selected genomes using InterProScan 5-RC7 and transformed into a absence-presence matrix. Distance was calculated using a standard Euclidean distance and clustering was performed by complete method using hclust. Tree comparison was performed by dendextend.

GENOME SEQUENCING INFORMATION

Genome project history

The *B. smithii* type strain was selected based on the biotechnological relevance of the species as described above. The initial Illumina sequencing was performed in March 2012 and the genome was closed by PacBio sequencing in June 2013. The final, closed genome sequence consisting of 1 chromosome and 1 plasmid was deposited in GenBank (nr CP012024 and CP012025) and released for public access on 8 July 2015. A summary of the project information and its association with MIGS version 2.0 compliance (Field et al., 2008) is shown in Table 2.

Growth conditions and genomic DNA preparation

B. smithii DSM 4216^T was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). DNA was isolated from *B. smithii* DSM 4216^T cultures grown overnight at 55°C in 100 mL LB2 and TVMY-glucose (Bosma et al., 2015a) in a 250 mL Erlenmeyer. 10 mL of the cultures was harvested by centrifugation for 15 min at 4°C and 4816 x *g*, after which DNA was isolated using the Epicentre Master Pure Gram Positive DNA Purification kit according to the manufacturer's protocol. DNA integrity was confirmed on a 1.0 % agarose gel and concentrations were measured using Qubit (Life Technologies), after which DNA integrity was re-evaluated by the sequencing company before sequencing.

Table 2. Project information of the whole genome sequence of *B. smithii* DSM 4216^T.

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	Mate-pair (average 4,260 bp), paired-end (average 273 bp), PacBio (2,075 and 2,775 kbp)
MIGS 29	Sequencing platforms	Illumina and PacBio
MIGS 31.2	Fold coverage	Illumina paired-end: 187x, Illumina mate pair: 311x, PacBio: 56x
MIGS 30	Assemblers	CLCbio Genomics Workbench 5.0, SSPACE Premium 2.0, GapFiller 1.10
MIGS 32	Gene calling method	RAST and protein domain analysis
	Locus Tag	BSM4216
	Genbank ID	CP012024 (chromosome) and CP012025 (plasmid)
	GenBank Date of Release	8 July 2015
	BIOPROJECT	PRJNA258357
MIGS 13	Source Material Identifier	Biotechnological
	Project relevance	DSM 4216^T

Genome sequencing and assembly

The genome of *B. smithii* DSM 4216^T was sequenced by BaseClear BV (NL) using Illumina HiSeq2000 mate-pair and paired-end sequencing for the initial sequencing and assembly, followed by PacBio sequencing to fully close the genome sequence. The average length of the paired-end samples was 273 bp and that of the mate-pair samples 4,260 bp. The sequence reads were filtered and trimmed based on Phred quality scores, assembled into contigs using the "De Novo Assembly" option of the CLCbio Genomics Workbench version 5.0 and further assembled into scaffolds using SSPACE Premium version 2.0 (Boetzer et al., 2011). This initial sequencing resulted in 6,185,516 reads, which were assembled into 214 contigs and 27 scaffolds. The coverage of the paired-end reads was 187x and that of the mate pair reads was 311x. For gap closure, sequencing was performed using a PacBio SMRT cell and quality was again assessed based on Phred scores. PacBio sequencing resulted in 90,013 reads with an average read length of 2,075 kbp and a coverage of 56x. The contigs were assembled into super-scaffolds using alignment of the PacBio reads with BLASR (Chaisson & Tesler, 2012), which was then used to determine the order of and distance between the contigs using a modified SSPACE Premium version 2.3 (Boetzer et al., 2011). This resulted in 5 scaffolds, after which a second PacBio run was performed, which resulted in 114,294 reads with an average length of 2,775 kbp. These results were analyzed in the same way as the first PacBio-round, after which gaps in the super-scaffolds were closed using GapFiller 1.10 (Boetzer & Pirovano, 2012), resulting in the final genome of 1 chromosome and 1 plasmid. Two small scaffolds (<450 bp) were found to be contaminants and removed from the data set.

Structural variations (SVs; small nucleotide polymorphism and small insertions and deletions) in the paired end and mate paired Illumina reads were compared to the PacBio scaffolds at the CMBI Nijmegen using an in-house developed tool RoVar (S. A. F. T. van Hijum, V. C. L. de Jager, B. Renckens, and R. J. Siezen, unpublished data; http://trac.nbic.nl/rovar). Repeat masking of the reference sequence was done by (i) creating 30-bp fragments, (ii) aligning these fragments to the PacBio reference sequence by using BLAT (Kent, 2002) with a tile size of 6, and (iii) masking regions to which fragments align perfectly in multiple positions in the reference sequence. Illumina read alignment performed by BLAT with a tile size of 6 and alignment events were allowed provided that SVs were at least 4 bp from the end of a given read. SVs were used for further analysis provided that they were supported by at least 5 unique forward and 5 unique reverse reads and at most 1% of the reads were allowed to suggest an alternative allele. A total of 14 SVs were corrected in the *B. smithii* 4216^T PacBio assembly.

Genome annotation

The corrected PacBio assembly was subjected to RAST annotation (Aziz et al., 2008) using default parameters. The following tools were used to predict gene functions (Table 4): Aragorn for tRNAs (http://mbio-serv2.mbioekol.lu.se/ARAGORN/), RNAmmer (http://www.cbs.dtu.dk/services/RNAmmer/) for rRNA, and CRISPR-finder for CRISPR repeats and spacers (http://crispr.u-psud.fr/Server/). The annotation was manually curated by running a BLAST of all genes and comparing starts and stops to the best hits. Via this method, also pseudogenes were manually identified.

As several pathways commonly found in bacilli were not identified by RAST in *B. smithii*, an analysis based on protein domains was performed on the *B. smithii* DSM 4216^T genome using InterProScan 5 (version 5RC7, 27th January 2014) (Koehorst & Van Dam, in prep.). This has been shown to be a powerful tool for identifying previously unknown protein functions, for example in determining microbial syntrophic interactions (Worm et al., 2014). The domain-based annotation was compared to the manually curated RAST annotation, after which duplicates were removed and genes identified uniquely via the domain-analysis were added. In total 142 extra genes were annotated via this method, of which all except 4 were hypothetical proteins. For 209 genes, the protein domain annotation resulted in the addition of EC-numbers to the annotation that had not been assigned via RAST.

Genome Properties

The genome of *B. smithii* DSM 4216^T consists of a circular chromosome of 3,368,778 base pairs with a GC content of 40.8% and a plasmid of 12,514 base pairs and a GC content of 35.9% (Table 3). Figure 3 shows a map of the DSM 4216^T plasmid and chromosome. On the chromosome, a total of 3880 genes were identified, of which 3627 were annotated as protein-coding genes, of which 81 are assigned 'putative' or 'probable' functions, 1472 are hypotheticals or genes with unknown function (38.2%) and the remaining had a defined function. Out of the total chromosomal genes, 126 genes are pseudogenes and 94 are tRNAs, 33 are rRNA genes, 122 are genes with signal sequences for secretion and 795 are genes with a transmembrane domain (Table 4). The rRNA genes are clustered in 11 operons, which is relatively many and is thought to be linked to the capacity to grow fast in different conditions (Lee et al., 2009). Eight of these operons were found on the forward strand and 3 on the reverse strand. Six of the operons appear to be positioned approximately opposite of each other on the two strands, while the remaining five are located very closely to the origin and to each other on the forward strand. The plasmid DNA was predicted to contain 18 genes, of which 5 have a function assigned, 11 are hypotheticals and 2 are mobile element associated proteins. The COG-distribution of genes is shown in Table 5.

Table 3. Summary of the *B. smithii* DSM 4216^T genome: one chromosome and one plasmid.

Label	Size (Mb)	Topology	GenBank accession nr
Chromosome	3,368,778	Circular	CP012024
Plasmid	12,514	Circular	CP012025

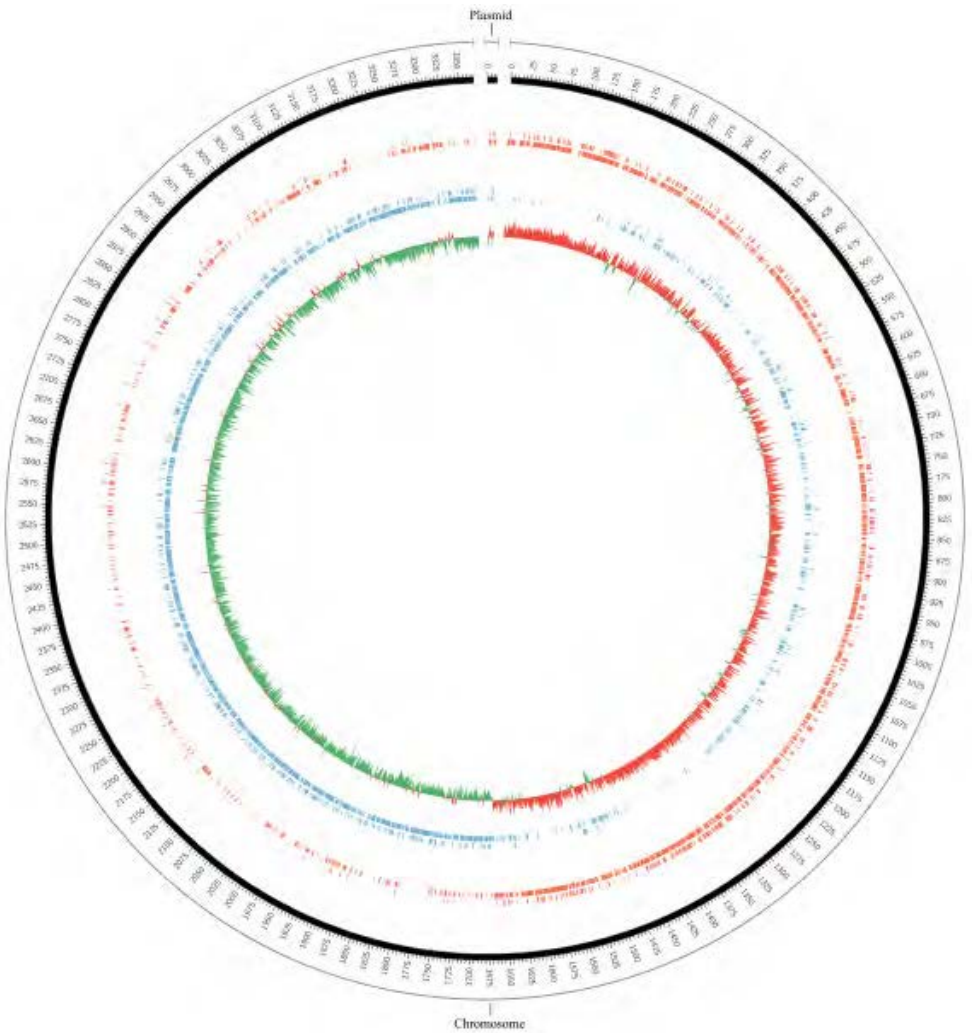

Figure 3. Chromosome and plasmid map of *B. smithii* DSM 4216^T. The outer circle represents base pair numbers; red are genes on the forward strand and blue on the reverse; the inner circle represents GC skew in which red is a positive GC content and green a negative.

Attribute	Value	% of Total
Genome size (bp)	3,381,292	100.0
DNA coding (bp)	2,799,365	82.8
DNA G+C (bp)	1,378,026	40.8
DNA scaffolds	2	
Total genes	3880	100.0
Protein coding genes	3627[a]	93.5
RNA genes	127	3.3
Pseudo genes	126	3.2
Genes in internal clusters	ND	
Genes with function prediction	2063	53.1
Genes assigned to COGs	2619	67.4
Genes with Pfam domains	2596	66.8
Genes with signal peptides	122	3.1
Genes with transmembrane helices	795	20.5
CRISPR repeats	69	

Table 4. Genome statistics of *B. smithii* DSM 4216^T.
[a] This is excluding 126 pseudogenes.

Code	Value	%age	Description
J	162	4.46	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	179	4.92	Transcription
L	160	4.40	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	28	0.77	Cell cycle control, Cell division, chromosome partitioning
V	31	0.85	Defense mechanisms
T	125	3.44	Signal transduction mechanisms
M	132	3.63	Cell wall/membrane biogenesis
N	64	1.76	Cell motility
U	42	1.16	Intracellular trafficking and secretion
O	92	2.53	Posttranslational modification, protein turnover, chaperones
C	156	4.29	Energy production and conversion
G	174	4.79	Carbohydrate transport and metabolism
E	291	8.01	Amino acid transport and metabolism
F	74	2.04	Nucleotide transport and metabolism
H	107	2.94	Coenzyme transport and metabolism
I	94	2.59	Lipid transport and metabolism
P	154	4.24	Inorganic ion transport and metabolism
Q	70	1.93	Secondary metabolites biosynthesis, transport and catabolism
R	382	10.51	General function prediction only
S	236	6.49	Function unknown
-	1321	36.34	Not in COGs

Table 5. Number of genes associated with general COG functional categories.
The total is based on the total number of protein coding genes in the RAST-annotated genome (3635) with an e-value cut-off of 0.00001.

COMPARISON TO OTHER GENOMES

As the number of available genome sequences from thermophilic bacilli is still rather limited and *B. smithii* also grows at mesophilic temperatures, we compared its genome properties to those of thermophilic bacilli as well as to those of several commonly studied mesophilic bacilli (Table 6). Compared to its close relative *B. coagulans*, *B. smithii* has a slightly larger genome with a lower GC content. Compared to most mesophilic bacilli, its genome is smaller and it has a higher GC content than *B. cereus* but lower than *B. halodurans* and *B. subtilis*. As will be discussed in the next section, the genome content differs from its close relatives in several ways.

Table 6. Comparison of several published complete genome sequences of the genus *Bacillus*.

Species / strain	Genome size (bp)	GC %[1]	ORFs[2]	Plasmid number	Growth[3]	Reference
B. smithii DSM 4216^T	3,368,778	40.8	3635	1	TT	This study
B. coagulans DSM1^T*	3,018,045	47.2	3437	0	TT	(Su et al., 2012)
B. coagulans 36D1	3,552,226	46.5	3306	0	TT	(Rhee et al., 2011)
B. coagulans 2-6	3,073,079	47.3	2985	1	TT	(Su et al., 2011a)
A. flavithermus WK1	2,846,746	41.8	2863	0	TT	(Saw et al., 2008)
B. licheniformis 10-1	4,317,010	45.9	4650	0	TT	(Li et al., 2012; Li et al., 2013)
B. licheniformis DSM13^T	4,222,748	46.2	4286	0	TT	(Veith et al., 2004)
B. cereus ATCC 14579	5,426,909	35.3	5366	1	MP	(Ivanova et al., 2003)
B. halodurans C-125	4,202,353	43.7	4066	0	MP	(Takami et al., 2000)
B. subtilis 168^T	4,214,810	43.5	4100	0	MP	(Kunst et al., 1997)
G. thermoglucosidans TNO-09.020*	3,75 Mb	43.9	4300	0	TP	(Zhao et al., 2012b)
G. thermodenitrificans NG80-2	3,550,319	48.9	3499	1	TP	(Feng et al., 2007)
G. kaustophilus HTA426	3,544,776	52.0	3498	1	TP	(Takami et al., 2004a; Takami et al., 2004b)
G. thermoleovorans CCB_US3_UF5	3,596,620	52.3	3887	0	TP	(Sakaff et al., 2012)

Currently available thermophilic *Bacillus* genomes are shown, as well as a selection of genomes of mesophilic model organisms.
*Sequence not fully closed.
[1] GC% of chromosome and plasmid combined as weighted average.
[2] Open Reading Frames as a total on chromosome and plasmid(s).
[3] MP: mesophile, TP: thermophile, TT: thermotolerant (grows at mesophilic as well as thermophilic temperatures).

EXTENDED INSIGHTS

CENTRAL CARBON METABOLISM AND MAIN PRODUCT PATHWAYS

To be able to use *B. smithii* as a host for biotechnological purposes, it is important to understand its metabolic pathways. In the *B. smithii* DSM 4216^T genome, all genes involved in glycolysis, gluconeogenesis, pentose phosphate pathway, TCA-cycle and glyoxylate shunt could be identified, but not the complete sets of genes for the phosphoketolase and Entner-Doudoroff pathways. Uptake systems for all sugars shown to support growth in the API-test were annotated by the RAST annotation. The organization of the xylose catabolic operon is similar to that found in *B. coagulans* XZL4 (Su & Xu, 2014). A reconstruction of the central carbon metabolism of *B. smithii* DSM 4216^T is shown in Figure 4. An L-lactate dehydrogenase gene was annotated, which is in accordance with L-lactate being the major fermentation product of *B. smithii* (Bosma et al., 2015a; Bosma et al., 2015b). After RAST annotation, the methylglyoxal pathway was identified only towards D-lactate, but an in-depth analysis of protein domains also revealed the presence of all genes necessary for L-lactate production via methylglyoxal. Based on 16S rRNA gene and complete protein domain analysis (Figure 2), the closest relative of *B. smithii* is *B. coagulans*. However, when reconstructing the metabolic network of *B. smithii*, several remarkable differences between *B. smithii* and *B. coagulans* as well as other bacilli were observed. The most striking difference with bacilli in general is the absence of the genes coding for phosphotransacetylase and acetate kinase, which form the standard acetate production pathway in bacteria. This was confirmed by the domain-based analysis. Moreover, we also could not identify these two genes in the genome sequence of *B. smithii* strain 7_3_47FAA, which is available from a metagenome database. The fact that *B. smithii* produces significant amounts of acetate from glucose (Bosma et al., 2015a; Bosma et al., 2015b) indicates that an alternative pathway is involved, which is currently being investigated. Furthermore, candidate genes for pyruvate formate lyase, pyruvate decarboxylase and pyruvate oxidoreductase could not be found in the genomes of both DSM 4216^T and 7_3_47FAA via either RAST or domain-based analysis. Therefore, *pdhc*-encoded pyruvate dehydrogenase complex is most likely the only enzyme responsible for the conversion of pyruvate to acetyl-CoA. This is confirmed by a *pdhA*-knockout strain of *B. smithii* strain ET 138, which is unable to grow without acetate supplementation and did not produce any acetate (Bosma et al., 2015b).

Another difference with *B. coagulans* is the lack of a catabolic *alsSD*-operon in *B. smithii*, coding for the enzymes acetolactate synthase and acetolactate decarboxylase. This is in accordance with the absence of 2,3-butanediol production (Bosma et al., 2015a; Bosma et al., 2015b). The anabolic acetolactate synthase small and large subunit genes *ilvBH* (also called α-acetohydroxyacid synthase) are present. These genes are mainly involved in the

isoleucine and valine biosynthetic pathways (Blomqvist et al., 1993). On the other hand, both an S- and an S/R-acetoin specific 2,3-butanediol dehydrogenase gene were identified in the genome. Although several alcohol dehydrogenases were found in the genome, no bifunctional acetaldehyde dehydrogenase-alcohol dehydrogenase *adhE* could be found, which is in accordance with the absence of alcohol production in the majority of *B. smithii* fermentations (Bosma et al., 2015a).

Amino acid and vitamin biosynthesis pathways

Microorganisms used for biotechnological purposes should have minimal nutrient requirements, as the addition of yeast extract, vitamins or amino acids is costly. The organisms should therefore preferably contain the pathways to synthesize vitamins, amino acids, purines and pyrimidines. In *B. smithii* DSM 4216^T, all amino acid biosynthetic pathways could be identified. Pathways for *de novo* synthesis and salvage pathways of pyrimidines and purines were also identified. Complete vitamin biosynthesis pathways were identified for cobalamin, riboflavin, tetrahydrofolate, panthothenate, *p*-aminobenzoic acid, nicotinic acid and pyridoxal, but not for thiamine, ascorbate, pyridoxamine and D-biotin.

Host-defense systems

Robustness against infection is crucial for industrial microorganisms. Host-defense systems can confer such robustness, but might also hinder genetic accessibility of the organism. In the genome of *B. smithii* DSM 4216^T, several host-defense systems are annotated: a type II-s restriction endonuclease, a 5-methylcytosine-specific restriction related enzyme, a type I restriction-modification system and a CRISPR-Cas Type I-B system. The CRISPR-Cas genes show the typical type I-B gene arrangement (Makarova et al., 2011), but seem to be partly duplicated around the CRISPR locus with a second locus containing cas6, cas8a/cst1, cas7 (originally annotated as 'CRISPR-associated negative autoregulator) and cas5 after the CRISPR repeats. The CRISPR-finder tool (http://crispr.u-psud.fr/Server/) was used to identify CRISPR repeats and spacers in the area around the Cas-genes (bp 2772458-2798619). Three CRISPR-loci were identified (CRISPR 1, 3 and 4) as well as one questionable locus (CRISPR 2). Using CRISPRTarget (bioanalysis.otago.ac.nz/ CRISPRTarget/), some of the spacers were found to have hits with potential protospacer target sequences against *Bacillus sp.* and *B. subtilis* plasmid DNA, and against *S. thermophilus*, *Lactococcus*, *Enterococcus* and *Campylobacter* phage DNA.

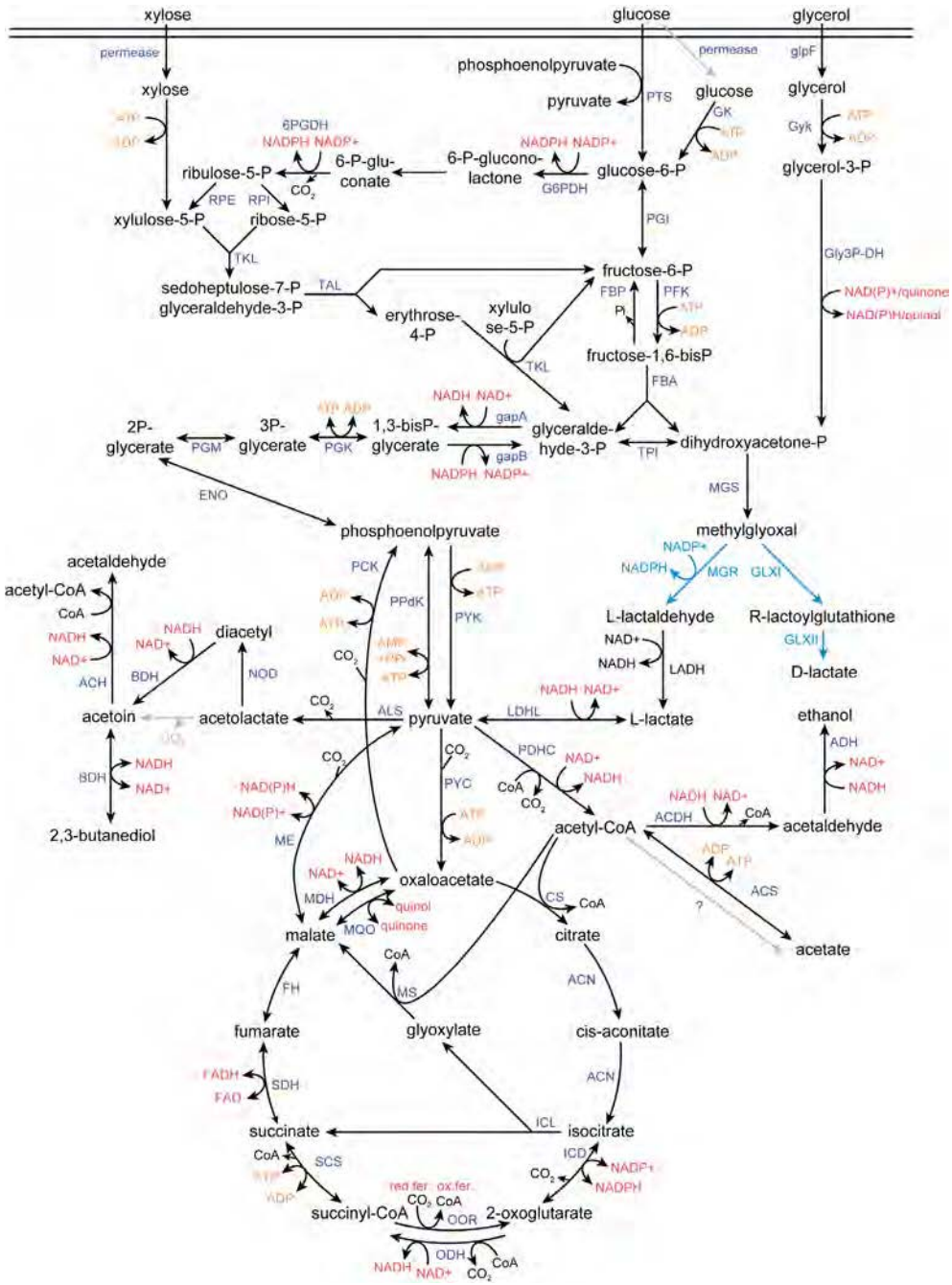

Figure 4. Reconstruction of central carbon metabolism of *B. smithii* DSM 4216^T.
Legend see next page. A similar figure can be found in the seperate sheet in this thesis.

Figure 4. Reconstruction of central carbon metabolism of *B. smithii* DSM 4216^T. Blue lines indicate pathways of which the EC-number was identified only via domainome analysis; grey lines indicate pathways unidentified by both RAST annotation and domainome analysis.
Abbreviations: XI: xylose isomerase; XK: xylulokinase; PTS: phosphotransferase system; GK: glucokinase; glpF: glycerol facilitator; glyK: glycerol kinase; Gly3P-DH: glycerol-3-phosphate dehydrogenase; PGI: glucose-6-phosphate isomerase; G6PDH: glucose-6-phosphate dehydrogenase; 6PGDH: 6-phosphogluconate dehydrogenase; RPI: phosphopentose isomerase; RPE: phosphopentose epimerase; TKL: transketolase; TAL: transaldolase; FBP: fructose bisphosphatase; PFK: phosphofructokinase; FBA: fructose bis-phosphate aldolase; TPI: triosephosphate isomerase; GAP: glyceraldehyde 3-phosphate dehydrogenase; PGK: phosphoglycerate kinase; PGM: phosphoglycerate mutase; ENO: enolase; PCK: phosphoenol pyruvate carboxykinase; PPC: phosphoenol pyruvate carboxylase; PYK: pyruvate kinase; PYC: pyruvate carboxylase; PDHC: pyruvate dehydrogenase complex; ME: malic enzyme; MDH: malate dehydrogenase; MQO: malate:quinone oxidoreductase; CS: citrate synthase; ACN: aconitase; ICL: isocitrate lyase; MS: malate synthase; ICD: isocitrate dehydrogenase; OOR: 2-oxoglutarate reductase; ODH: 2-oxoglutratae dehydrogenase; SCS: succinyl-CoA synthetase; SDH: succinate dehydrogenase; FH: fumarate hydratase; ALS: acetolactate synthase; NOD: non-enzymatic oxidative decarboxylation; BDH: butanediol dehydrogenase; ACH: acetoin dehydrogenase; LDHL: L-lactate dehydrogenase; ACDH: acetyl-CoA dehydrogenase; ADH: alcohol dehydrogenase; ACS: acetyl-CoA synthetase; MGS: methylglyoxal synthase; MGR: methylglyoxal reductase; GLXI: glyoxalase I; GLXII: glyoxalase II; LADH: lactaldehyde dehydrogenase.

CONCLUSIONS

This report describes the complete genome sequence of *Bacillus smithii* type strain DSM 4216^T. The species has biotechnological potential due to its efficient conversion of both C_5 and C_6 sugars at 55°C to lactic acid, combined with its genetic accessibility. Its central carbon metabolism is different from its close relative *B. coagulans* as it lacks the *alsSD* operon, as well as the *pta-ack* acetate production pathway and the *pfl* gene.

AUTHORS' CONTRIBUTIONS

EFB and AHPvdW performed the microbiology and molecular biology studies. BV, RvK, JJK, PJS, SAFTvH and BR performed bio-informatics analyses. EFB, RvK and BV were involved in the preparation and submission of the genome to NCBI. EFB drafted the manuscript, with input and revisions from RvK, JJK, PJS, SAFTvH, WdV, JvdO and RvK. All authors participated in the design of the study, which was coordinated by RvK, WdV and JvdO. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This project was financially supported by Corbion (NL).

Chapter 5

Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains

This chapter has been published as:
Bosma, E.F., van de Weijer, A.H.P., van der Vlist, L., de Vos, W.M., van der Oost, J., van Kranenburg, R. 2015. Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains. *Microbial Cell Factories*, 14(1), 99.

ABSTRACT

Microbial conversion of biomass to fuels or chemicals is an attractive alternative for fossil-based fuels and chemicals. Thermophilic microorganisms have several operational advantages as a production host over mesophilic organisms, such as low cooling costs, reduced contamination risks and a process temperature matching that of commercial hydrolytic enzymes, enabling simultaneous saccharification and fermentation at higher efficiencies and with less enzymes. However, genetic tools for biotechnologically relevant thermophiles are still in their infancy. In this study we developed a markerless gene deletion method for the thermophile *Bacillus smithii* and we report the first metabolic engineering of a potential platform organism. Clean deletions of the *ldhL* gene were made in two *B. smithii* strains (DSM 4216^T and compost isolate ET 138) by homologous recombination. Whereas both wild-type strains produced mainly L-lactate, deletion of the *ldhL* gene blocked L-lactate production and caused impaired anaerobic growth and acid production. To facilitate the mutagenesis process, we established a counter-selection system for efficient plasmid removal based on *lacZ*-mediated X-gal toxicity. This counter-selection system was applied to construct a sporulation-deficient *B. smithii* Δ*ldhL* Δ*sigF* mutant strain. Next, we demonstrated that the system can be used repetitively by creating *B. smithii* triple mutant strain ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*, from which also the gene for the α-subunit of the E1 component of the pyruvate dehydrogenase complex is deleted. This triple mutant strain produced no acetate and is auxotrophic for acetate, indicating that pyruvate dehydrogenase is the major route from pyruvate to acetyl-CoA. In this study, we developed a markerless gene deletion method including a counter-selection system for thermophilic *Bacillus smithii*, constituting the first report of metabolic engineering in this species. The described markerless gene deletion system paves the way for more extensive metabolic engineering of *B. smithii*. This enables the development of this species into a platform organism and provides tools for studying its metabolism, which appears to be different from its close relatives such as *B. coagulans* and other bacilli.

KEYWORDS

Bacillus smithii; thermophile; lactate dehydrogenase; sporulation; pyruvate dehydrogenase; counter-selection system

INTRODUCTION

Microbial conversion of biomass to fuels such as ethanol or hydrogen, or to green chemical building blocks such as organic acids has gained increasing attention over the last decade (Bozell & Petersen, 2010; Werpy et al., 2004). Thermophilic microorganisms have several advantages over mesophilic organisms for use as microbial production hosts. Fermentation at high temperatures lowers cooling costs and contamination risks and increases product and substrate solubility (Bhalla et al., 2013; Ma et al., 2014; Studholme, 2014). Moreover, the optimum temperature of moderate thermophiles matches that of commercial hydrolytic enzymes, enabling simultaneous saccharification and fermentation at higher efficiencies and with less enzymes compared to mesophilic bacteria (Ou et al., 2009).

Despite the aforementioned advantages of thermophiles, mesophilic model organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae* are still preferred production organisms, as these are well-studied and genetic tools are available to enable their use as versatile platform organisms (Chen et al., 2013; Hong & Nielsen, 2012). Genetic tools for biotechnologically relevant thermophiles are recently emerging for different species, but most are still in their infancy or highly strain-specific. Several strictly and facultatively anaerobic thermophiles have been engineered for green chemical and fuel production, as has been reviewed recently (Bosma et al., 2013; Taylor et al., 2011). Most engineering efforts in thermophiles have so far been directed at ethanol production, but recently also examples for chemical production have been shown such as *Thermoanaerobacterium aotearoense* for lactate production (Yang et al., 2013), *Bacillus licheniformis* for 2,3-butanediol production (Wang et al., 2012b), and *Bacillus coagulans* for D-lactate production (Kovacs et al., 2010; Wang et al., 2011a). The development of genetic tools for thermophilic organisms is crucial to fully understand their metabolic versatility and to establish a thermophilic production platform for green chemical and fuel production. For industrial applications, markerless gene deletions should be made such that no antibiotic resistance genes or other scars are introduced into the target genome. This is especially important when working with thermophilic organisms as the number of available markers is limited, requiring re-use of the marker (Bosma et al., 2013; Taylor et al., 2011).

Recently, we isolated a thermophilic *Bacillus smithii* strain capable of degrading C_5 and C_6 sugars at a wide range of temperatures and pHs (Bosma et al., 2015a) and demonstrated electrotransformation of several *Bacillus smithii* strains with plasmid pNW33n. In the current study, we developed a clean gene deletion method and counter-selection system for this species and applied this to create multiple markerless gene deletions both in the previously isolated *B. smithii* ET 138 (Bosma et al., 2015a) and in the type strain *B. smithii* DSM 4216^T.

RESULTS

CONSTRUCTION OF MARKERLESS *LDHL* DELETION MUTANTS

B. smithii ET 138 can be transformed with *E. coli-Bacillus* shuttle vector pNW33n with an efficiency of $5x10^3$ colonies per µg DNA (Bosma et al., 2015a). To obtain mutants in strain ET 138, we planned to use a protocol similar to that used for *Geobacillus thermoglucosidans* (recently renamed from *G. thermoglucosidasius* (Coorevits et al., 2012)), which applies pNW33n-derivatives as thermosensitive integration plasmid (Cripps et al., 2009). To create a markerless L-lactate dehydrogenase (*ldhL*) knockout strain from which the *ldhL* gene was entirely deleted, ~1000 bp regions flanking the *ldhL* gene and including the start and stop codon were cloned and fused together in plasmid pNW33n. Double homologous recombination of this plasmid with the ET 138 chromosome will fuse the start and stop codons of the gene, thereby removing the entire gene in-frame without leaving any marker (Figure 1). *B. smithii* ET 138 was transformed with pWUR732 and colonies were transferred once at 55°C on LB2 plates containing chloramphenicol. Subsequent PCR analysis of 7 colonies already showed integration of the plasmid DNA without the temperature increase normally performed with thermosensitive integration systems (Cripps et al., 2009). A mixture of single crossover integrants via both upstream and downstream regions together with no-integration (either caused by replicating plasmids or randomly integrated plasmids) genotype was observed in 1 colony, 1 colony showed a mixture of downstream crossover and wild-type genotype, and 5 colonies showed no single crossovers but only wild-type genotype. Serial transfer of the colonies containing single crossovers in liquid medium combined with replica plating to identify double recombinants repeatedly resulted in only wild-type double crossover mutants. The mixed genotype persisted after several subculturings on plates containing 7 and 9 µg/mL chloramphenicol in an attempt to obtain pure genotypes. After 4 transfers, however, also a colony was found that contained a mixture of double crossover knockout genotype together with upstream single crossover and wild-type genotype. After this point, we added glycerol or acetate as carbon sources to allow for a metabolism with minimal impact of the *ldhL* deletion. After streaking this colony to an LB2 plate containing 10 g/L glycerol, colonies were obtained that had lost the wild-type genotype but contained a mixture of both single crossovers and a double crossover knockout genotypes. A pure double crossover knockout genotype was observed after two transfers on the more defined TVMY medium supplemented with acetate at 65°C, creating strain ET 138 Δ*ldhL* (Figure 2A).

Similar to strain ET 138, also type strain *B. smithii* DSM 4216^T is transformable with pNW33n, with efficiencies of around $2x10^2$ colonies per µg DNA (Bosma et al., 2015a). After transformation of strain DSM 4216^T with *ldhL*-knockout construct pWUR733 and transfer of

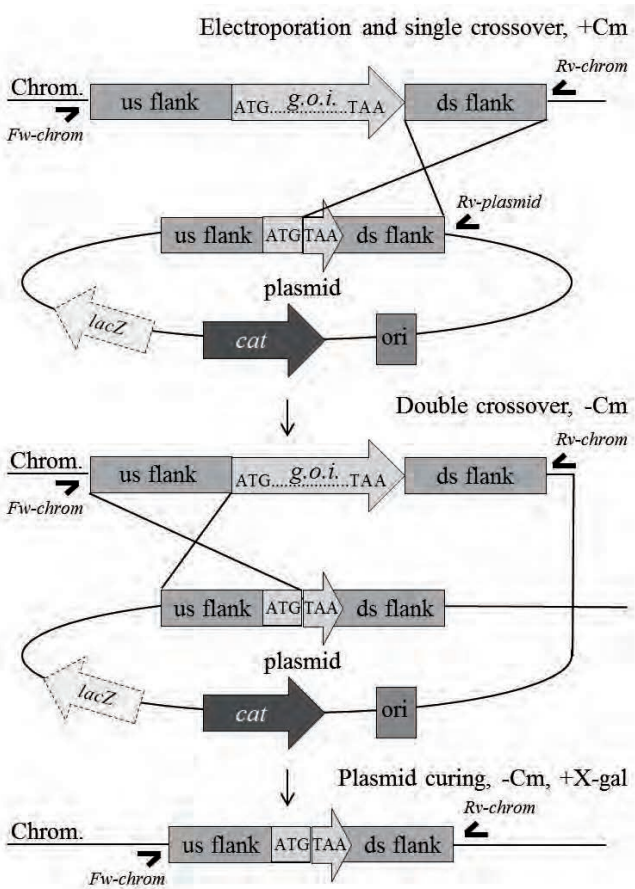

Figure 1. Illustration of markerless knockout construction in *B. smithii*. A pNW33n-derived plasmid containing the 1000 bp-flanking regions of a gene of interest ('*g.o.i.*') is introduced into the cell via electroporation, plated on LB2 containing chloramphenicol at 55°C and subjected to PCR analysis to check for single crossovers using primers 'Fw-chrom' and 'Rv-plasmid'. Colonies containing single crossovers are transferred to plates without chloramphenicol, after which PCR screening is performed to check for double crossovers using primers 'Fw-chrom' and 'Rv-chrom'. If the 2nd recombination occurs via the same flank as the 1st recombination, a wild-type genotype will be the result. If the 2nd recombination occurs via the other flank, this results in a knockout genotype. To delete the *ldhL* gene, only PCR screening was performed and no *lacZ* gene was present on the plasmid. For creating the *sigF* and *pdhA* mutants, the plasmid also contained the *lacZ* gene. In those cases, *lacZ* counter-selection was performed by plating an overnight culture on LB2 plates containing 100 µg/mL X-gal, resulting in toxic concentrations of the X-gal cleavage product in the presence of *lacZ,* resulting in small blue colonies still containing the plasmid (either replicating plasmid or inserted in the genome via single crossover) and white colonies that have cured the plasmid. Abbreviations: *g.o.i.*: gene of interest; *cat*: chloramphenicol acetyltransferase; chrom.: chromosome; us: upstream; ds: downstream; Cm: chloramphenicol; primers are indicated with black hooks.

colonies to a new plate at 55°C, PCR on 30 colonies showed mixtures of wild-type genotype and both single crossovers for all tested colonies. Contrary to what was observed with strain ET 138, several subculturings on TVMY supplemented with acetate at temperatures varying between 55°C and 65°C did not result in a pure mutant genotype for derivatives of DSM 4216^T. However, after the substrate was changed from acetate to lactate, a pure double-crossover knockout colony was obtained after two transfers, creating strain DSM 4216^T Δ*ldhL* (Figure 2A).

ESTABLISHMENT OF A *LACZ*-COUNTER-SELECTION SYSTEM

Purifying the mixtures of plasmid integrated into the *B. smithii* genome via single and double crossovers during the construction of the *ldhL*-mutant strains required laborious PCR-screening. To simplify the screening procedure, a counter-selection tool was desirable to select against plasmid presence. A *lacZ*-counter-selection method has been described for the Gram-negative mesophile *Paracoccus denitrificans*, which is based on toxicity of high X-gal concentrations in the presence of β-galactosidase activity encoded by a *lacZ*-gene on the integration plasmid (van Spanning et al., 1991). In the genome sequence of *B. smithii* ET 138 (unpublished data) no *lacZ* gene could be identified and the strain did not form blue colonies on plates containing X-gal. The *lacZ* gene from *B. coagulans* under control of the *B. coagulans pta*-promoter (Kovacs et al., 2010) was cloned into pNW33n, creating plasmid pWUR734. Introduction of pWUR734 into *B. smithii* ET 138 resulted in blue colonies in the presence of 25 mg/L X-gal. When *B. smithii* ET 138 harbouring pWUR734 was grown in the presence of 100 mg/L X-gal, the blue colonies were significantly smaller than on 25 mg/L,

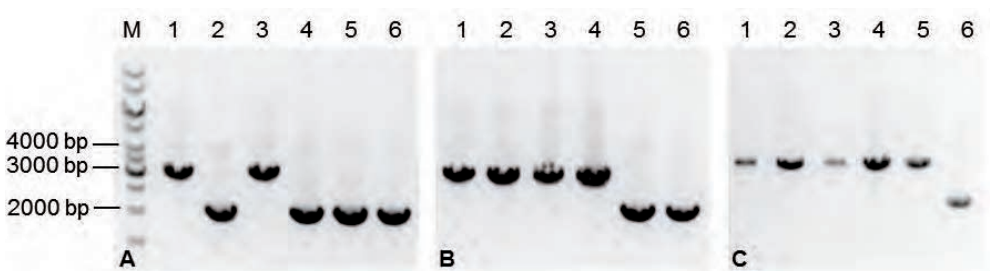

Figure 2. Gel-electrophoresis of PCR products from amplified target genes *ldhL* (A), *sigF* (B) and *pdhA* (C). Order of strains in all three parts of the picture: M: Thermo Scientific 1kb DNA ladder, 1: DSM 4216^T wild-type, 2: DSM 4216^T Δ*ldhL*, 3: ET 138 wild-type, 4: ET 138 Δ*ldhL*, 5: ET 138 Δ*ldhL* Δ*sigF*, 6: ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*. The original gel pictures without cropping are provided in Figure S1 (for Figure 2A and B) and S2 (for Figure 2C). **(A)** Amplification of the region 1000 bp up- and downstream of the *ldhL* gene using primers BG 3663 and 3669. The wild-type genotype results in a product of 3036 bp, whereas the complete deletion of the *ldhL* gene is confirmed by a shift of the product to 2094 bp. **(B)** Amplification of the region 1000 bp up- and downstream of the *sigF* gene using primers BG 3990 and 3991. The wild-type genotype results in a product of 3040 bp, whereas the complete deletion of the *sigF* gene is confirmed by a shift of the product to 2278 bp. **(C)** Amplification of the region 1000 bp up- and downstream of the *pdhA* gene using primers BG 4563 and 4564. The wild-type genotype results in a product of 3390 bp, whereas the complete deletion of the *pdhA* gene is confirmed by a shift of the product to 2280 bp.

indicating toxicity of the X-gal cleavage product at high X-gal-concentrations (Figure S3). To test the *lacZ*-counter-selection system, we chose *sigF* (*spoIIAC*) as the first knockout target gene, which is involved in the onset of *Bacillus* sporulation (Yudkin, 1987) (Figure 1). Transformation of strain ET 138 Δ*ldhL* with *sigF*-knockout vector pWUR735 containing the ~1000 bp flanking regions of *sigF* yielded only blue and no white colonies, indicating functional expression of the *B. coagulans pta::lacZ* construct. Colony PCR on 8 colonies showed a mixture of single crossovers, wild-type and double-crossover knockout genotypes for 6 colonies and no single crossover but only double-crossover knockout and wild-type genotypes for 2 colonies. The latter 2 colonies, however, failed to grow after transfer to new LB2 plates. To obtain pure knockout strains, the counter-selection was applied by growing the colonies containing the mixed genotype of single and double crossovers overnight in 10 mL LB2 at 55°C, after which dilution series were plated on LB2 supplemented with 100 mg/L X-gal. A mix of large white (1-3 mm) and small blue (≤ 1 mm) colonies was obtained for all 6 cultures (Figure S3). Colony PCR on 3 white colonies from one of the cultures showed the presence of one pure knockout, one pure wild-type and one mix of wild-type and single-crossover genotypes. The colony showing the pure knockout genotype was inoculated into liquid LB2, after which DNA was isolated and PCR analysis confirmed the knockout genotype and absence of plasmid, creating strain ET 138 Δ*ldhL* Δ*sigF* (Figure 2B).

Construction of markerless triple mutant

To evaluate whether the *lacZ*-counter-selection method can be used repeatedly to delete multiple genes and to evaluate acetate production pathways in *B. smithii* ET 138, the α-subunit of the E1 component of the pyruvate dehydrogenase complex *pdhA* was targeted for deletion in strain ET 138 Δ*ldhL* Δ*sigF* using the *lacZ*-counter-selection system. Based on genome analysis, pyruvate dehydrogenase appears to be the only route to acetyl-CoA in *B. smithii* (unpublished data). Therefore, the mutant strain was expected to be dependent on acetate to form acetyl-CoA and the medium was supplemented with acetate at all times after transformation. After transformation with *pdhA*-knockout vector pWUR737 containing the ~1000 bp flanking regions of *pdhA*, cells were plated on TVMY supplemented with acetate and chloramphenicol. PCR on 15 colonies showed 13 colonies with a mixture of wild-type genotype together with both single crossovers, 1 colony with a mixture of wild-type and downstream crossover and 1 colony with a pure downstream crossover. After one transfer of the downstream crossover colony on LB2 medium supplemented with acetate without chloramphenicol, a colony was picked showing a mixture of downstream crossover, wild-type and double crossover knockout genotypes. This colony was subjected to the counter-selection protocol by plating on 100 mg/L X-gal after overnight growth in liquid LB2, resulting in a mixture of small blue and large white colonies. From the 32 white colonies tested in PCR, 12 still showed single crossovers, 16 returned to wild-type and 4

showed a clean double crossover knockout genotype, creating triple mutant ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA* (Figure 2C).

Confirmation of sporulation deficiency

To confirm that strain ET 138 Δ*ldhL* Δ*sigF* was unable to form spores, a Schaeffer-Fulton staining was performed (Figure 3). In the wild-type and the *ldhL*-mutant (Figure 3A and B) many spores were observed as indicated by the presence of green spheres, whereas no spores were observed in the *ldhL-sigF*-double mutant (Figure 3C). Pasteurisation of cultures of the wild-type and the Δ*ldhL*-strain resulted in colony counts of >5x10^5 and 3x10^5 per mL of cells, respectively. As expected, no colonies were observed after Pasteurisation of a culture of the Δ*ldhL* Δ*sigF* double mutant, while colony counts for the control treatment were >5x10^5 per mL of cells for all 3 strains. Both assays confirm that the removal of the *sigF* gene results in a sporulation-deficient phenotype.

Growth and production characteristics of mutant strains

To evaluate growth and fermentation characteristics of the mutant strains, both strains ET 138 and DSM 4216^T and all derived mutants were grown in tubes for 24 h under micro-aerobic conditions (Table 1). Whereas both ET 138 and DSM 4216^T wild-type strains produce mainly L-lactate (± 95.5% of total products), the deletion of *ldhL* reduces L-lactate production to values around or below the detection limit, as shown by HPLC analysis combined with D- and L-lactate-specific enzyme assays (Table 1). For both strains, the main

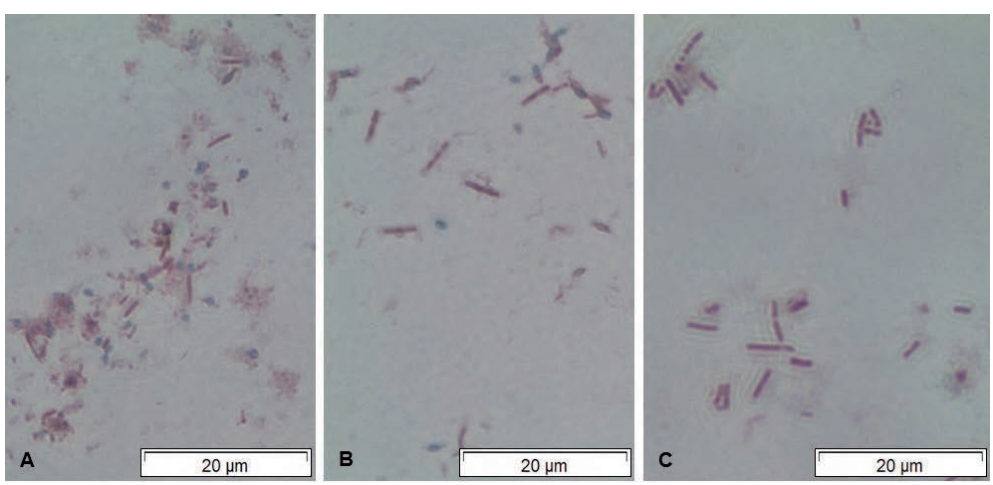

Figure 3. Schaeffer-Fulton staining on ET 138 wild-type and mutants. Cultures used were grown aerobically overnight in LB2 medium at 55°C and subsequently kept at room temperature for 24h, after which the staining was performed. Pink-stained cells indicate intact cells, whereas spores are green-blue. **A:** strain ET 138 wild-type, in which sporulation is observed. **B:** strain ET 138 Δ*ldhL*, in which sporulation is observed. **C:** strain ET 138 Δ*ldhL* Δ*sigF*, where no sporulation is observed due to removal of the *sigF* gene.

product shifts from L-lactate to acetate, with minor amounts of D-lactate, malate and succinate and increased concentrations of pyruvate compared to the wild-types. Both the OD_{600} and the final product titre of the mutants are about half that of the wild-types. Strain ET 138 Δ*ldhL* Δ*sigF* did not show significant differences compared to strain ET 138 Δ*ldhL* (Table 1). When the volume was increased from 25 to 40 mL in 50 mL tubes to further decrease the amount of oxygen present, the mutant strains showed even further reduced growth and production compared to the wild-type (data not shown). Similar results were obtained in 1 L pH controlled reactors (Table S1). Complementation of strain ET 138 Δ*ldhL* Δ*sigF* with its native *ldhL* gene and promoter expressed from pNW33n restored growth and L-lactate production to around wild-type levels (Table 1). Triple mutant ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA* was unable to grow without acetate supplementation and produced mainly D-lactate and pyruvate. In this strain, very minor amounts of L-lactate (0.39 ±0.02 mM) were observed. The final OD_{600} of this strain was on average comparable to the single and double mutant, but it showed less variation (Table 1).

Table 1. HPLC analysis of *B. smithii* ET 138 and DSM 4216^T wild-type and mutant strains.

	Products (mM) and measurement method[1]							
	Enzymatic		HPLC					
Strain	**L-lac**	**D-lac**	**Lac**	**Ace**	**Pyr**	**Mal**	**Suc**	**OD_{600}**
DSM 4216 wild-type	15.85 *±2.22*	0.73 *±0.08*	20.67 *±1.25*	9.12 *±0.49*	0.21 *±0.11*	0.06 *±0.02*	0.27 *±0.10*	0.785 *±0.209*
DSM 4216 Δ*ldhL*	0.02 *±0.03*	4.11 *±0.86*	4.95 *±3.35*	12.40 *±3.16*	1.48 *±1.14*	0.35 *±0.07*	0.57 *±0.22*	0.466 *±0.057*
ET 138 wild-type	20.55 *±4.09*	0.87 *±0.05*	26.18 *±7.66*	8.25 *±2.59*	0.78 *±0.31*	0.05 *±0.08*	0.61 *±0.16*	1.036 *±0.136*
ET 138 Δ*ldhL*	0.08 *±0.09*	1.08 *±0.32*	1.73 *±0.48*	14.33 *±5.54*	0.76 *±0.82*	0.33 *±0.15*	0.72 *±0.22*	0.581 *±0.289*
ET 138 Δ*ldhL* Δ*sigF*	0.06 *±0.05*	0.95 *±0.20*	1.58 *±0.62*	10.97 *±2.13*	0.70 *±0.41*	0.27 *±0.08*	0.60 *±0.16*	0.672 *±0.102*
ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*[2]	0.39 *±0.02*	11.17 *±0.47*	11.50 *±2.07*	-3.19[2] *±1.24*	8.26 *±0.82*	1.12 *±0.14*	1.85 *±0.37*	0.696 *±0.119*
ET 138 Δ*ldhL* Δ*sigF* + pWUR736	18.13 ±3.82	1.11 *±0.40*	18.99 *±5.19*	7.32 *±1.33*	0.24 *±0.16*	0.08 *±0.18*	0.39 *±0.17*	0.831 *±0.200*
ET 138 Δ*ldhL* Δ*sigF* + pNW33n	0.14 *±0.04*	2.57 *±1.30*	3.17 ±1.13	10.14 ±0.91	0.94 *±0.53*	0.07 *±0.12*	0.36 *±0.08*	0.535 *±0.062*
ET 138 wild-type + pNW33n	22.12 *±3.69*	0.69 *±0.10*	21.42 *±1.51*	5.63 *±1.15*	0.65 *±0.07*	0.05 ±0.12	0.34 *±0.14*	0.849 *±0.076*

Strains were grown in 25 mL TVMY supplemented with 10 g/L glucose in 50 mL Greiner tubes at 55°C for 24 h after transfer from a 10 mL-overnight culture. D- and L-lactate were distinguished via enzymatic assays, for which the lowest detection limit was 0.04 mM. The values shown are the results of three to fourteen independent experiments; numbers in italics are standard deviations.

[1] Abbreviations: L-lac: L-lactate, D-lac: D-lactate, Ace: acetate, Pyr: pyruvate, Mal: malate, Suc: succinate.

[2] These cultures were supplemented with 3 g/L ammonium acetate.

DISCUSSION

In this study, we developed an integration and counter-selection system for markerless and consecutive gene deletion in thermophilic *B. smithii*. As L-lactate is the main fermentation product of *B. smithii*, the *ldhL*-gene was selected as the first knockout target in order to use this bacterium for the production of other products. Wang *et al.* reported a laborious screening procedure when constructing a *B. coagulans* Δ*ldhL* strain and indicated that only 1 in 5000 colonies showed a knockout genotype after double crossover (Wang et al., 2011a). We observed a similar bias in *B. smithii* for the second crossover to result only in wild-type revertants. Furthermore, in single colonies we observed mixtures of either upstream or downstream single crossovers, wild-type and double crossover knockout genotype, even after several transfers targeted at purifying the colony. These mixed genotypes were not only observed for the *ldhL* deletion, but also for the *sigF* and *pdhA* deletions. For *E. coli*, such mixed genotypes have been described to occur during recombineering with linear DNA fragments, where it has been attributed to polyploidy, *i.e.* the existence of multiple chromosomes (Boyle et al., 2013). The copy number of the genome of our organism is currently unknown, but in general this number is likely to be higher when cells are grown in rich medium compared to minimal medium (Boyle et al., 2013). Transformation attempts with *B. smithii* grown on minimal medium were not successful (data not shown), but when cells were grown on minimal medium after crossovers had occurred, the purification of pure genotypes was relatively fast and easy. Switching to minimal medium to overcome mixed genotype issues might be a useful approach for other species as well.

In other thermophilic bacilli, integration events were reported after the growth temperature had been increased (Cripps et al., 2009; van Kranenburg et al., 2007). Integration of the pNW33n-derived knockout plasmids in *B. smithii* without increasing the temperature indicates either a highly efficient recombination machinery, or plasmid instability already at 55°C, although this temperature is regarded as permissive for pNW33n (Cripps et al., 2009). In both *B. smithii* strains, we observed very stable plasmid integration into the genome via single crossover recombination in colonies grown on plates without antibiotic pressure. As this integrational stability hampered purification of pure double crossover knockouts, a counter-selection method was developed to select against plasmid presence. Most frequently used counter-selection systems are based on either auxotrophy or antibiotic resistance. The *lacZ*-system is fundamentally different in enabling clean gene deletions and re-use of the marker without inducing auxotrophies, as has been demonstrated in the mesophilic α-proteobacterium *Paracoccus denitrificans* (van Spanning et al., 1991). The system can readily be used without making any prior gene deletions if the target strain does not possess β-glycosidase activity, as is the case in *B. smithii*. The system is based on the formation of toxic concentrations of inodxyl derivatives such as 5-bromo-4-chloro-3-indol, which is the

cleavage product of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). Recently, indoxyl derivatives were shown to inhibit the growth of a wide variety of species growing at different temperatures, indicating that this system might be more widely applicable (Angelov et al., 2013).

The *lacZ*-counter selection system considerably simplified the purification of single genotypes and enabled rapid and clean deletion of the *sigF* gene of ET 138 Δ*ldhL* to create a sporulation-deficient strain. The system can be used repetitively to make multiple sequential deletions in the same strain as shown by the generated triple mutant ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*. For both the *sigF* and the *pdhA* deletion, around 33% of the tested colonies were false positive after counter-selection, as they were white while still having a single crossover, which might be due to spontaneous disruption of the *lac*Z coding sequence or its promoter. Even with 33% false-positive white colonies, the screening for mutants is significantly simplified when using the *lacZ*-counter-selection system. A drawback of the system is that it does not force the double crossover in the direction of the knockout and can result in wild-type revertants. Whenever possible, cultivation conditions should be chosen such that the knockout cells will not have a large disadvantage over the wild-type cells. During this study, culturing in liquid medium resulted only in wild-type revertants, whereas knockouts were successfully obtained when the cultures were kept on plates. Supplementation of the plates with the gluconeogenic substrates acetate or lactate might have further reduced the disadvantage of knockouts over wild-types.

Removal of the *sigF* gene from *B. smithii* ET 138 resulted in a sporulation-deficient strain. Sporulation deficiency is desired in industrial settings both for practical and safety reasons. *B. smithii* was shown to have highly thermo-resistant spores (Lücking et al., 2013; Stoeckel et al., 2014) and was found to be the most dominant species together with *Geobacillus pallidus* (recently renamed to *Aeribacillus pallidus* (Coorevits et al., 2012)) as highly thermostable spores in food. Both species were found to be non-cytotoxic (Lücking et al., 2013). The *sigF*-gene has been targeted successfully to create sporulation-deficient strains in several other bacilli as well as in *Clostridia*, such as in *B. coagulans* (Kovacs et al., 2010), *B. licheniformis* (Fleming et al., 1995; Wang et al., 2005), *B. subtilis* (Yudkin, 1987) and *Clostridium acetobutylicum* (Jones et al., 2011).

Removal of the *ldhL*-gene from *B. smithii* ET 138 and DSM 4216^T eliminated L-lactate production in both strains to values around or below the detection limit, which is similar to an *ldhL*-knockout *B. coagulans* (Wang et al., 2011a), a close relative of *B. smithii*. A lactate racemase was not found in the ET 138 genome, but the methylglyoxal pathway was annotated towards both D- and L-lactate (unpublished data) and this pathway is most likely the origin of the trace amounts of lactate observed in triple mutant ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*. Products such as acetoin, 2,3-butanediol, formate and ethanol have not been detected in any of the *B. smithii* mutant strains, while these were observed in *B. coagulans*

Δ*ldhL* (Wang et al., 2011a). The absence of these products is in line with the absence of the acetolactate decarboxylase and pyruvate-formate lyase genes from the *B. smithii* genomes (unpublished data). The *B. smithii ldhL*-mutants produced mainly acetate and some D-lactate and showed reduced growth and acid production under micro-aerobic conditions, which was restored by plasmid-based complementation in strain ET 138 Δ*ldhL* Δ*sigF*. The deficiency in anaerobic capacities is likely to be caused by the redox imbalance that results from the elimination of its main NAD^+-regeneration pathway and the apparent lack of an alternative NAD^+-regeneration pathway, such as that to 2,3-butanediol. The lack of these pathways potentially makes *B. smithii* an interesting platform organism as only the *ldhL* gene needs to be removed in order to eliminate production, after which the desired product pathways can be inserted.

Acetate was the main fermentation product of the *B. smithii* strains lacking the *ldhL* gene, but the standard pathway to acetate via acetate kinase and phosphotransacetylase as well as a pyruvate-formate lyase gene are absent in the genomes of both strains ET 138 and DSM 4216^T (unpublished data). To remove acetate production, the *pdhA* gene encoding the α-subunit of the E1 component of the pyruvate dehydrogenase complex, was removed from strain ET 138 Δ*ldhL* Δ*sigF*. The resulting strain ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA* did not produce acetate and was unable to grow without acetate supplementation. This implies that pyruvate dehydrogenase is the main enzyme responsible for pyruvate to acetyl-CoA, which is in line with the absence of a *pfl* gene. Acetate utilization was previously suggested as a rescue pathway for redox balance in *Lactococcus lactis* Δ*ldhL* (Hols et al., 1999). We tested acetate supplementation of strain ET 138 Δ*ldhL* Δ*sigF* but this did not improve growth or production (data not shown).

CONCLUSIONS

In this study, we established a clean gene deletion system for *B. smithii* using *lacZ* counter-selection. We constructed *ldhL* mutants in the type strain *B. smithii* DSM 4216^T and compost isolate strain ET 138. In the latter strain, triple mutant Δ*ldhL* Δ*sigF* Δ*pdhA* was constructed, which does not produce L-lactate and acetate and is no longer capable of forming spores. Although further studies and modifications are needed to restore anaerobic growth and production capacities, the *lacZ*-counter-selection system combined with mutant-specific culturing strategies provides a tool for the construction of markerless gene deletions in thermophilic *B. smithii*. This enables the development of this species into a platform organism and provides tools for studying its metabolism, which appears to be different from its close relatives such as *B. coagulans* and other bacilli.

MATERIALS AND METHODS

BACTERIAL STRAINS AND GROWTH CONDITIONS

Strains used in this study are listed in Table 2. All *B. smithii* strains were routinely cultured at 55°C unless stated otherwise. *E. coli* DH5α was grown at 37°C. For growth experiments with strain ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*, 3 g/L ammonium acetate was added to all media at all times. For all tube and plate cultures, carbon substrates were used in a concentration of 10 g/L unless stated otherwise. Substrates and acetate were added separately as 50% autoclaved solutions after autoclavation of the medium. For plates, 5 g/L Gelrite (Roth) was added. Unless indicated otherwise, chloramphenicol was added in concentrations of 25 µg/mL for *E. coli* and 7 µg/mL for *B. smithii*.

Thermophile Vitamin Medium with Yeast extract (TVMY) contained per L: 8.37 g MOPS; 0.5 g yeast extract (Roth), 100 mL 10x concentrated Eight Salt Solution (ESS), 1 mL filter sterile 1000x concentrated vitamin solution, and 1 mL filter sterile 1000x concentrated metal mix. ESS contained per L: 2.3 g K_2HPO_4; 5.1 g NH_4Cl; 50 g NaCl; 14.7 g Na_2SO_4; 0.8 g $NaHCO_3$; 2.5 g KCl; 18.7 g $MgCl_2.6H_2O$; 4.1 g $CaCl_2.2H_2O$). 1000x concentrated metal mix contained per L: 16.0 g $MnCl_2.6H_2O$; 1.0 g $ZnSO_4$; 2.0 g H_3BO_3; 0.1 g $CuSO_4.5H_2O$; 0.1 g $Na_2MoO_4.2H_20$; 1.0 g $CoCl_2.6H_2O$; 7.0 g $FeSO_4.7H_2O$. 1000x concentrated vitamin mix contained per L: 0.1 g thiamine; 0.1 g riboflavin; 0.5 g nicotinic acid; 0.1 g panthothenic 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. The pH of TVMY was set to 6.94 at room temperature and the medium was autoclaved for 20 min at 121°C, after which vitamin solution, metal mix and substrate were added.

LB2 medium contained per L: 10 g tryptone (Oxoid), 5 g yeast extract (Roth), 100 mL ESS. The pH was set to 6.95 at room temperature and the medium was autoclaved for 20 min at 121°C. For all mutant strains, vitamins and metals as described above for TVMY were also added to LB2.

Table 2. *B. smithii* strains used in this study.

Strain	Description	Reference/origin
DSM 4216^T	Wild-type, type strain of the species	DSMZ
DSM 4216^T Δ*ldhL*	DSM 4216^T with clean *ldhL*-deletion	This study
ET 138	Wild-type, natural isolate	(Bosma et al., 2015a)
ET 138 Δ*ldhL*	ET 138 with clean *ldhL*-deletion	This study
ET 138 Δ*ldhL* Δ*sigF*	ET 138 Δ*ldhL* with clean *sigF*-deletion	This study
ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*	ET 138 Δ*ldhL* Δ*sigF* with clean *pdhA*-deletion	This study

DSMZ: Deutsche Sammlung von Microorganismen und Zellkulturen, Germany.

To evaluate product profiles and growth, cells were inoculated from glycerol stock into 10 mL TVMY supplemented with 10 g/L glucose in a 50 mL Greiner tube and grown overnight at 55°C and 150 rpm. Next morning, 250 µL cells was transferred to 25 mL of the same medium in 50 mL Greiner tubes and incubated at 55°C and 150 rpm for 24 h, after which OD_{600} was measured and fermentation products were analysed.

Plasmid construction

Plasmids and primers used in this study are shown in Table 3 and 4. Genomic DNA from *B. smithii* strains was isolated using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre). *E. coli* DH5α heat shock transformation was transformed according to standard procedures (Sambrook et al., 1989). All restriction enzymes and polymerases were obtained from Thermo Scientific. PCR products were gel-purified from a 0.8% agarose gel using the Zymoclean™ Gel DNA Recovery Kit.

For the construction of ET 138 *ldhL*-knockout vector pWUR732, the flanking regions of the *ldhL* gene were PCR-amplified from genomic DNA using primers BG3633 and BG3637 (upstream, 923 bp) and BG3638 and BG3636 (downstream, 928 bp) using Phusion polymerase. DSM 4216^T *ldhL*-knockout vector pWUR733 was created using the same primers, resulting in a 913 bp upstream region and 935 downstream. After gel-purification, an overlap extension PCR was performed in which the upstream and downstream region were fused using primers BG3633 and BG3636, making use of the complementary overhang in primers BG3637 and BG3638. The resulting PCR product was again gel-purified and subsequently cut with EcoRI and SalI using the restriction sites included in primers BG3633 and BG3636, as was plasmid pNW33n. After restriction, the fusion product and pNW33n

Table 3. Plasmids used in this study.

Plasmid	Description	Reference/origin
pNW33n	*E. coli-Bacillus* shuttle vector, cloning vector, Cm^R	BGSC
pWUR732	*ldhL*-KO vector for ET138: pNW33n+*ldhL*-flanks	This study
pWUR733	*ldhL*-KO vector for DSM 4216: pNW33n+*ldhL*-flanks	This study
pWUR734	pNW33n+*B. coagulans* P_{pta} -*lacZ*	(Kovacs et al., 2010); this study
pWUR735	*sigF*-KO vector for ET138: pWUR734+ET 138 *sigF*-flanks	This study
pWUR736	*ldhL*-restoration vector for ET138: pNW33n+*ldhL*-gene from ET 138 under its native promoter (525 bp us of *ldhL*)	This study
pWUR737	*pdhA*-KO vector for ET138: pWUR734+ET 138 *pdhA*-flanks	This study

Cm^R: chloramphenicol resistance gene; KO: knockout; BGSC: Bacillus Genetic Stock Centre, USA; us: upstream; bp: base pairs.

were ligated using T4 ligase (Thermo Scientific) for 1 h at room temperature and transformed to heat shock competent *E. coli* DH5α.

To create *lacZ*-containing plasmid pWUR734, primers BG3887 and BG3888 were used to generate the *B. coagulans* P_{pta}-*lacZ* promoter-gene fusion fragment by PCR using plasmid pPTA-LAC as template (Kovacs et al., 2010). The resulting fragment was cloned into pNW33n using SacI and BspHI and transformed to heat shock competent *E. coli* DH5α.

The ET 138 *sigF* flanking regions fragments of the knockout-plasmid pWUR735 were generated by using primers BG3971 and BG3972 (downstream, 970 bp) and BG3973 and BG3974 (upstream, 976 bp). The flanks were fused by PCR using primers BG3971 and BG3974, cloned into pWUR734 using EcoRI and SalI and transformed to heat shock competent *E. coli* DH5α. Plasmid pWUR737 containing the *pdhA* flanks in pWUR734 was constructed in a similar manner, using primers BG4522 and BG4523 to generate the *pdhA* upstream flank (1011 bp) and BG4524 and BG4525 to generate the *pdhA* downstream flank (1037 bp) by PCR. The flanks were fused by PCR using primers BG4522 and 4525 and cloned into pWUR737 using EcoRI and SalI.

For construction of ET 138 *ldhL*-complementation plasmid pWUR736, the *ldhL* gene with its native promotor (until 525 bp upstream of the gene) was amplified from the ET 138 genome using primers BG4534 and BG4535. The fragment was cloned into pNW33n using HindIII and XbaI.

Transformed *E. coli* DH5α colonies were picked and inoculated into 5 mL LB containing 25 µg/mL chloramphenicol, after which plasmids were isolated using the GeneJET Plasmid Miniprep Kit, Thermo Scientific and the integrity of the cloned fragments was confirmed by DNA sequencing (GATC, Germany). Plasmids for transforming ET 138 and DSM 4216^T were extracted from DH5α via maxiprep isolation (Genomed Jetstar 2.0).

Competent cell preparation and electroporation of *B. smithii*

B. smithii was transformed by electroporation as described previously (Bosma et al., 2015a). In brief, ET 138 cells were grown overnight at 55°C in 10 mL LB2 in a 50 mL Greiner tube and next morning diluted to an OD_{600} of 0.08 in 100 mL LB2 in a 500 mL Erlenmeyer flask or 1 L bottle. Cells were grown to an OD_{600} between 0.45 and 0.65 and made competent as described previously (Rhee et al., 2007). Electroporation was performed applying settings of 2.0 kV, 25 µF and 400 Ω in a 2 mm cuvette for ET 138 and 1.5 kV, 25 µF and 600 Ω in a 1 mm cuvette for $DSM4216^T$ (ECM 630 electroporator, GeneTronics Inc.). 2-5 µg plasmid DNA was added to the cells for electroporation and LB2 medium was used for recovery at 52°C for 3 h. After overnight growth on LB2 plates containing 7 µg/mL chloramphenicol and in the case of *lacZ*-containing plasmids also 20 µg/mL X-gal at 52°C, several colonies were streaked to a fresh plate and grown overnight at 55°C, after which colony PCR was performed to confirm the presence of the plasmid and check for integration.

Table 4. Primers used in this study.

BG nr	Sequence 5'-3'	Purpose
3464	AACTCTCCGTCGCTATTGTAACCA	Check plasmid presence
3465	TATGCGTGCAACGGAAGTGAC	Check plasmid presence
3633	GCCGTCGACCATTTGCAGTAGGTCTCGATC	*ldhL*-us-Fw
3636	GCCGAATTCTAGGTCACCAAAGACGAAATTG	*ldhL*-ds-Rv
3637	GCTCCCTTTGTATGGTCGTTTACATAATAAGAAACTCCTTTCGTCATTTC	*ldhL*-us-Rv
3638	GAAATGACGAAAGGAGTTTCTTATTATGTAAACGACCATACAAAGG-GAGC	*ldhL*-ds-Fw
3664	AGGGCTCGCCTTTGGGAAG	Int. check, in plasmid
3663	ATCGCGTGAAATGTTCTAATGG	Int. check *ldhL* on chr.-Fw
3669	AACCGATGCCGTTGATTAAAG	Int. check *ldhL* on chr.-Rv
3887	GCCGAGCTCTTGCCGGAATTCTTTCAC	*pta-lacZ*-Fw
3888	GCCTCATGACTATTTTTCAATTACCTGCAAAATTTTC	*pta-lacZ*-Rv
3971	GCCGAATTCAGCTAATCTTGTTGACGGTTTTC	*sigF*-ds-Fw
3972	GTAACTAAGGAGTCGTGCCTTAACGATTCATGTGCTTTTTTTTG	*sigF*-ds-Rv
3973	CAAAAAAAAGCACATGAATCGTTAAGGCACGACTCCTTAGTTAC	*sigF*-us-Fw
3974	GCCGTCGACCTCTGATTTAGAAGATGGAGGTTTT	*sigF*-us-Rv
3990	CGCCTATTCTTTTCGCTAAAATCGG	Int. check *sigF* on chr.-Fw
3991	ATAAGCTGCAGAGGGATATACAC	Int. check *sigF* on chr.-Rv
4534	GCCTCTAGAATTGGTCATTTGATTAGA	ET 138 *ldhL*+ prom.-Fw
4535	GCCAAGCTTTTAAGAAAGTACTTTATT	ET 138 *ldhL*+ prom.-Rv
4522	GCCGAATTCGAGGTACATAGCCCGGAATC	*pdhA*-us-Fw
4523	GTCATTTGCGGCATGGCTTACATTCGTGTCACCTCTTCCTTTC	*pdhA*-us-Rv
4524	GAAAGGAAGAGGTGACACGAATGTAAGCCATGCCGCAAATGAC	*pdhA*-ds-Fw
4525	GCCGTCGACCATCCTCATAACGGCCATCC	*pdhA*-ds-Rv
4563	GTTTCACATACCATTTAACGATTT	Int. check *pdhc* on chr.-Rv
4564	GTCAATAGGTGCAAATGGATTTTC	Int. check *pdhc* on chr.-Fw

us: upstream flanking region; ds: downstream flanking region; Fw: forward primer; Rv: reverse primer; Int.: integration; chr.: chromosome; prom.: promoter.

B. SMITHII COLONY PCR

Colony PCR on *B. smithii* was performed using the InstaGene Matrix protocol (BioRad) with several modifications: colonies were picked and resuspended in 200 µL MQ water in a 1.5 mL Eppendorf tube and centrifuged at 13200 rpm for 2 min. The supernatant was removed, 100 µL InstaGene Matrix was added to the pellet and this was incubated at 55°C for 30 min. After this, the mixtures were vortexed at high speed for 10 s and incubated at 99°C in a heat block (Eppendorf) for 8 min, vortexed again for 10 s and centrifuged at 13200 for 3 min. Subsequently, 10 µL of the resulting supernatant was used per 25 µL PCR reaction and the remainder was stored at -20°C for later use.

CONSTRUCTION OF *B. SMITHII* ET 138 AND DSM 4216^T *LDHL* MUTANTS

B. smithii ET 138 was made competent and electroporated with pWUR732. After transformation, colonies were subjected to colony PCR using primers BG3663 and BG3669 to distinguish double crossover wild-type and knockout genotypes, BG3664 and BG3669 to check chromosomal integration of the plasmid and BG3464 and BG3465 to check plasmid presence. A colony showing both upstream and downstream integration as well as wild-type genotype was picked and streaked to a fresh LB2 plate supplemented with 7 µg/mL chloramphenicol and grown overnight at 55°C, which was repeated one more time at 7 µg/mL chloramphenicol and then twice at 9 µg/mL chloramphenicol. From the last plate, a colony was picked that showed both wild-type and double-crossover knockout genome, as well as a single crossover via the upstream region. After overnight growth on LB2 supplemented with 7 µg/mL chloramphenicol and 1% glycerol at 55°C, a colony was picked that did no longer show the wild-type genotype. This colony was transferred twice on TVMY supplemented with 50 mM ammonium acetate at 65°C, resulting in a pure knockout genotype. Genomic DNA isolation was performed on liquid cultures grown overnight in TVMY containing 50 mM ammonium acetate to confirm the knockout genotype and lack of plasmid. The PCR product from primers BG3663 and BG3669 was purified (Zymo DNA Clean & Concentrator) to confirm correct deletion of the gene by sequencing.

B. smithii DSM4216^T was made competent and transformed with pWUR733, colonies were streaked to a new LB2 plate with 7 µg/mL chloramphenicol and checked for integrations as described for *B. smithii* ET 138. A colony showing wild-type genotype as well as both upstream and downstream integration was picked and streaked to a fresh LB2 plate supplemented with 9 µg/mL chloramphenicol and grown overnight at 55°C. Subsequently, it was transferred 3 more times on the same medium at 55°C and once at 65°C, after which 1 transfer was performed on LB2 containing 7 µg/mL chloramphenicol and 1% (v/v) glycerol at 55°C. Next, the colony was transferred several times on TVMY containing 50 mM ammonium acetate at 55°C and 65°C. During the whole procedure, colony PCR using

the above-mentioned primers was performed and only colonies showing single crossover (combined with wild-type genotype) were transferred. Subsequently, a colony showing double crossover knockout genotype mixed with wild-type and single crossovers was purified by transferring to TVMY containing 50 mM ammonium acetate 5 more times at 60°C and then 2 times on TVMY containing 50 mM lactate. Genomic DNA isolation was performed on liquid cultures grown overnight in TVMY containing 50 mM lactate to confirm the knockout genotype and lack of plasmid. The PCR product from primers BG3663 and BG3669 was purified (Zymo DNA Clean & Concentrator) to confirm correct deletion of the gene by sequencing.

Construction of *B. smithii* ET 138 Δ*ldhL* Δ*sigF* mutant using *lacZ* counter-selection

After transformation of *B. smithii* ET 138 Δl*dhL* with plasmid pWUR735, blue colonies were transferred to new LB2 plates supplemented with 7 µg/mL chloramphenicol twice at 55°C. Subsequently, colony PCR was performed using primers BG3990 and BG3991 to distinguish double crossover wild-type and knockout genotypes, and primers BG3990 and BG3664 to check chromosomal integration of the plasmid and BG3464 and BG3465 to check plasmid presence. Several colonies showing a mixture of single crossovers, wild-type and double-crossover knockout genotype were transferred to 10 mL LB2 in 50 mL tubes and grown overnight at 55°C, after which dilution series were plated on LB2 supplemented with 100 µg/mL X-gal. After overnight growth, white colonies were picked and transferred twice for overnight growth at 55°C on LB2 plates, after which colony PCR was performed to distinguish wild-type from double-crossover knockout genotype. Genomic DNA isolation was performed on liquid culture grown overnight in LB2 to confirm the knockout genotype and lack of plasmid. The resulting PCR product was purified (Zymo DNA Clean & Concentrator) to confirm correct deletion of the gene by sequencing and glycerol stocks were made.

Construction of *B. smithii* ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA* mutant using lacZ counter-selection

B. smithii ET 138 Δ*ldhL* Δ*sigF* was transformed with plasmid pWUR737 and recovery and plating was performed at 52°C on TVMY containing 50 mM ammonium acetate, 7 µg/mL Cm and 20 µg/mL X-gal. Next day, blue colonies were streaked to new plates containing the same medium without X-gal and grown overnight, after which colony PCR was performed using primers BG4563 and BG4564 to distinguish double crossover wild-type and knockout genotypes and BG4564 and BG3664 to check chromosomal integration of the plasmid. Several colonies showing either single crossover and/or double crossover were transferred to fresh TVMY containing 50 mM ammonium acetate plates containing antibi-

otics and grown overnight at 52°C, after which the same colony PCR was performed on the new colonies. One colony that showed only downstream crossover, whereas the others showed only upstream crossover, did not grow any further on TVMY containing 50 mM ammonium acetate and thus was transferred to LB2 containing 50 mM ammonium acetate plates without antibiotics and subjected again to colony PCR after overnight growth. One colony showing a strong double crossover knockout band was transferred to LB2 containing 50 mM ammonium acetate and grown overnight, after which it was inoculated into 10 mL liquid LB2 containing 50 mM ammonium acetate in a 50 mL tube and grown overnight at 55°C and 150 rpm. Subsequently, dilution series were plated on LB2 containing 50 mM ammonium acetate supplemented with 100 µg/mL X-gal. After overnight growth, white colonies were streaked to a new plate LB2 containing 50 mM ammonium acetate and subjected to colony PCR. Several colonies showing pure knockout genotype were inoculated into 10 mL LB2 containing 50 mM ammonium acetate and grown overnight. Genomic DNA isolation was performed on liquid cultures grown overnight in LB2 to confirm the knockout genotype and lack of plasmid, after which the resulting PCR product was purified (Zymo DNA Clean & Concentrator) to confirm correct deletion of the gene by sequencing.

SPORULATION ASSAYS

For the Schaeffer-Fulton staining (Schaeffer & Fulton, 1933), *B. smithii* strains ET 138 wild-type, ET 138 Δ*ldhL* and ET 138 Δ*ldhL* Δ*sigF* were grown aerobically overnight in LB2 medium at 55°C and subsequently kept at room temperature for 24h, after which a droplet of the cell culture was added to a microscopy slide and allowed to air-dry. The sample was heat fixed above a gas flame and covered with a piece of absorbance paper, after which the slide was flooded with 50 g/L malachite green (4-[(4-dimethylaminophenyl)phenyl-methyl]-N,N-dimethylaniline) and heated to steam twice. The absorbance paper was removed and the slide was washed with tap water, after which it was flooded with 25 g/L safranin for 30 seconds, washed with tap water, dried with paper and evaluated under the microscope (Carl Zeiss Primo Star 1000x magnification with Olympus Soft Imaging Solutions Camera and analySIS 5.0 imaging software).

For the pasteurization assay, strains ET 138 wild-type, ET 138 Δ*ldhL* and ET 138 Δ*ldhL* Δ*sigF* were grown for 24h in 10 mL LB2 in a 50 mL Greiner tube. 1 mL culture was transferred to a 1.5 mL reaction tube in duplicate, of which one tube was incubated at 60°C for 45 minutes as a control and one at 85°C for 45 minutes. After this, a 100x dilution was plated on LB2 and incubated overnight at 55°C, after which colonies were counted.

Analytical methods

Sugar and fermentation products were quantified using a high-pressure liquid chromatography (HPLC) system (Thermo) equipped with a UV1000 detector operating on 210 nm and a RI-150 40°C refraction index detector and containing a Shodex RSpak KC-811cation-exchange column. The mobile phase consisted of 5 mM H_2SO_4 and the column was operated at 0.8 mL/min and 80°C. All samples were diluted 1:1 with 10 mM DMSO in 0.04N H_2SO_4. L-lactate and D-lactate kits from Megazyme (K-LATE and K-DATE) were used to distinguish between L-lactate and D-lactate according to the manufacturer's protocol.

CONFLICT OF INTEREST

The authors declare that they have no competing interests. RvK is employed by the commercial company Corbion (Gorinchem, The Netherlands).

AUTHORS' CONTRIBUTIONS

EFB designed, executed and analysed the experiments and wrote the manuscript. AHPvdW participated in the design of the experiments, performed the experimental and analytical part of the growth experiments and revised the manuscript. LvdV constructed and analysed the *sigF* mutant, designed and performed the sporulation assays and was involved in revision of the manuscript. JvdO, WMdV and RvK participated in the design and co-ordination of the study and in revision of the manuscript. All authors approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Sjuul Hegger and Martinus J.A. Daas for technical assistance, Corbion (Gorinchem, NL) for kindly providing plasmid P_{pta}-*lacZ* (Kovacs et al., 2010) and Dr. Ron Winkler from Dutch Technology Foundation STW for his involvement in an earlier phase of this work. This work was financially supported by Corbion.

SUPPLEMENTARY DATA

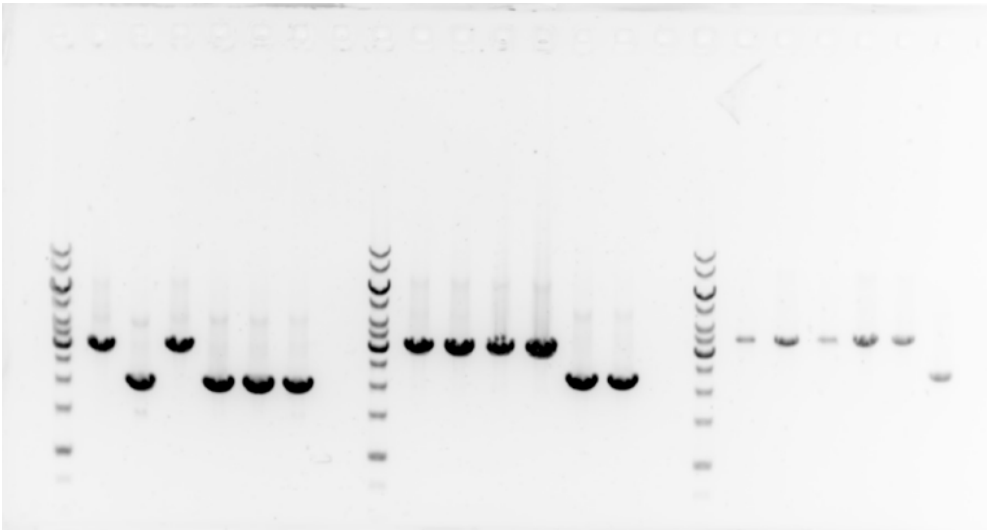

Figure S1. Original gel Figure 2 exposure 1 - for A and B. This figure shows the original gel for Figure 2 A and B. It is the same gel as in Figure C/additional file 2, but with a shorter exposure time.

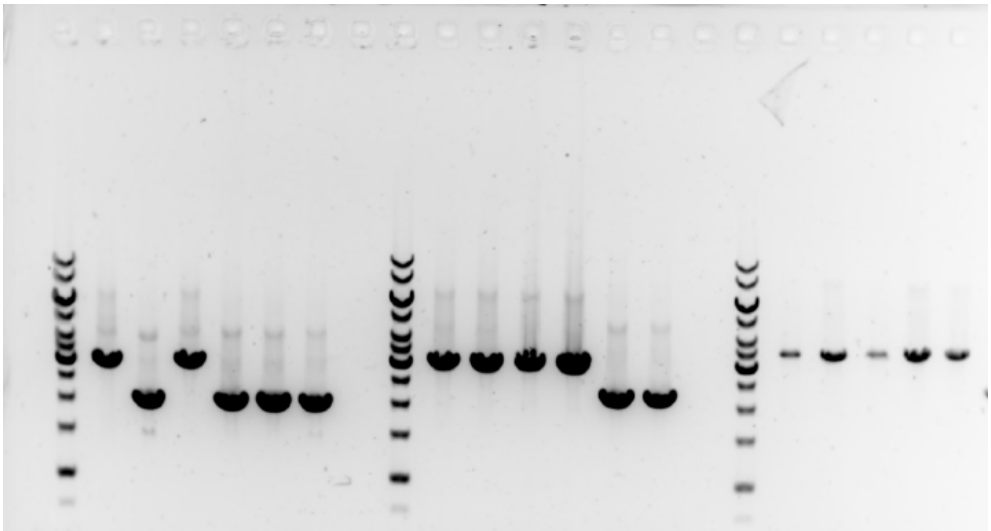

Figure S2. Original gel Figure 2 exposure 2 - for C. This file shows the original gel for Figure 2 C. It is the same gel as in Figure A-B/additional file 1, but with a longer exposure time.

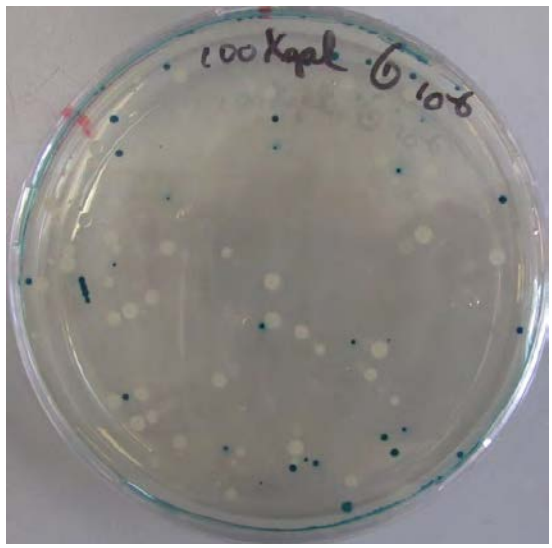

Figure S3. *B. smithii* Δ*ldhL* Δ*sigF* after counter-selection on plates containing 100 µg/mL X-gal. This figure shows the difference in colony size on 100 µg/mL X-gal after *lacZ* counter-selection. Blue colonies still contain the plasmid-encoded *lacZ* gene and are smaller than the white colonies. White colonies have lost the plasmid and thus are not inhibited by the high concentration of the X-gal cleavage product.

Table S1. Fermentation analysis of *B. smithii* ET 138 wild-type and mutant strains in 1 L pH-controlled bioreactors after ±24 h.

	Products (mM)[1]						
ET 138 strain	**Lac**	**Ace**	**Pyr**	**Mal**	**Suc**	**Total**	**OD_{600}**
wild-type	73.14	3.97	0.34	2.67	0.34	80.46	0.891
Δ*ldhL*	1.82	3.72	3.15	1.19	0.12	10.01	0.251
Δ*ldhL* Δ*sigF*	1.17	3.42	2.18	0.12	1.56	8.46	0.232
Δ*ldhL* Δ*sigF* Δ*pdhA*	1.03	-1.33	6.43	nd[2]	0.15	7.60	0.236

Cells were grown for 22.5 h (wild-type, single and double mutant) or 28 h (triple mutant) in 1 L TVMY medium containing 30 g/L glucose in a 2 L reactor at 55°C and pH 6.5, without any gas additions.

[1] Abbreviations: Lac: lactate, Ace: acetate, Pyr: pyruvate, Mal: malate, Suc: succinate.

[2] nd: not determined: in the case of strain Δ*ldhL* Δ*sigF* Δ*pdhA*, malate could not be separated precisely from pyruvate due to high pyruvate concentrations.

CHAPTER 6

METABOLIC CHARACTERIZATION OF THERMOPHILIC *BACILLUS SMITHII* AND MUTANT STRAINS

Manuscript in preparation.

Bosma, E.F., van de Weijer, A.H.P., Wahl, S.A., Diender, M., Kruis, A.J., van der Oost, J., de Vos, W.M., van Kranenburg, R.

ABSTRACT

Chemicals and fuels produced from renewable biomass by microbial fermentation are an attractive alternative for products derived from fossil resources. To reach economic feasibility, low production costs are required. We propose a thermophilic organism as production host to decrease production costs compared to a mesophilic host. However, the use of thermophiles as whole cell catalysts is hampered by (1) a lack of genetic tools and (2) insufficient knowledge about their metabolic properties. In this study, we used transcriptomics and metabolomics during controlled cultivation conditions to examine the metabolism of *Bacillus smithii* ET 138 and previously constructed mutant strains, lacking different combinations of genes coding for L-lactate dehydrogenase (*ldhL)*, sporulation-specific sigma factor *sigF,* and pyruvate dehydrogenase E1 α-component (*pdhA)*. *B. smithii* strains lacking the *ldhL* gene show a strong reduction in anaerobic growth rate and yield. We observed increased amounts of intracellular dihydroxyacetone phosphate and glyceraldehyde-3-phosphate as well as decreased amounts of intermediates of lower glycolysis and the TCA-cycle. Transcriptomics analysis furthermore showed increased expression of genes encoding most enzymes of glycolysis and gluconeogenesis, and of a cycle around pyruvate via pyruvate dehydrogenase, the glyoxylate shunt and phosphoenol pyruvate carboxykinase. Growth and production capacities were partially restored by expression of an NAD^+-generating D-lactate dehydrogenase gene from *Lactobacillus delbrueckii* subsp. *bulgaricus* in the *B. smithii ldhL* mutant. By simultaneously overexpressing native pyruvate carboxylase and malate dehydrogenase genes we attempted to redirect carbon flux towards dicarboxylic acid production and provide the cells with a means to regenerate NAD^+. However, the increase in malate and succinate production was low and pyruvate still accumulated. Enzymatic analysis indicated that this might be caused by an increased activity of malic enzyme, suggesting a similar redirection of the metabolism towards pyruvate as deduced from the transcriptomics data. The insight gained into the unusual metabolism of *B. smithii* and that of the mutants lacking *ldhL* can be used to design future engineering strategies to use *B. smithii* as a platform organism for the production of organic acids.

KEYWORDS

Bacillus smithii, thermophile, *ldhL*, *sigF*, *pdhA*, *pyc*, *mdh*, lactate, succinate, malate, pyruvate, genetic tool development, metabolomics, metabolic engineering, transcriptomics

INTRODUCTION

There is a growing demand for fuels and chemicals produced from renewable resources to replace products derived from fossil resources. To compete with their petrochemical counterparts in an economically feasible way, the costs to produce green chemicals and fuels still need to be further reduced. A prerequisite for the cost-effective production of a metabolite by microbial fermentation is a high productivity of the target compound with little or no side products, which can be achieved by rerouting of metabolic fluxes via metabolic engineering. To do so, an effective genetic engineering toolbox is required, as well as understanding of the metabolism of the host organism. As these two criteria are mainly met for model systems such as *E. coli* and *S. cerevisiae*, these organisms are among the most widely used platform organisms for biotechnological chemical and fuel production. To further reduce the costs of the process, the use of a moderately thermophilic, facultatively anaerobic host has several advantages such as reduced cooling costs, less contamination risk and reduction of the amount of required commercial sets of saccharolytic enzymes as these have their optimum around 50-60°C (Bosma et al., 2013; Ou et al., 2009; Taylor et al., 2009).

Previously, we isolated the genetically accessible moderate thermophile *B. smithii* ET 138 from compost (Bosma et al., 2015a) and established a markerless gene deletion system for it (Bosma et al., 2015b). Subsequently, we constructed three mutant strains, lacking different combinations of the genes coding for L-lactate dehydrogenase *ldhL*, sporulation-specific sigma factor *sigF* and pyruvate dehydrogenase E1 α-component *pdhA* (Bosma et al., 2015b). The removal of *ldhL* was found to lead to a drastic decrease in growth and acid production under anaerobic conditions, as apparently no alternative NAD^+-regenerating metabolic pathways are available or sufficiently expressed under these conditions in this organism. A similar modification in its close relative *B. coagulans* resulted in normal growth at neutral pH and either 2,3-butanediol production (Wang et al., 2011a) or ethanol production (Su et al., 2011b). The absence of the *alsSD* operon and the bifunctional *adhE* from the genome explains that *B. smithii* cannot produce such compounds. In thermophilic *Clostridium thermocellum*, removal of *ldhL* does not change growth of the strain, whereas simultaneous removal of *ldhL* and *pta* leads to a reduction in growth (Veen et al., 2013), although this is likely caused by disruption of the ATP-generating pathway to acetate rather than by a redox imbalance. Next to lacking alternative NAD^+-regenerating pathways besides *ldhL*, *B. smithii* does also not possess the canonical route to acetate production via *pta* and *ack*. Despite the importance of *ldh* and the highly variable behaviour of strains lacking this gene, only a limited amount of studies are available that investigate transcriptomics and metabolomics of these strains – examples of such studies being on for *Streptococcus mutans* (Mehmeti et al., 2011), *C. thermocellum* (Veen et al., 2013) and *Bacillus cereus* (Laouami et al., 2011). As

B. smithii is lacking several canonical pathways in its central carbon metabolism, it appears to differ from currently investigated strains lacking *ldhL*. In this study, we compared the wild-type strain and its isogenic knock-out mutants via transcriptomics and metabolomics. Furthermore, several genes were evaluated for the complementation of the mutant phenotype. The results indicate an accumulation of upper glycolytic intermediates, decreased amounts of intermediates of lower glycolysis and TCA-cycle intermediates, and a futile cycle around pyruvate, providing an explanation for the low yield of TCA-intermediate production after overexpression of pyruvate carboxylase and malate dehydrogenase.

RESULTS AND DISCUSSION

TRANSCRIPTOMICS ANALYSIS OF *B. SMITHII* WILD-TYPE AND MUTANTS

The main fermentation product of *B. smithii* ET 138 is L-lactate (Bosma et al., 2015a). To develop this species into a platform organism, we deleted the *ldhL* gene to eliminate L-lactate production. Subsequently, we constructed a double mutant which also lacks the *sigF* gene, creating a sporulation-deficient strain. A triple mutant Δ*ldhL*-Δ*sigF*-Δ*pdhA* was constructed to eliminate acetate production by removing the pyruvate dehydrogenase E1 α-component *pdhA* (Bosma et al., 2015b) (Table 1). Removal of the *ldhL* gene leads to a strong reduction in anaerobic growth and production capacities (Figure 1, Table 1). We

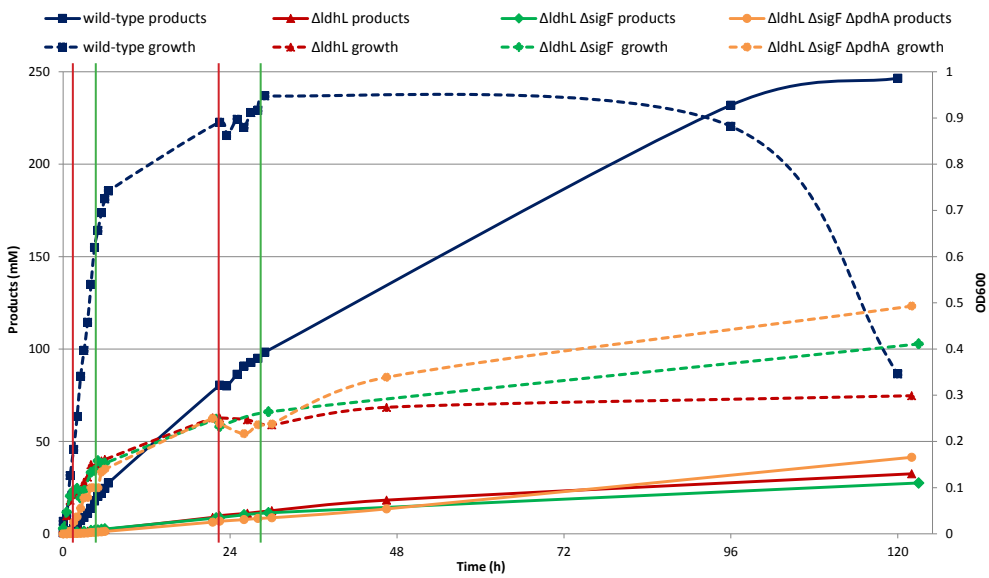

Figure 1. Fermentation graph for RNA-samples. Sampling points are indicated with a red bar for wild-type, single and double mutant and with a green bar for the triple mutant. The same sampling points were used for the metabolomics analysis.

Table 1. Fermentation analysis of *B. smithii* ET 138 wild-type and mutant cultures used for RNA-sequencing.

Time	ET 138 strain	Code[3]	Products (mM)[2]						Glc cons. (mM)	Av. prod. (mM/h)[4]	OD_{600}
			Lac	Ace	Pyr	Mal	Suc	Total			
2 h (WE-DE) or 3.5 h (TE)	wild-type	WG	4.03	0.47	0.12	0.24	0.07	4.93	3.19	2.47	0.254
	Δ*ldhL*	SG	0.08	0.49	0.16	nd	nd	0.72	nd[5]	0.36	0.093
	Δ*ldhL* Δ*sigF*	DG	0.08	0.49	0.16	0.05	0.21	0.99	nd[5]	0.50	0.098
	Δ*ldhL* Δ*sigF* Δ*pdhA*	TG	nd	nd	0.50	nd[5]	nd	0.50	nd[5]	0.14	0.079
22.5 h (WS-DS) or 28 h (TS)	wild-type	WP	73.14	3.97	0.34	2.67	0.34	80.46	41.8	3.58	0.891
	Δ*ldhL*	SP	1.82	3.72	3.15	1.19	0.12	10.01	10.1	0.44	0.251
	Δ*ldhL* Δ*sigF*	DP	1.17	3.42	2.18	0.12	1.56	8.46	6.87	0.38	0.232
	Δ*ldhL* Δ*sigF* Δ*pdhA*	TP	1.03	-1.33	6.43	nd[5]	0.15	7.60	2.88	0.27	0.236
122 h[7]	wild-type	na	231.25	17.69	0.70	5.87	0.89	256.41	123.12	2.14	0.948[8]
	Δ*ldhL*	na	6.08	14.42	9.31	2.66	0.13	32.60	27.19	0.27	0.299
	Δ*ldhL* Δ*sigF*	na	2.50	15.37	4.84	1.97	0.25	24.93	20.70	0.21	0.411
	Δ*ldhL* Δ*sigF* Δ*pdhA*	na	4.13	-4.47	27.16	nd[6]	0.70	31.99	34.67	0.27	0.493

Fermentation data of samples used for RNA-sequencing in TVMY medium containing 27-30 g/L (150-160 mM) glucose. Similar time points showing similar growth and production profiles were used for metabolomics analysis (Table S2). Data have been corrected for volume (sample taken and base added).

[2] Abbreviations: Lac: lactate, Ace: acetate, Pyr: pyruvate, Mal: malate, Suc: succinate. na: not applicable.

[3] In the two-letter code abbreviations, the first letter denotes the strain (W: wild-type; S: single knockout Δ*ldhL*; D: double knockout Δ*ldhL* Δ*sigF*; T: triple knockout Δ*ldhL* Δ*sigF* Δ*pdhA*) and the second letter denotes the growth phase or condition (G: growth phase, which is early exponential phase; P: 'production phase', which is early stationary/late-exponential phase)

[4] Av. prod.: average productivity, as calculated by dividing total products by fermentation time.

[5] nd: not determined: glucose could not be accurately measured around initial concentrations.

[6] nd: not determined: in the case of strain Δ*ldhL* Δ*sigF* Δ*pdhA*, malate could not be separated precisely from pyruvate due to high pyruvate concentrations.

[7] This time point was not used for further analysis but is shown for completeness of the dataset.

[8] This is the OD_{600} after 72 h, as for the wild-type strain the OD_{600} decreases after approximately 72 h until 0.350 at 120 h, while production continues (Bosma et al., 2015a) and therefore the maximum OD_{600} is shown.

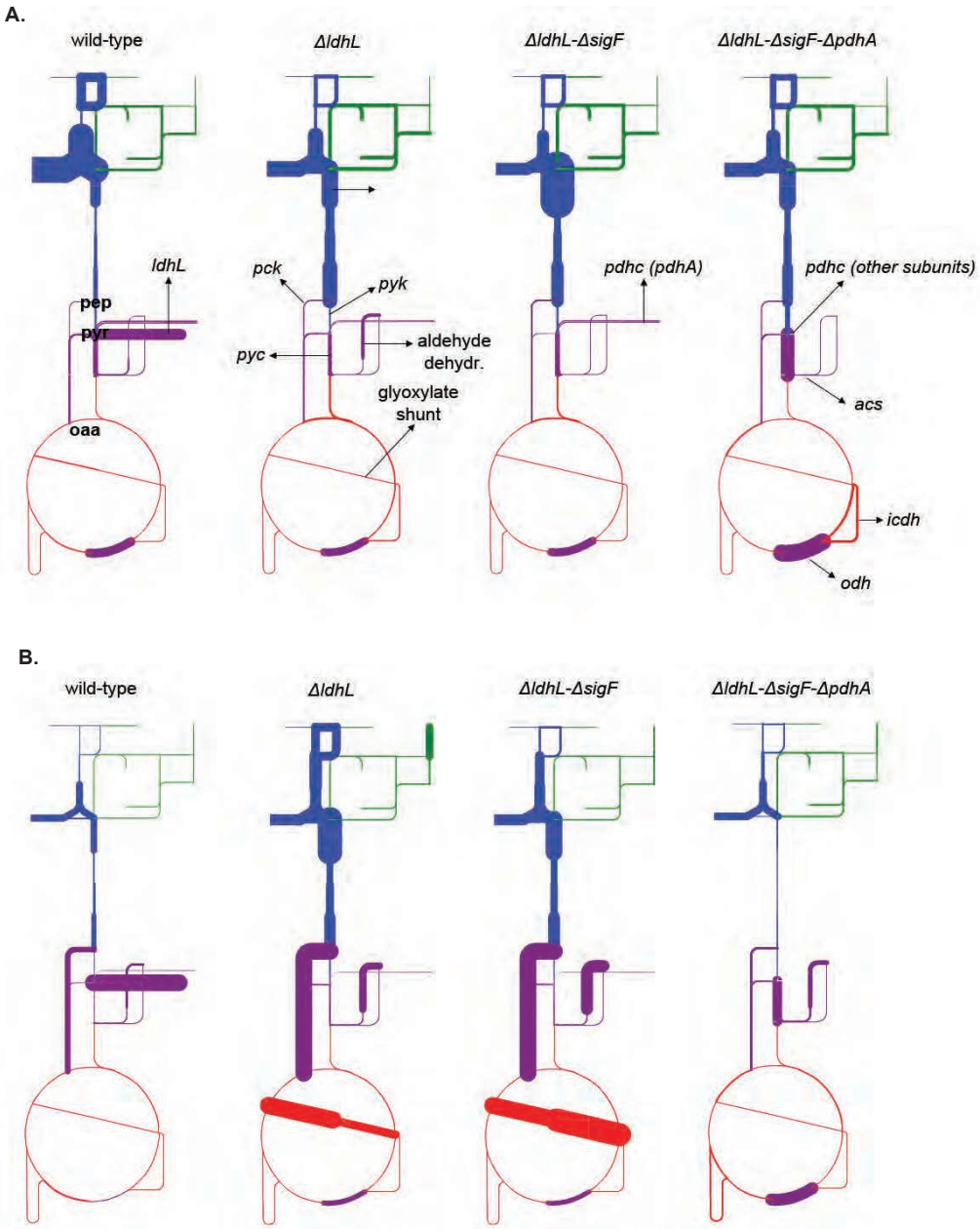

Figure 2. Mapped expression values in micro-aerobic fermentations of *B. smithii* ET 138 wild-type and mutant strains. Colours indicate sub-pathways: green is pentose phosphate pathway, blue glycolysis, purple pyruvate metabolism, red TCA-cycle. Line thickness corresponds to expression value. Pictures were generated using iPath. **A.** Growth phase (exponential phase, 2 hours). **B.** Production phase (late exponential/early stationary phase, 22.5 hours). Abbreviations: pep: phosphoenol pyruvate; pyr: pyruvate; oaa: oxaloacetate; *ldhL*: L-lactate dehydrogenase; *pck*: PEP carboxykinase; *pyk*: pyruvate kinase; *pyc*: pyruvate carboxylase; *pdhc*: pyruvate dehydrogenase complex; *acs*: acetyl-CoA synthetase; *odh*: 2-oxoglutarate dehydrogenase; *icdh*: isocitrate dehydrogenase; *6pgdh*: 6-phosphogluconate dehydrogenase; *gap*: glyceraldehyde-3-phosphate dehydrogenase; *fba*: fructose 1,6-bis-phosphate aldolase; *cs*: citrate synthase.

performed RNA-sequencing and metabolomics to gain insight into the impact of the *ldhL* deletion and its combination with the *sigF* and *pdhA* deletions on transcript and metabolite levels to identify metabolic bottlenecks in these mutant strains. RNA-sequencing was performed on the wild-type and all three mutant strains in pH-controlled reactors in 1 L working volume under micro-aerobic conditions during early exponential (2-3.5 h) and late exponential/early stationary (21-28 h) phase (Figure 1). The first sample is referred to as 'growth'. As the mutant cells did not show a clear transition from exponential to stationary phase but both mutant and wild-type cells still actively produced acids at the second sampling point, this sampling point is called 'production' (Table 1, Figure 1). The complete set of expression values can be found in Table S1. Comparison of wild-type and single (*ldhL*) mutant reveals a large set of differentially expressed genes (Table 2, Table S1). The additional deletion of the *sigF* gene in the *ldhL-sigF* double mutant does not have a large additional effect on gene expression compared to the single mutant (Table 2, Table S1). These two strains are expected to behave similarly as the only difference is sporulation gene *sigF*, which is known to not have any further metabolic effects (Fleming et al., 1995; Jones et al., 2011; Kovacs et al., 2010; Wang et al., 2005; Yudkin, 1987).

A global overview of the transcriptomics results for central carbon metabolism revealed specific patterns over time and marked differences between wild-type and mutants (Figure 2). A clear transition of activity from upper glycolytic genes to *ldhL* was observed in the wild-type strain over time. In the single and double mutant strains, the entire glycolysis is more active than in the wild-type at both time points and glycolytic genes are less de-activated over time than in the wild-type. In addition, the *pck* gene and the genes involved in the glyoxylate shunt are highly active in the production phase in the single and double mutants (Figure 2B). The glyoxylate shunt is generally regarded as a typical aerobic pathway, but it was reported to be activated under oxygen-limited conditions in an *E. coli* strain lacking *pta*, *ack* and *ldh* (Zhu et al., 2013). All of these genes are all absent in the *B. smithii* Δ*ldhL* mutants as the *pta* and *ack* genes are not present in the *B. smithii* genome (unpublished data).

Table 2. Number of differentially expressed genes in *B. smithii* ET 138 wild-type and mutant strains. For explanation of the 2-letter code abbreviations see Table 1. Differentially expressed genes are those that are at least 4-fold up-or downregulated. In the table, the columns are divided by the rows. For example, WG/SG=307 means that 307 genes are differentially higher expressed in WG than in SG and thus downregulated in SG compared to WG, and SG/WG=65 means that 65 genes are differentially higher expressed in SG than in WG and thus upregulated in SG compared to WG. Empty squares mean that these values have not been calculated.

	WG	SG	DG	TG	WP	SP	DP	TP
WG	x	65	102	80	345			
SG	307	x	22			252		
DG	407	27	x				349	
TG	244			x				249
WP	447				x	200	220	356
SP		241			341	x	32	
DP			233		346	41	x	
TP				47	237			x

The transcriptional profile of the triple mutant is more similar to the wild-type than to that of the single and double mutants, except for an increased expression of the genes encoding the non-deleted parts of the pyruvate dehydrogenase complex, as well as that coding for isocitrate dehydrogenase and 2-oxoglutarate dehydrogenase.

A more detailed overview of central carbon metabolism transcriptional and metabolite profiles in the *ldhL* and *ldhL-sigF* mutants compared to the wild-type in the production phase is shown in Figure 3. In the single and double mutant, the *gapB* gene coding for the NADPH-dependent, gluconeogenic, glyceraldehyde 3-phosphate dehydrogenase is 5-7 times upregulated compared to the wild-type in the production phase, whereas the expression of this gene is not changed in the triple mutant. The NAD^+-dependent, glycolytic, *gapA* is 4-fold upregulated in the single and double mutant but 3-fold downregulated in the triple mutant at both time points. As this is the NAD^+-dependent step in glycolysis, it is likely that this step creates a bottleneck in the *ldhL*-knockouts as these are hampered in regenerating NAD^+. The upregulation of the *gapA* gene might indicate that the cells attempt to increase this reaction by upregulating gene expression while not being able to perform it due to a lack of NAD^+. The bottleneck at the *gapA* step is supported by metabolomics data, in which glyceraldehyde-3-P accumulates in the *ldhL*-mutant (Figure 3). The lack of *gapA* upregulation in the triple mutant strain might be caused by the fact that this strain does not form extra NADH via pyruvate dehydrogenase (Figure 3) and thus faces a less severe NAD^+-shortage than the single and double mutants. The observation that this strain does not grow better than the single and double mutant, despite an apparently improved redox balance, is likely because it can only form acetyl-CoA via energy-requiring acetate uptake. Although the *ldhL* mutants are likely to suffer from redox imbalance, the redox-sensitive transcriptional regulator *rex* was not found to be differentially expressed (Table S1), whereas this was found in for example an *E. feacalis ldhL* mutant (Mehmeti et al., 2011). In all three mutant strains at both time points, a strong upregulation was observed of the operon encoding glucose-6-phosphate isomerase and the two oxidative pentose phosphate pathway (oPPP) genes encoding 6-phosphogluconate dehydrogenase and glucose-6-phosphate-1-dehydrogenase, the enzymes that form NADPH. These enzymes might form a futile cycle together with GapB, or they might shuttle carbon into other pathways such as the purine and xanthine pathways, which were also found to be upregulated. The *B. smithii* genome does not code for a transhydrogenase, but cycles created by the oPPP, gapA-B and malic enzyme might provide a means for the cells to perform transhydrogenase-like activities. A seemingly futile cycle of highly activated genes in the *ldhL* and *ldhL-sigF* mutants compared to the wild-type was observed in the genes surrounding pyruvate via upregulated expression of *pyk*/*ppdk*, *pdhc*, *icl*, *ms*, *mdh* and *pck* (Figure 3). In order to gain more insight into the metabolism, the double mutant Δ*ldhL*-Δ*sigF* and the wild-type were subjected to a metabolomics analysis, after which the transcriptomics and metabolomics data were combined (Figure 3).

Metabolome analysis of central carbon metabolism

Metabolomics analysis was performed on the wild-type and the Δ*ldhL*-Δ*sigF* strain on samples from the same reactor conditions and similar time points as the RNA-sequencing analysis (Figure 3). Initial metabolomics studies were performed by a non-targeted approach on triplicate samples of *B. smithii* ET 138 wild-type and Δ*ldhL*-Δ*sigF* at 2 h and 21 h (see Table S2 for fermentation data). Samples were analysed both from the medium and the cells. To evaluate both the influence of different variables and the statistical significance of the data, a PCA-analysis was performed. PCA analysis of the fermentation medium samples showed that the largest differences were caused by the strains, followed by the time (Figure S1). PCA-analysis of the cell samples showed a large heterogeneity of the wild-type samples at 21 h and one outlier sample for each strain at the early sampling time point (Figure S2). Due to the large heterogeneity, only few metabolites could be identified that changed significantly in the untargeted metabolomics approach. Moreover, most metabolites in central carbon metabolism could not be identified via this approach. To improve the identification of these compounds, a second round of analysis was performed using a targeted approach on a subset of the samples, excluding the outliers based on the PCA analysis (Table S2). We compared both analytic platforms by calculating Pearson's coefficients for metabolites that were measured with both platforms, which were mainly TCA-cycle intermediates and amino acids and showed a good overall correspondence between the methods (Table S3). Most glycolytic and pentose phosphate pathway (PPP) intermediates were measured only with the targeted approach and could not be compared.

Combining the transcriptomics data with the targeted metabolomics data of the production phase samples showed increased amounts of upper glycolytic intermediates and decreased amounts of lower glycolytic intermediates in the double mutant compared to the wild-type (Table 3, Figure 3). Fructose-6-P and fructose-1,6-bis-P were similar in concentration, whereas that of the triose-phosphates (glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP)) increased ~3-fold in the double mutant. The concentration of 1,3-bis-phosphoglycerate was 2-fold lower in the double mutant compared to the wild-type, whereas that of the lower glycolytic intermediates from 2-phosphoglycerate until PEP was ~5-8 decreased (Table 3, Figure 3). This supports the bottleneck at the NADH-producing GAP-dehydrogenase step that was also suggested by the transcriptomics data. The observed increased transcription of *pyk* and decrease of phosphorylated intermediates 2-PGA, 3-PGA and PEP in the *ldhL-sigF* mutant strain is in line with the decrease in the concentrations of 3-PGA and PEP in stationary cells of *L. lactis* overexpressing *pyk*, although that strain was not deficient in *ldh* (Ramos et al., 2004). The transcriptomics data also suggested futile cycling at the level of pyruvate (Figure 3). Although pyruvate inside the cell is lower in the mutant strain, extracellular levels are around 19-fold higher than in the

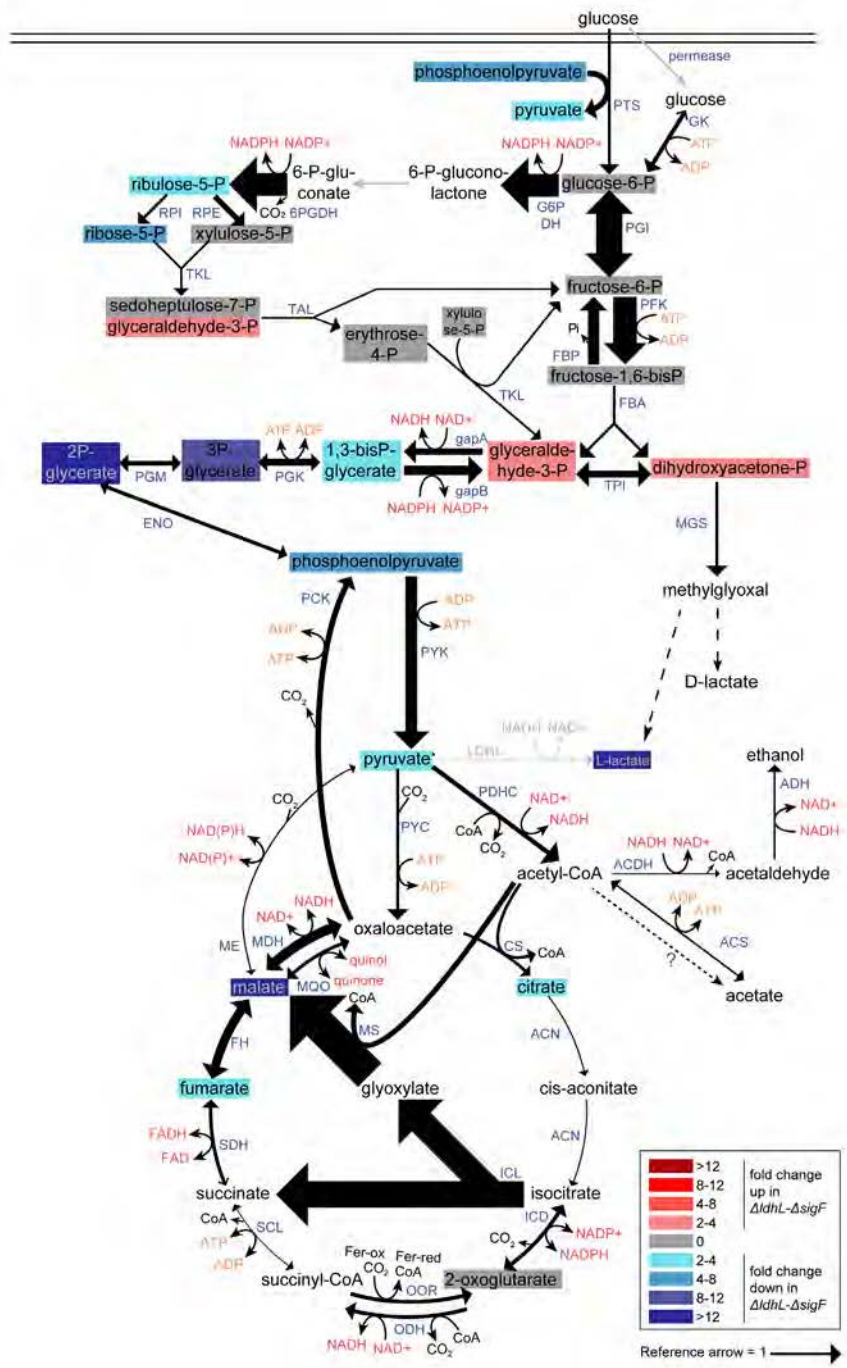

Figure 3. Detailed mapping of relative expression values and intra- and extracellular metabolite concentrations in *B. smithii* ET 138 Δ*ldhL*-Δ*sigF* compared to wild-type at 21 hours ('P' samples). Legend see next page.

Table 3. Intracellular metabolite measurements in *B. smithii* ET 138 fermentations.

Pathway	Compound	Concentrations in µmol/g cell dry weight[1]			
		wild-type t=2	wild-type t=21	ΔldhL-ΔsigF t=2	ΔldhL-ΔsigF t=21
Glucose uptake	UDP-Glucose	0.01 ±*0.00*	0.01 ±*0.00*	0.00 ±*0.00*	0.00 ±*0.00*
	Glucose	15.61 ±*5.56*	>	>	34.58 ±*4.83*
	Trehalose	<	<	<	<
Glycolysis and pyruvate metabolism	Glucose-6P	<	0.19 ±*0.04*	<	0.20 ±*0.01*
	Fructose-6P	<	0.05 ±*0.01*	<	0.07 ±*0.00*
	Fructose-1,6-bisP	0.03 ±*0.04*	0.44 ±*0.14*	0.00 ±*0.01*	0.51 ±*0.04*
	Dihydroxyacetone-P	0.15 ±*0.07*	0.39 ±*0.10*	0.12 ±*0.02*	1.02 ±*0.45*
	Glyceraldehyde-3P	0.03 ±*0.01*	0.08 ±*0.04*	0.01 ±*0.00*	0.25 ±*0.09*
	1,3-bis-phosphoglycerate	0.01 ±0.01	0.04 ±0.01	0.00 ±0.00	0.02 ±0.01
	3-phosphoglycerate	<	1.10 ±*0.24*	<	0.13 ±*0.02*
	2-phosphoglycerate	0.01	0.09 ±*0.06*	<	0.01
	Phosphoenolpyruvate	0.00 ±*0.00*	0.01 ±*0.00*	0.00 ±*0.00*	0.00 ±*0.00*
	Pyruvate	0.06 ±*0.01*	1.19 ±*0.38*	0.08 ±*0.01*	0.44 ±*0.08*
	Lactate[2]	44.44 ±*11.14*	884.05 ±*364.17*	31.82 ±*2.04*	56.05 ±*13.48*
TCA cycle	Citrate	0.13	0.51 ±*0.20*	0.11 ±*0.00*	0.15 ±*0.02*
	Fumarate	0.03	0.36 ±*0.01*	0.01 ±*0.00*	0.12 ±*0.00*
	Malate	<	0.26 ±0.07	<	<
Pentose phosphate pathway	Ribulose-5P	<	0.02	<	<
	Ribose-5P	0.03	0.09 ±*0.01*	<	0.02
	Erythorse-4P	0.01	0.01 ±*0.00*	<	0.01 ±*0.00*
	Xylose-5P	<	<	<	<
	Sepheptulose-7P	<	<	<	<

[1] Values in italics behind the ± sign are standard deviations, all values are averages of duplicate experiments and technical triplicates. Light grey values were close to or below the lowest standard. A '<' indicates a value that was below the lowest standard and could not be measured, a '>' means above highest standard. Glyceraldehyde-3P and Citrate measurements were compromised by high glucose concentrations.
[2] Lactate levels were measured in peak areas and not quantified.

Figure 3 (p. 136). Detailed mapping of relative expression values and intra- and extracellular metabolite concentrations in *B. smithii* ET 138 Δ*ldhL*-Δ*sigF* compared to wild-type at 21 hours ('P' samples). Arrow thickness indicates relative expression of ET 138 Δ*ldhL*-Δ*sigF* compared to the wild-type. Boxed metabolites represent intracellular metabolites, whereas circles next to a metabolite indicate extracellular data. No box or circle means not measured. Colours indicate fold changes up or down in ET 138 Δ*ldhL*-Δ*sig*F compared to wild-type. Abbreviations: PTS: phosphotransferase system; GK: glucokinase; PGI: glucose-6-phosphate isomerase; G6PDH: glucose-6-phosphate dehydrogenase; 6PGDH: 6-phosphogluconate dehydrogenase; RPI: phosphopentose isomerase; RPE: phosphopentose epimerase; TKL: transketolase; TAL: transaldolase; FBP: fructose bisphosphatase; PFK: phosphofructokinase; FBA: fructose bis-phosphate aldolase; TPI: triosephosphate isomerase; GAP: glyceraldehyde 3-phosphate dehydrogenase; PGK: phosphoglycerate kinase; PGM: phosphoglycerate mutase; ENO: enolase; PCK: phosphoenol pyruvate carboxykinase; PPC: phosphoenol pyruvate carboxylase; PYK: pyruvate kinase; PYC: pyruvate carboxylase; PDHC: pyruvate dehydrogenase complex; ME: malic enzyme; MDH: malate dehydrogenase; MQO: malate:quinone oxidoreductase; CS: citrate synthase; ACN: aconitase; ICL: isocitrate lyase; MS: malate synthase; ICD: isocitrate dehydrogenase; OOR: 2-oxoglutarate reductase; ODH: 2-oxoglutratae dehydrogenase; SCS: succinyl-CoA synthetase; SDH: succinate dehydrogenase; FH: fumarate hydratase; LDHL: L-lactate dehydrogenase; ACDH: acetyl-CoA dehydrogenase; ALDH: acetaldehyde dehydrogenase; ADH: alcohol dehydrogenase; ACS: acetyl-CoA synthetase; MGS: methylglyoxal synthase.

wild-type (Table 1 and 3). It is unclear whether the low intracellular concentrations of PEP, pyruvate and C_4-dicarboxylic acids are a result or a cause of the observed cycle around pyruvate, but the block in the upper glycolysis and subsequent lower concentrations of all downstream metabolites might cause an upregulation of these genes in order to replenish intermediates. The latter option together with the upregulation of both *gapA* and *gapB* suggests that improved growth may be achieved by engineering of the *gapA* cofactor specificity towards $NADP^+$ and providing the cells with both the native and the engineered variant, as has been reported previously for other systems (Bommareddy et al., 2014; Fillinger et al., 2000).

The non-targeted metabolomics approach furthermore revealed a 2-fold increase in several fatty acids in the double mutant samples in the production phase. As fatty acid biosynthesis produces $NADP^+$, its increased activity might be related to the upregulation of the oPPP which requires $NADP^+$. The energy charge based on measured AMP, ADP and ATP concentrations was comparable in wild-type and mutant cells with 0.936 and 0.958 at 2 h and 0.779 and 0.761 at 21 h, respectively. Such a slight decrease at the end of the log phase has been generally observed (Chapman et al., 1971; Kahru et al., 1982).

In the PPP, metabolite levels of ribose-5-P and ribulose-5-P were also slightly decreased in the double mutant (Table 3, Figure 3), which is in line with the strong upregulation of the genes of the oPPP and the subsequent purine/xanthine metabolism (Table S1), which might withdraw metabolites from the oPPP. Xanthine permease was also highly upregulated in the double mutant, but no higher levels of xanthine were observed by the untargeted approach in the medium in the production phase compared to the growth phase or the blanc sample at 0 h. The purine biosynthesis pathway was among the most strongly upregulated pathways in all mutant cells, whereas the pyrimidine pathway was not affected. In the wild-type, this pathway is active in the growth phase but almost completely inactivated in the production phase, while for all three mutant strains this pathway remains equally active at the second time point. This might be explained by the fact that the mutants do not fully reach stationary phase or do not show typical exponential growth, but keep on growing with very low growth rate, while the wild-type is about to enter stationary phase at 22.5 h and therefore needs to synthesize less building blocks for growth (Figure 1).

Metabolome and transcriptome analysis of amino acid metabolism

In several organisms, amino acid production has been shown to be a way of maintaining redox balance under fluctuating environmental conditions or after gene deletions (Kengen & Stams, 1994; Kim et al., 2013; Shimizu et al., 2010; Veen et al., 2013). In *B. smithii* ET 138 Δ*ldhL*-Δ*sigF*, only alanine and glycine were produced and excreted (Table 4/S4). Between 2 and 21 h, the wild-type produced a negligible amount of 6.23 µM of alanine (0.008% of total C-moles produced) and consumed instead of produced glycine. The mutant produced

592.39 µM alanine, constituting ~8% of total C-moles produced and a very small amount of glycine (21.08 µM, constituting 0.2% of total C-moles) (Table 1/S2/S4). This is different from observations in *C. thermocellum*, in which an *ldhL-pta* mutant produced amino acids to a lesser total percentage but in a greater variety. In that strain, especially amino acids were produced that require little ATP and generate NAD^+ or $NADP^+$ (Veen et al., 2013). Alanine production from pyruvate via alanine dehydrogenase consumes NADH and yields NAD^+ but requires ammonia. Alanine dehydrogenase showed no significant changes in expression in the double mutant compared to the wild-type at 21 h, but its expression was 7.7-fold increased at 2 h. Nevertheless, the flux through this NAD^+-generating pathway is apparently insufficient under the tested conditions to support normal growth of the mutant cells.

Table 4. Intracellular amino acid measurements in *B. smithii* ET 138 fermentations.

	Concentrations in µmol/g cell dry weight[1]			
Amino acid	**wild-type t=2**	**wild-type t=21**	**ΔldhL- ΔsigF t=2**	**ΔldhL- ΔsigF t=21**
valine	<	0.23 ±*0.04*	<	0.12
leucine	0.03 ±*0.01*	0.05 ±*0.01*	0.06 ±*0.00*	0.17 ±*0.08*
isoleucine	0.01 ±*0.00*	0.02 ±*0.00*	0.03 ±*0.00*	0.07 ±*0.03*
alanine	<	3.29 ±*0.78*	<	0.61 ±*0.12*
glycine	<	0.13 ±*0.03*	0.07 ±*0.00*	0.10 ±*0.00*
phenylalanine	0.02 ±*0.02*	0.45 ±*0.04*	0.03 ±*0.00*	0.06 ±*0.01*
tryptophan	<	<	<	0.03 ±*0.01*
tyrosine	0.02	0.38 ±*0.01*	0.02 ±*0.00*	0.17 ±*0.23*
proline	<	<	<	<
serine	<	0.15 ±*0.05*	<	<
threonine	<	<	<	0.13
cysteine	<	<	<	<
methionine	<	<	<	0.03 ±*0.01*
asparagine	<	<	<	<
glutamine	<	0.58 ±	<	1.18 ±*0.23*
aspartic acid	<	0.48 ±*0.17*	<	<
glutamic acid	<	<	<	1.85 ±*0.29*
lysine	0.25	3.87 ±*0.43*	0.12 ±*0.04*	0.88 ±*0.05*
histidine	<	0.06	<	0.06

[1] Values in italics behind the ± sign are standard deviations, all values are averages of duplicate experiments and technical triplicates. Orange values were close to or below the lowest standard. A '<' indicates a value that was below the lowest standard and could not be measured. Arginine was not measured.

Most other amino acids were present either in higher or in equal amounts outside the cell in the mutant compared to the wild-type (Table S4), meaning that the mutant consumes less amino acids from the medium. This is likely caused by the fact that the mutant grows less well than the wild-type and therefore needs less amino acids from the medium. Inside the cells there were also differences in amino acid contents between the strains (Table 4), although the reasons for this are unclear. Glutamate, tryptophan, threonine and methionine were not detectable in the wild-type but accumulate over time in the mutant, whereas serine and aspartate showed the opposite (Table 4). Also alanine was present in higher amounts in the wild-type in the production phase, whereas this is excreted only by the mutant. Several amino acids are completely depleted from the medium in the production phase by the wild-type (Table S4), making amino acid supplementation an option for medium optimization.

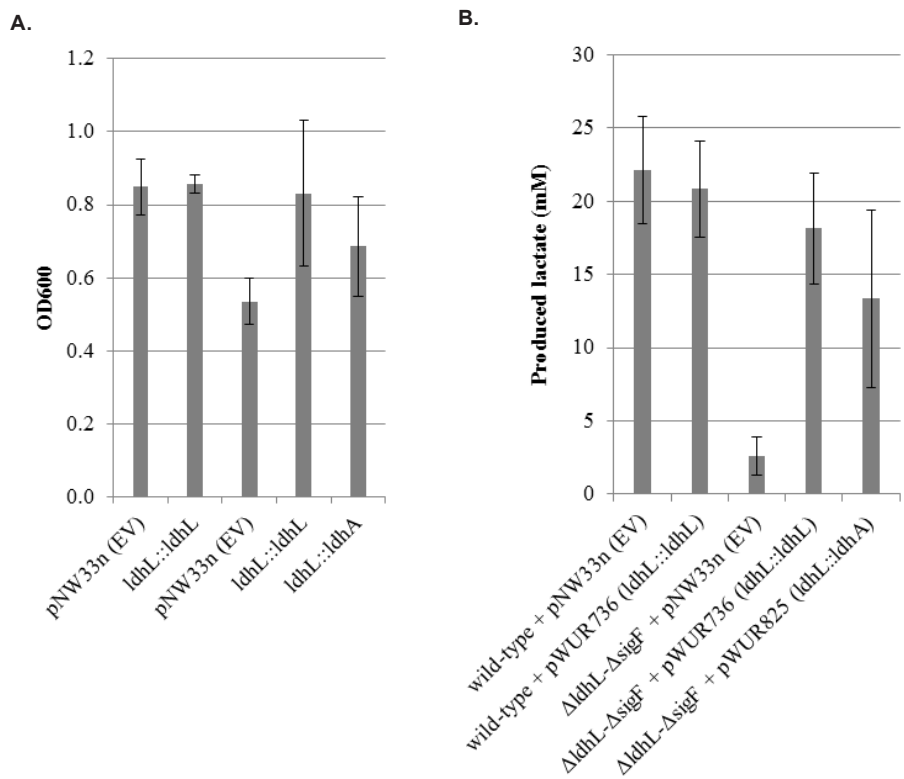

Figure 4. Plasmid-based expression of *B. smithii* ET 138 *ldhL* and *Lactobacillus delbrueckii ldhA* in *B. smithii* strains. For comparison, the previously published data of native *ldhL* expression in the mutant strain of ET 138 are also included in the graph (Bosma et al., 2015b). Samples were taken after 24 h of growth in 25 mL TVMY medium with glucose in 50 mL tubes, resembling micro-aerobic conditions. Data shown are averages of 6-14 experiments and error bars indicate standard deviations. **A.** OD_{600}. **B.** Lactate production. Lactate was determined via D- and L-lactate specific enzymatic assays; only the main isomer of lactate is shown in the graph as the other form is produced in negligible levels close to the detection limit (Bosma et al., 2015b).

Heterologous expression of D-lactate dehydrogenase

The activity of L-lactate dehydrogenase (LDH-L) is likely the main NAD^+-regenerating route under oxygen-limited conditions in *B. smithii* and based on transcriptomics and metabolomics, the observed limitation of growth and production in strains lacking *ldhL* are largely caused by the inability to regenerate NAD^+. To further test this hypothesis and simultaneously redirect metabolism to another industrially relevant product, we expressed the *Lactobacillus delbrueckii* subsp. *bulgaricus ldhA* gene under the *B. smithii ldhL* promoter. This *ldhA* gene has successfully been used for D-lactate production in *B. coagulans* (Kovacs et al., 2010; van Kranenburg et al., 2007). Heterologous expression of *ldhA* in ET 138 showed a significantly lower average production compared to plasmid-based expression of the native *ldhL* gene in the *ldhL* mutant, but with a very large standard deviation (Figure 4). The plasmid-based expression of the *ldhL* gene in the *ldhL* mutant also showed a rather large standard deviation, but less pronounced than D-lactate expression and the average difference with the wild-type is insignificant. The rather poor reproducibility might be caused by yet unexplained differences between different batches of cells, which might be related to fluctuating plasmid expression efficiencies or plasmid copy number (unpublished observations). Remarkably, the wild-type strain could not be transformed with the plasmid containing the *ldhA* gene.

Altogether, *ldhA* does not give full complementation but this might be explained by the factors mentioned above. As both the plasmids containing D- and L-lactate dehydrogenases significantly improve growth and production compared to the mutant containing the empty vector, NAD^+-regeneration might be a major cause of the growth problems in our *ldhL*-mutant strains. Alternatively, NAD^+-regeneration is not the only problem in the mutant cells and *ldhL* or L-lactate play other roles as well. In *B. subtilis*, LDH-L was shown to use both NADH and NADPH (Romero et al., 2007). Expression of the *E. coli* transhydrogenase *udhA* restored anaerobic growth capacities almost to wild-type levels in these cells, suggesting that LDH-L plays a transhydrogenase-like role in maintaining NADH/NADPH balance (Romero et al., 2007). Enzyme assays in our strain indicated that *B. smithii* LDH-L is unable to use NADPH, while its close relative *B. coagulans* showed similar enzyme activities with NADH and NADPH (data not shown). Thus, LDH-L does not play a transhydrogenase-like role in *B. smithii*. Also, the addition of L-lactate to the growth medium did not enhance growth of the mutant strain (data not shown). Complementation of the mutant strain with a gene coding for a catalytically dead LDH-L might provide further insight in a possibly regulatory role of LDH-L in this organism.

OVEREXPRESSION OF NAD$^+$-GENERATING PATHWAYS TO PRODUCE DICARBOXYLIC ACIDS

Next, we aimed to provide the mutant cells with a different NAD$^+$-regenerating route, which would simultaneously redirect metabolism towards a product of biotechnological relevance. Dicarboxylic acids such as malate and succinate can be used as green building block chemicals and have a large market potential (Bozell & Petersen, 2010; Werpy et al., 2004). To redirect metabolism in our mutant strains towards these reductive TCA-cycle products, we overexpressed the genes for either native pyruvate carboxylase *pyc*, malate dehydrogenase *mdh*, or *pyc-mdh* simultaneously in strain ET 138 Δ*ldhL*-Δ*sigF* under the *B. coagulans pta*-promoter. This promoter has been shown previously to drive *lacZ* expression in *B. smithii* (Bosma et al., 2015b) and initial tests showed it was also efficiently driving *lacZ* if this was placed behind another gene together with a Shine-Dalgarno sequence when grown on glucose or xylose (data not shown). The enzyme MDH interconverts oxaloacetate and malate, with an NAD$^+$ produced for each molecule of malate. Overexpression of this enzyme has been successfully used in *E. coli* to enhance succinate production (Liang

Table 5. Product profiles after 24 h of growth of *B. smithii* ET 138 Δ*ldhL*-Δ*sigF* strains overexpressing several native genes.

Growth condition	Plasmid	Gene on plasmid	Products (mM)[1]					
			Lac	Ace	Pyr	Mal	Suc	OD_{600}
Bottles[2]	pNW33n	none	1.00 *±0.32*	9.05 *±2.45*	0.68 *±0.94*	0.37 *±0.65*	0.24 *±0.12*	0.636 *±0.022*
	pWUR826	*pyc*	1.38 *±0.63*	9.28 *±1.69*	0.72 *±0.43*	0.12 *±0.16*	0.55* *±0.07*	0.639 *±0.109*
	pWUR827	*mdh*	0.94 *±0.08*	9.30 *±1.25*	0.92 *±0.76*	0.17 *±0.25*	0.33 *±0.17*	0.642 *±0.022*
	pWUR828	*pyc-mdh*	1.69 *±0.50*	10.84 *±0.15*	0.76 *±0.73*	0.84 *±0.56*	0.85* *±0.13*	0.614 *±0.050*
Reactors[3]	pNW33n	none	0.40	6.01	0.91	0.47	0.06	0.430
	pWUR826	*pyc*	1.31	8.19	2.55	0.46	0.04	0.509
	pWUR828	*pyc-mdh*	1.52	5.67	2.19	1.52	0.29	0.370

[1] Abbreviations: Lac: lactate; Ace: acetate; Pyr: pyruvate; Mal: malate; Suc: succinate; nd: not determined.
[2] Strains were grown in 50 mL TVMY medium with glucose in 120 mL rubber-capped bottles for 24 h. Data are the average of 3 experiments for pNW33n, pWUR826 and pWUR828 and 2 for pWUR827; numbers in italics indicate standard deviations.
[3] Strains were grown in 1 L TVMY medium with glucose in pH-controlled reactors for 24 h.
* Indicates a significant difference ($p < 0.05$) with the empty vector (pNW33n) strain.

et al., 2011; Zhu et al., 2013). As our organism lacks a fumarate reductase, and succinate dehydrogenase strongly favours conversion in the direction of fumarate (Cecchini et al., 2002), it might not be possible to produce succinate without heterologous expression of a fumarate reductase. Nevertheless, overexpression of malate dehydrogenase in our strain is expected to lead to malate production. To ensure sufficient flux towards oxaloacetate and in an attempt to relieve to observed pyruvate accumulation (Table 1), also co-expression of *mdh* with the native ET 138 *pyc* was investigated.

Initial tests were performed in 120 mL rubber-capped bottles containing 50 mL TVMY supplemented with 10 g/L glucose (Table 5). Expression of *pyc* was confirmed by SDS-PAGE analysis (data not shown) and resulted in a 2.3-fold increase of succinate production from 0.24 ±0.12 mM to 0.55 ±0.07 mM (Table 5). Expression of *mdh* was also confirmed on protein gel (data not shown) but did not alter the production profile significantly (Table 5). When *pyc* and *mdh* were co-expressed, succinate production became slightly higher compared to the *pyc*-strain and was increased 3.5-fold compared to the empty vector. Malate production did not significantly change but was slightly lower in the *pyc* and *mdh* strains and highest in the *pyc-mdh* strain. The addition of CO_2 in the form of $NaHCO_3$ in different concentrations did not significantly alter the production profile (data not shown), indicating that CO_2-supply is not limiting at this point. To further test these strains under more controlled conditions, they were grown in 1 L pH-controlled reactors for 24 h (Table 5). Similar observations were made as in the bottle test, with the highest increase in malate and succinate production being 5-fold compared to the empty vector strain in the *pyc-mdh* strain (Table 5). Furthermore, pyruvate levels were increased compared to the empty vector control. This is unexpected as *pyc-mdh* would supply the cells with a means to further metabolize pyruvate and regenerate NAD^+. Also, contrary to our observation in *B. smithii*, in an *E. coli ldh-pfl* double mutant which was unable to grow anaerobically, *mdh* overexpression resulted in improved growth and high succinate production (Liang et al., 2011). *E. coli* lacks a *pyc* gene, but heterologous expression of this gene has been shown to result in increased succinate production (Gokarn et al., 2000; Sánchez et al., 2005). In our strains overexpressing *pyc* or *pyc-mdh*, growth was only moderately improved and succinate and malate levels were only slightly increased. Transport of malate does not appear to be the major bottleneck at this point, as the wild-type is still able to excrete more malate (Table 1).

To evaluate whether the observed pyruvate accumulation and lack of dicarboxylic acid production was due to reflux from malate to pyruvate via malic enzyme, enzyme assays were performed on samples taken from the reactor for the empty vector and *pyc*-overexpressing strains after 50 h of cultivation. At that time point, the *pyc* strain had produced 7.10 mM pyruvate, 0.98 mM succinate and 0.38 mM malate, whereas this was only 2.00 mM, 0.15 mM and 0.01 mM respectively for the empty vector strain. The genome of *B. smithii* ET 138 encodes both NAD(H)- and NADP(H)-dependent malic enzymes and malic enzyme is

known to be able to work in both directions between malate and pyruvate, although in nature the most common direction is towards pyruvate (Lerondel et al., 2006; Stols & Donnelly, 1997), so both cofactors were tested in both directions (Figure 5). Malic enzyme activity in the direction to pyruvate was observed with $NADP^+$ as a co-factor and was 4-fold higher in the *pyc*-overexpressing strain than in the empty vector strain (Figure 5A). In the reverse direction, from pyruvate to malate, activity was observed with both cofactors in both strains and the differences between the two strains were less pronounced, although also in this case the *pyc*-overexpressing strain showed a slightly higher activity (Figure 5B). Altogether, these results indicate an increased activity of mainly $NADP^+$-dependent malic enzyme in the *pyc*-overexpressing strain compared to the empty vector strain. These results might explain the observed pyruvate accumulation in the *pyc*-strain (and *pyc-mdh*-strain) via reflux through $NADP^+$-dependent malic enzyme. This mechanism shows similarities to the malate shunt in *C. thermocellum*, which uses MDH and ME to produce pyruvate in the absence of PYK (Taillefer et al., 2015; Zhou et al., 2013a). Also, this suggests a cycle between malate and pyruvate via the overexpressed pyruvate carboxylase and the native malic enzyme, and might explain why the increase in growth and in malate (or succinate) production in the overexpression mutants is less than expected. These observations are in line with the observed cycle around pyruvate via *pyk/ppdk*, *pdhc*, *icl, ms, mdh* and *pck* as suggested by the transcriptomics data. If *pyc* is overexpressed, malic enzyme takes over or adds to the role of *pck* and flux is again re-directed to pyruvate.

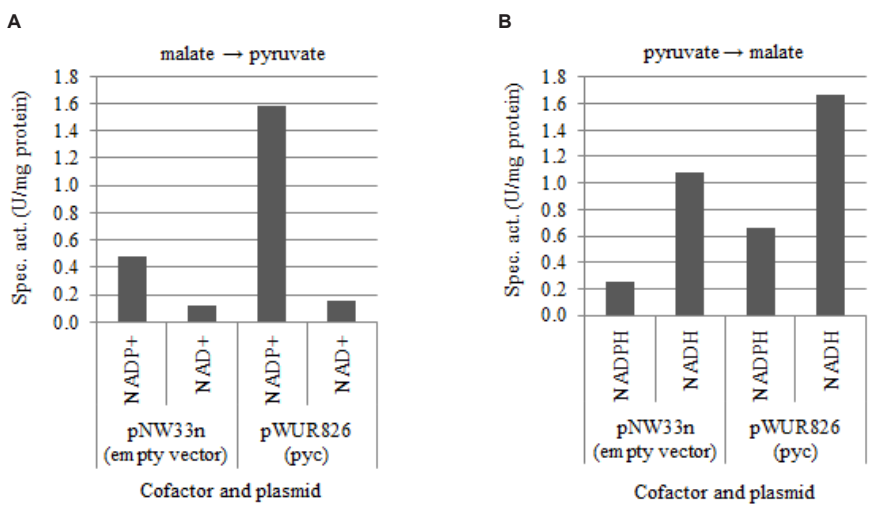

Figure 5. Malic enzyme activity assays in *B. smithii* ET 138 Δ*ldhL* Δ*sigF* containing empty vector pNW33n and overexpressing native *pyc* from plasmid pWUR826. Samples were taken after 50 h of growth in 1 L pH-controlled fermentations (Table 5), after which cell free extracts were prepared which were used in the assay. Abbreviations: Spec. act.: specific activity (in Units per mg protein). **A.** Assay using malate as initiator and oxidized co-factor. **B.** Assay using pyruvate as initiator and reduced co-factor.

Although NAD^+-regeneration appears to be the major bottleneck in the mutant strains lacking ldhL, overexpression of mdh caused a reflux to pyruvate while providing an alternative NAD+-regenerating route. To prevent redirection of flux to pyruvate and force flux into the TCA-cycle, most likely malic enzyme and PEP-carboxykinase have to be deleted or engineered to favour the reverse reaction, and a reductive TCA-branch needs to be heterologously inserted. Alternatively, the bottleneck is created already in upper glycolysis at the GAPDH level. To solve this, the co-factor specificity of GapA could be engineered to use $NADP^+$ instead of NAD^+ and both the native and engineered enzymes could be inserted to allow for more flexibility at this step.

CONCLUSIONS

In this study, we examined the metabolism of *B. smithii* after deletion or overexpression of metabolic genes by transcriptomics, metabolomics and enzyme assays. In strains lacking the *ldhL* gene, alternative NAD^+-regeneration pathways do not seem to be sufficiently available to maintain a wild-type level of anaerobic growth and production. The main bottleneck appears to be *gapA*, the NADH-producing step in glycolysis. Metabolites upstream of this enzyme accumulate, whereas metabolites downstream of this enzyme were found to be less present in the mutant compared to the wild-type. Furthermore, transcriptomics revealed futile cycling around pyruvate. If pyruvate was forced into the TCA-cycle via overexpression of *pyc*, an increase in malic enzyme activity was observed, recreating the reflux towards pyruvate. Further research for example aiming at co-factor engineering of *gapA* should show if the cycling around pyruvate is related to the bottleneck in upper glycolysis, or that *pck* and malic enzyme should also be targets of future metabolic engineering.

Table 6. Strains used in this study.

Strain	Description	Reference/origin
E. coli DH5α	Cloning host	Lab stock
E. coli TG90	Cloning host	O.P. Kuipers, RUG[1] (Lopilato et al., 1986)
B. smithii strains:		
ET 138	Wild-type, natural isolate	(Bosma et al., 2015a)
ET 138 Δ*ldhL*	ET 138 with clean *ldhL*-deletion	(Bosma et al., 2015b)
ET 138 Δ*ldhL* Δ*sigF*	ET 138 Δ*ldhL* with clean *sigF*-deletion	(Bosma et al., 2015b)
ET 138 Δ*ldhL* Δ*sigF* Δ*pdhA*	ET 138 Δ*ldhL* Δ*sigF* with clean *pdhA*-deletion	(Bosma et al., 2015b)

[1] RUG: Rijksuniversiteit Groningen/University of Groningen.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Strains used in this study are listed in Table 6. All *B. smithii* strains were routinely cultured at 55°C, unless stated otherwise. *E. coli* DH5α and TG90 were grown at 37°C. For strain ET 138 Δ*ldhL*-Δ*sigF*-Δ*pdhA*, 3 g/L ammonium acetate was added to all media at all times. For all tube and plate cultures, substrates were used in a concentration of 10 g/L unless stated otherwise. Substrates and ammonium acetate were added separately as 50% autoclaved solutions after autoclavation of the medium. For plates, 5 g/L gelrite (Roth) was added for *B. smithii* strains and 10 g/L agar (Difco) for *E. coli* strains. Unless indicated otherwise, chloramphenicol was added in concentrations of 25 µg/mL for *E. coli* DH5α, 15 µg/mL for *E. coli* TG90, and 7 µg/mL for *B. smithii* strains.

Thermophile Vitamin Medium with Yeast extract (TVMY) and LB2 medium were used as described previously (Bosma et al., 2015a).

For enzyme assays, cells were either sampled from the reactors after 50 h, or grown overnight from glycerol stocks in 10 ml TVMY containing 10 g/L glucose at 55°C in a 50 mL Greiner tube at 150 rpm. After overnight growth, 100 µL was transferred to 10 mL fresh medium and cells were grown for 4-6 h, after which cell extracts were prepared.

Fermentations

Fermentations were performed as described previously with minor modifications (Bosma et al., 2015a). Pre-cultures were grown overnight from glycerol stocks in 10 ml TVMY containing 10 g/L glucose at 55°C in a 50 mL Greiner tube at 150 rpm. Next morning, 1 mL of the culture was transferred to 50 mL TVMY containing 10 g/L glucose in a 250 mL Erlenmeyer flask and incubated at 55°C and 150 rpm. When the cells reached an OD_{600} of 0.3-0.4, 20 mL was used to inoculate the reactor containing 1 L TVMY medium without MOPS and supplemented with 27 g/L glucose.

Fermentations were all in batch mode and performed in 2 L glass reactors (Applikon) under control of an ADI 1025 Bio-console and an ADI 1010 Bio-controller (Applikon) containing 1 L working volume. Temperature was controlled at 55°C, stirring speed was controlled at 150 rpm, pH was maintained at 6.5 by addition of 3 M KOH and growth was monitored by absorbance at 600 nm (OD_{600}). Sugar and fermentation products were measured by HPLC. For plasmid-containing strains, chloramphenicol was added twice per day to maintain a concentration of 7 mg/L, assuming a half-life time of 100 h for chloramphenicol at 55°C (Taylor et al., 2011).

Cell dry weight (CDW) of the fermentation broth was determined of 25 mL sample, which was centrifuged at 4816 x *g* at 4°C for 15 min, after which the supernatant was discarded.

The pellet was resuspended in 10 mL dH2O and pelleted again. A second resuspension was performed in 1 mL dH2O and the cells were transferred to an aluminium dish of which the tare weight was determined, after which the cell suspension was allowed to dry at 120°C overnight and weight was determined.

Sampling procedure for metabolomics

The sampling procedure was performed according to the protocol as described by Metabolomic Discoveries (Meißner et al., 2014), with some modifications as follows. From the fermentation broth, a sample of 50-100 mL was taken (sampling time ~2 s), of which 15 or 50 mL was immediately used for quenching, while the remainder was kept on ice for determination of cell dry weight (CDW), microscopic analysis and cell count. The 15/50 mL sample was transferred to a pre-cooled 50 mL Falcon tube, which was immediately dipped into a -40°C ethanol bath for 50 s with intervals of 10 s and inverting the tubes in between to cool the samples to ~4°C. The cooled cell suspension was poured onto a filter (Pall 0.45 µm PVDF) in a 47 mm filter system (Millipore Sterifil aseptic system) attached to a vacuum pump. The filter had been equilibrated just prior to sample by operating it twice with 5 mL ice-cold 0.9% NaCl. After filtration of the sample, the filter was washed twice with ice-cold 0.9% NaCl, after which the filter was added to a Falcon tube, immediately dipped into liquid nitrogen, kept on dry ice and subsequently stored at -80°C until further processing. Extraction buffer (80% methanol + ^{13}C Sorbitol as internal standard) was pipetted ~10 times over the filters in dry-ice to obtain a concentration of 10 mg/mL CDW. After extraction, samples were shipped on dry ice for analysis. Prior to the first analysis, the extraction buffer was washed again several times over the filter, the cell suspension was treated in a ball mill and cell debris was removed.

RNA isolation and library preparation

RNA isolation was performed by the phenol extraction based on (van Hijum et al., 2005). For each RNA isolation, 25-100 mL culture was centrifuged at 4°C and 4816 x *g* for 15 min and immediately used for RNA isolation. After removal of the medium, cells were resuspended in 0.5 mL ice cold TE buffer (pH 8.0) and kept on ice. All samples were divided into two 2 mL screw-capped tubes containing 0.5 g zirconium beads, 30 µL 10% SDS, 30 µL 3 M sodium acetate (pH 5.2), and 500 µL Roti-Phenol (pH 4.5-5.0, Roth). Cells were disrupted using a Fastprep apparatus (MP Biomedicals) at speed 6 for 40 seconds and centrifuged at 4°C and 10,000 rpm for 5 min. 400 µL of the water phase from each tube was transferred to a new tube, to which 400 µL chloroform-isoamyl alcohol (Roth) was added, after which samples were centrifuged at 4°C and 18400 x *g* for 3 min. 300 µL of the aqueous phase was transferred to a new tube and mixed with 300 µL of the Lysis buffer

from the High Pure RNA Isolation Kit (Roche). Subsequently, the rest of the procedure from this kit was performed according to the manufacturer's protocol, except for the DNAse incubation step, which was performed for 45 min. Integrity and concentration of the isolated RNA was checked on Nanodrop-1000 and on the Bioanalyzer (BioRad experion RNA Std-Sens) according to the manufacturer's protocol. All samples had a RIN value of at least 8.0, had a concentration above 200 ng/µL and contained minimally a total of 15 µg.

Further processing of the samples and RNA-sequencing were performed by Base-Clear (NL). The rRNA fraction was depleted from the total RNA using the Ribo Zero Kit (Epicentre), after which the TruSeq RNA-Seq kit (Illumina) was used for mRNA library preparation. Subsequently, the mRNA library was fragmented and converted to double-stranded cDNA with fragment lengths of 100-200 bp. DNA adapters were ligated to both ends of the fragments and subjected to PCR amplification, after which the resultant library was checked on the Bioanalyzer. RNA-sequencing was performed using the Illumina HiSeq 2500 platform using 50 cycles single-read protocol.

Plasmid construction, transformations and colony PCR

Plasmids and primers used in this study are shown in Table 7 and 8. Genomic DNA from *B. smithii* strains was isolated using the MasterPure™ Gram Positive DNA Purification Kit (Epicentre). Transformation and colony PCR of *B. smithii* strains was performed as described previously (Bosma et al., 2015a). Heat shock transformation of *E. coli* strains was performed according to standard procedures (Sambrook et al., 1989). Transformed *E. coli* colonies were picked and inoculated into 5 mL LB containing chloramphenicol, after which plasmid was isolated (using the GeneJET Plasmid Miniprep Kit, Thermo Scientific) and subsequently subjected to PCR analysis, after which correct plasmids were sent for sequencing (GATC, Germany) to confirm the sequence of the inserted regions. Plasmids for transforming *B. smithii* were extracted from *E. coli* via maxiprep isolation (Genomed Jetstar 2.0). All restriction enzymes and polymerases were obtained from Thermo Scientific. Purification of PCR products was performed after running on a 0.8% agarose gel using the Zymoclean™ Gel DNA Recovery Kit. For all PCR reactions, Phusion polymerase with HF-buffer was used.

For the construction of *P::ldhL-ldhA* vector pWUR825, primers BG 5478 and 5479 were used to amplify the *ldhL* promoter from genomic DNA of *B. smithii* ET 138 and primers BG 5476 and 5683 for the *ldhA* gene from plasmid pJS26 (van Kranenburg et al., 2007). After gel purification, an overlap extension PCR was performed in which the promoter and gene were fused using primers BG 5478 and 5683, making use of the complementary overhang in primers BG 5479 and 5476. The resulting PCR product was again gel-purified and subsequently cut with XbaI and BamHI using the restriction sites included in primers BG 5479 and 5476, as was plasmid pNW33n. After restriction, the fusion product and pNW33n were

Table 7. Plasmids used in this study.

Plasmid	Description	E. coli host[1]	Reference/origin
pNW33n	*E. coli-Bacillus* shuttle vector, cloning vector, Cm^R		BGSC
pWUR736	*ldhL*-restoration vector for ET138: pNW33n with *ldhL* gene from ET 138 under its native promoter (525 bp us of *ldhL*)	DH5α	(Bosma et al., 2015b)
pWUR825	pNW33n with *L. delbrueckii ldhA* gene under *B. smithii* ET 138 *ldhL*-promoter (525 bp us of *ldhL*)	TG90	This study; (Kovacs et al., 2010; van Kranenburg et al., 2007)
pWUR826	pNW33n with *B. smithii* ET 138 *pyc* gene under *B. coagulans* DSM 1 *pta*-promoter	DH5α	This study; (Kovacs et al., 2010)
pWUR827	pNW33n with *B. smithii* ET 138 *mdh* gene under *B. coagulans* DSM 1 *pta*-promoter	DH5α	This study; (Kovacs et al., 2010)
pWUR828	pNW33n with *B. smithii* ET 138 *pyc* and *mdh* genes under *B. coagulans* DSM 1 *pta*-promoter	TG90	This study; (Kovacs et al., 2010)

[1] The *E. coli* strain that was used for the construction of the plasmid.
Abbreviations: Cm^R: chloramphenicol resistance gene (chloramphenicol acetyltransferase); BGSC: Bacillus Genetic Stock Centre, USA; us: upstream; bp: base pairs.

Table 8. Primers used in this study.

BG nr	Sequence 5'-3'	Purpose
4122	GAGCTCCGTTGCCGGAATTCTTTCAC	*P_pta*-Fw, SacI site
4268	GCCTCTAGAAAGTTTATACTAAAGAAGCCATGAC	*mdh*-Rv, XbaI site
4269	GATTTTTTTACGCTTCAACGCCAACACGCATTCCTCCAAGTCCAT	*P_pta*-Rv, overhang with 4270
4270	ATGGACTTGGAGGAATGCGTGTTGGCGTTGAAGCGTAAAAAAATC	*mdh*-Fw, overhang with BG 4269
4441	GCCTCTAGATTGCCGGAATTCTTTCACAG	*P_pta*-Fw, XbaI site
4446	GACTTGGAGGAATGCGTGATGAAAAAGATCGAGAAAGTG	*pyc*-Fw, overhang with BG 4448
4447	GCCGTCGACTTATTTTTGCAGCTCAATCAGTA	*pyc*-Rv, SalI site
4448	CACTTTCTCGATCTTTTTCATCACGCATTCCTCCAAGTC	*P_pta*-Rv, overhang with BG 4446
4449	GAAGTTCCTCCTTATTATGGATTTATATTTATTTTTGCAGCTCAAT-CAGTA	*pyc*-Rv, overhang with BG 4450
4450	TACTGATTGAGCTGCAAAAATAAATATAAATCCATAATAAGGAG-GAACTTC	*mdh*-Fw, overhang with BG 4449
4451	GCCGTCGACTTATACTAAAGAAGCCATGACG	*mdh*-Rv, SalI site
5476	AATGACGAAAGGAGTTTCTTATTATGACTAAAATTTTTGCTTAC	*ldhA*-Fw, overhang with BG 5479
5478	GCCTCTAGAATTGGTCATTTGATTAGATCCCCTTA	*P_ldhL*-Fw, XbaI site
5479	GTAAGCAAAAATTTTAGTCATAATAAGAAACTCCTTTCGTCATT	*P_ldhL*-Rv, overhang with BG 5476
5683	GCCGGATCCTTAGCCAACCTTAACTGGAG	*ldhA*-Rv, BamHI site

Fw: forward primer; Rv: reverse primer; *P_*: promoter.

ligated using T4 ligase (Thermo Scientific) for 1 h at room temperature and transformed to heat shock competent *E. coli* TG90.

A similar procedure was followed for construction of the other plasmids. For *pyc*-overexpression plasmid pWUR826, primers BG 4441 and 4448 were used to amplify the *pta* promoter from plasmid pPTA-LAC (Kovacs et al., 2010) and primers BG 4446 and 4447 for the *pyc* gene from genomic DNA of *B. smithii* ET 138. The fragments were fused using BG 4441 and 4447 and cloned into pNW33n using XbaI and SalI. To construct the *pyc-mdh* double overexpression plasmid pWUR828, the same procedure as for pWUR826 was followed, but primer BG 4447 was replaced by BG 4449, which contains a Shine-Dalgarno sequence and has overhang with the forward primer for the *mdh* gene BG 4450. BG 4450 and 4451 were used to amplify the *mdh* gene from genomic DNA of *B. smithii* ET 138. Subsequently, the *P-pta-pyc* and *mdh* fragments were fused using primers BG 4441 and 4451, after which the fragment was cloned into pNW33n using XbaI and SalI.

For *mdh*-overexpression plasmid pWUR827, primers BG 4122 and 4269 were used to amplify the *pta* promoter from plasmid pPTA-LAC (Kovacs et al., 2010) and primers BG 4270 and 4268 for the *mdh* gene from genomic DNA of *B. smithii* ET 138. The fragments were fused using BG 4122 and 4268 and cloned into pNW33n using XbaI and SalI.

Cell extract preparation and enzyme assays

For cell free extract (CFE) preparation, 10 mL culture was centrifuged at 4816 x *g* at 4°C for 10 min. The cell pellet was resuspended in 1 mL ice-cold buffer (25 mM Tris-HCl pH 7.0) and transferred to a 2 mL Eppendorf tube, after which cells were disrupted via sonication for 3x 30 s on 20% power, with 30 s breaks in between sonication rounds using a Sonifier B-12 Power Supply (Branson Sonic Power, 3 mm pin). Cells were kept on ice at all times. Subsequently, the cell suspension was centrifuged for 10 min at 16,000 x *g* at 4°C, after which the supernatant was transferred to a new Eppendorf tube and used as CFE for enzyme assays and protein determination. Protein determination was performed using Roti Nanoquant (Roth) according to the manufacturer's protocol.

Unless stated otherwise, enzyme assays were performed in a HITACHI U-2001 spectrophotometer in plastic cuvettes at 340 nm at 55°C, water and buffer were pre-warmed to 55°C and total volume was 1 mL. Negative controls were measured without adding the initiator to measure the background of the CFE and this value was subtracted from the value with initiator.

Malic enzyme assays were performed as described previously in quartz cuvettes (Sridhar et al., 2000). The complete reaction mixture consisted of: 0.1 M Tris-HCl (pH 7), 0.5 mM NAD(P)H or $NAD(P)^+$, 5 mM $MnCl_2$, 20 mM NH_4Cl, 50µl cell free extract, and 2 mM malate or pyruvate was used as initiator. The preferred direction was determined by using malate and oxidized co-factor or pyruvate and reduced co-factor. Co-factor preference of the en-

zyme was determined by either using NAD^+ or $NADP^+$. Additional control was performed by determining the composition of the assay mixture by HPLC before and after adding the initiator (pyruvate or malate).

Lactate dehydrogenase assays were performed as described previously (Goel et al., 2012) and contained 0.1 M Tris-HCl (pH 7), 2 mM $MgSO_4$, 3 mM fructose-1,6-bis-phosphate, 0.3 mM NADH or NADPH, while 6 mM pyruvate was used as initiator.

Analytical methods

Sugar and fermentation products were quantified as described previously (Bosma et al., 2015a).

Non-targeted metabolomics analysis was performed by Metabolomic Discoveries (DE). Derivatisation and analyses of metabolites by a GC-MS 7890A mass spectrometer (Agilent, Santa Clara, USA) were carried out as described (Lisec et al., 2006). Metabolites were identified in comparison to the Metabolomic Discoveries' database of standards. The LC separation was performed using hydrophilic interaction chromatography with a ZIC-HILIC 3.5 µm, 200 A column (Merck Sequant, Umeå Sweden), operated by an Agilent 1290 UPLC system (Agilent, Santa Clara, USA). The LC mobile phase was a linear gradient from 90% to 70% acetonitrile over 15 min, followed by linear gradient from 70% to 10% acetonitrile over 1 min, 3 min wash with 10% and 3 min re-equilibration with 90% acetonitrile. The flow rate was 400 µl/min, injection volume 1 µl. The mass spectrometry was performed using a 6540 QTOF/MS Detector (Agilent, Santa Clara, USA). The measured metabolite concentration was normalized to the internal standard. Metabolites that were present in less than 20% of all samples independent of the replicate group were filtered out, as were metabolites that were present in less than 33% within a replicate group. Significant concentration changes of metabolites in different samples were analyzed using appropriate statistical test procedures. All statistical analyses were performed using R-project (Team, 2012). A p-value of $p<0.05$ was considered as significant. ATP, ADP and AMP (Sigma-Aldrich) were measured by a targeted approach using in-matrix calibrations with authentic standards.

Targeted metabolomics analysis was performed at the Technical University Delft (NL) on a subset of the samples also used at Metabolomic Discoveries (Table S2). Metabolite concentrations were determined by isotope dilution mass spectrometry (Wu et al., 2005), *i.e.* 20 µL of ^{13}C yeast cell extract was added to 50 µL of sample prior to MS analysis as internal standard. The amino acids Ala, Gly, Val, Leu, Ile, Pro, Ser, Thr, Meth, Asp, Phe, Cys, Glu, Lys, Asn, Gln, Tyr, His, Trp were measured by GC-MS (GC 7890A, MS 5975C Agilent Technologies, USA) with PTV injection system (Gerstel GmbH, Germany). The freeze-dried sample was derivatized using 80 µL MTBSTFA (N-methyl-N-t-butyldimethyl-silyltrifluoroacetamide, purchased from Sigma-Aldrich) reacting for 1 h at 50°C in a heating block.

The central carbon metabolism intermediates Pyr, Fum, Mal, aKG, DHAP, G6P and

trehalose were also determined using GC-MS, but the derivatization was performed using 50 µL MOX (O-methoxyamine hydrochloride at 20 g/L in pyridine) and 80 µL of MSTFA (N-methyl-N-trimethylsilyltrifuoroacteamide, Thermo Scientific). The metabolites FBP, PEP, Succ, T6P and G1P were determined using LC-MS with settings described in (van Dam et al., 2002; Wahl et al., 2014). The concentrations were determined based on the $^{12}C/^{13}C$ ratio and a standard calibration line using 100 µL of sample and 20 µL of ^{13}C yeast cell extract (Wahl et al., 2014).

Transcriptomics bioinformatics analysis was performed as follows. The FASTQ sequence reads were generated using the Illumina Casava pipeline version 1.8.3. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control signal were removed using a BaseClear-in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.10.0. The quality of the FASTQ sequences was enhanced by trimming off low-quality bases using the "Trim sequences" option of the CLC Genomics Workbench version 6.0.4. The quality-filtered sequence reads were used for further analysis with the CLC Genomics Workbench. First an alignment against the *B. smithii* ET 138 genomes and calculation of the expression values was been performed using the "RNA-Seq" option. Subsequent comparison of expression values and statistical analysis was performed with the "Expression analysis" option. The used expression value was the Reads per Kilobase of exon model per Million mapped reads (RPKM) (Mortazavi et al., 2008).

AUTHORS' CONTRIBUTIONS

EFB, AHPvdW, MD and AJK were involved in the design and execution of the experiments and in the data analysis; EFB wrote the manuscript; SAW was involved in the experimental design and data analysis of the metabolomics study; JvdO, WMdV and RvK were involved in the design and coordination of the study and in revision of the manuscript. All authors have read and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare to have no conflict of interest. RvK is employed by the company Corbion (Gorinchem, The Netherlands).

ACKNOWLEDGEMENTS

We thank Oscar Kuipers and Siger Holsappel (University of Groningen) for kindly providing *E. coli* TG90. We thank BaseClear (NL) for their help in analysing the RNA-sequencing data and Metabolomic Discoveries (DE) for their help in analysing the metabolomics data. This work was financially supported by Corbion.

SUPPLEMENTARY DATA

Table S1. Absolute and relative expression values in *B. smithii* ET 138 strains.
Separate Excel file, available on request.

Table S2. Fermentation data for metabolomics samples.

Strain	Nr[1]	Time (h)	Products (mM)					OD_{600}	Cell count/mL	Dry weight (g/L)	Sample volume (mL)	Filtration time (s)
			Lac	Ace	Pyr	Mal	Suc					
DKO	2	2	0.11	0.00	0.00	0.86	0.00	0.145	445313	0.450	50	10
DKO	2	*21*	*1.52*	*3.55*	*2.28*	*0.15*	*0.06*	*0.317*	*1437500*	*0.635*	50	20
DKO	3	2	0.59	0.52	0.18	0.25	0.00	0.144	304688	0.425	50	10
DKO	3	21	1.18	3.21	2.49	1.19	0.00	0.270	1148438	0.340	50	16
DKO	*4*	*2*	*0.84*	*0.13*	*0.15*	*0.27*	*0.00*	*0.115*	*406250*	*0.325*	50	15
DKO	4	21	2.39	2.62	2.42	0.00	0.00	0.234	1242188	0.315	50	25
WT	2	2	2.95	0.58	0.12	0.00	0.28	0.253	1187500	0.435	50	15
WT	2	21	73.96	4.59	0.21	1.47	0.78	1.045	4237500	0.530	15	21
WT	3	*2*	*3.31*	*0.38*	*0.10*	*0.05*	*0.00*	*0.323*	*1140625*	*0.355*	50	43
WT	3	21	74.09	3.82	0.28	3.65	0.00	0.923	6640625	0.485	15	39
WT	4	2	3.93	0.48	0.07	0.11	0.00	0.315	1335938	0.485	50	36
WT	*4*	*21*	*69.18*	*4.13*	*0.00*	*2.36*	*0.25*	*0.843*	*6328125*	*0.600*	15	29

[1] The number indicates the replicate number (nr 1 was discarded due to slow sampling).
Samples not used for quantitative measurements by the targeted approach are indicated in italics.
Abbreviations: DKO: strain ET 138 Δ*ldhL*-Δ*sigF* ('double knockout'); WT: strain ET 138 wild-type; Lac: lactate; Ace: acetate; Pyr: pyruvate;Mal: malate; Suc: succinate.

Table S3. Comparison of both measurement platforms via Pearson's coefficients.

	untargeted measurements				targeted measurements				Pearsons
Metabolite extracellular	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**coefficient**
Alanine	0.24	0.20	0.29	1.48	126.41	132.64	164.55	756.94	1.00
Asparagine	0.14	0.07	0.17	0.06	23.83	0.36	34.43	2.08	0.99
Aspartic acid	0.20	0.24	0.15	0.08	27.36	0.00	84.09	0.00	-0.07
Glutamic acid	0.13	0.09	0.15	0.08	50.82	0.00	95.31	0.00	0.97
Glycine	1.61	0.09	1.55	1.60	58.57	2.15	65.02	86.10	0.95
Isoleucine	0.36	0.01	0.45	0.12	52.30	0.00	67.45	32.04	0.97
Leucine	237.27	4.07	247.27	113.59	89.81	0.62	122.41	52.44	0.97
Lysine	0.84	0.31	0.78	0.31	57.17	1.75	57.57	48.84	0.69
Phenylalanine	0.04	0.02	0.08	0.03	29.39	1.07	47.26	9.86	0.97
Proline	0.06	0.01	0.06	0.02	18.87	0.00	21.66	14.35	0.89
Serine	0.42	0.03	0.42	0.02	54.36	0.00	78.82	0.00	0.97
Threonine	0.25	0.07	0.31	0.05	44.14	0.00	59.96	12.00	0.97
Tryptophan	1.77	1.84	2.06	1.22	6.46	0.88	9.30	4.63	0.32
Tyrosine	3.01	3.42	2.89	1.17	34.31	1.02	45.24	20.80	-0.07
Valine	0.78	0.07	0.91	0.31	74.77	8.98	93.50	58.40	0.94
	untargeted measurements				**targeted measurements**				**Pearsons**
Metabolite extracellular	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**coefficient**
Fumaric acid	12.10	25.85	11.78	12.31	0.86	14.30	0.87	1.34	1.00
Malic acid	0.03	0.22	0.03	0.01	2.11	29.65	2.36	1.96	1.00
Succinic acid	0.24	1.29	0.34	0.23	42.72	259.18	40.58	40.81	1.00
D-alpha,alpha'-Trehalose	1.61	1.00	2.12	0.97	39.99	0.00	65.60	34.11	0.84
dihydroxyacetone phosphat	0.96	6.83	1.12	1.53	0.58	0.93	0.20	0.59	0.80
	untargeted measurements				**targeted measurements**				**Pearsons**
Metabolite intracellular	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**coefficient**
Alanine	0.01	0.03	0.00	0.01	0.00	3.29	0.00	0.61	0.99
Glycine	0.01	0.03	0.02	0.01	0.00	0.13	0.07	0.10	0.89
Valine	0.00	0.01	0.00	0.01	0.00	0.23	0.00	0.12	0.99
Leucine	0.00	0.00	0.00	0.01	0.03	0.05	0.06	0.17	0.99
Isoleucine	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.07	0.97
Proline	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Serine	0.00	0.01	0.00	0.00	0.00	0.15	0.00	0.00	0.98
Threonine	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.13	0.90
Methionine	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.03	0.81
Aspartic acid	0.00	0.03	0.00	0.04	0.00	0.48	0.00	0.00	0.43
Phenylalanine	0.00	0.05	0.00	0.01	0.02	0.45	0.03	0.06	1.00
Glutamic acid	0.00	0.01	0.00	0.17	0.00	0.00	0.00	1.85	1.00
Lysine	0.01	0.21	0.02	0.06	0.25	3.87	0.12	0.88	1.00
Asparagine	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	
Tyrosine	0.00	0.02	0.01	0.01	0.02	0.38	0.02	0.17	0.62
Histidine	2.71	3.20	2.55	4.67	0.00	0.06	0.00	0.06	0.73
Tryptophan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.96
Arginine	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	untargeted measurements				**targeted measurements**				**Pearsons**
Metabolite intracellular	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**aver wt 2**	**aver wt 21**	**aver dko 2**	**aver dko 21**	**coefficient**
Glucose-6-phosphate	0.00	0.01	0.00	0.02	0.00	0.19	0.00	0.20	0.80
Fumaric acid	7.92	8.70	8.37	8.54	0.03	0.36	0.01	0.12	0.74
Malic acid	0.00	0.01	0.00	0.00	0.00	0.26	0.00	0.00	0.99
2-oxo-Glutaric acid	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
dihydroxyacetone phosphat	0.63	0.76	0.65	0.99	0.15	0.39	0.12	1.02	1.00

Untargeted measurements are in relative peak areas; targeted measurements are in µmoles. 'aver' means 'average', in which for both measurements the same duplo is used. Values in blue were below the detection limit in the targeted approach and therefore set to 0 for calculations. Empty cells for the Pearson's coefficient mean that it could not be calculated due to values being 0.

Colour legend:
0.90-1.00
0.70-0.89
0.50-0.69
0.00-0.49
<0.00

Table S4. Extracellular amino acid measurements in *B. smithii* fermentations.

	Extracellular amino acid concentration (µM)																		
name	Ala	Gly	Val	Leu	Ile	Pro	Ser	Thr	Meth	Asp	Phe	Cys	Glu	Lys	Asn	Gln	Tyr	His	Trp
WT t2 average	126.41	58.57	74.77	89.81	52.30	18.87	54.36	44.14	6.46	27.36	29.39	0.57	50.82	57.17	23.83	24.89	34.31	11.88	6.46
WT t2 stdev	*23.94*	*3.54*	*12.22*	*17.87*	*9.34*	*1.49*	*12.38*	*8.02*	*2.46*	*3.13*	*5.14*	*0.03*	*15.64*	*9.11*	*7.44*	*6.66*	*5.85*	*1.32*	*1.64*
WT t21 average	132.64	2.15	8.98	0.62	0.00	0.00	0.00	0.00	0.00	0.00	1.07	0.11	0.00	1.75	0.36	2.07	1.02	0.00	0.88
WT t21 stdev	*31.69*	*0.45*	*5.91*	*0.18*	<	<	<	<	<	<	*0.62*	*0.02*	<	*0.50*	*0.01*	<	*0.61*	<	*0.08*
DKO t2 average	164.55	65.02	93.50	122.41	67.45	21.66	78.82	59.96	16.15	84.09	47.26	0.79	95.31	57.57	34.43	27.09	45.24	14.52	9.30
DKO t2 stdev	*33.12*	*14.76*	*18.72*	*13.38*	*11.54*	*4.00*	*6.19*	*8.80*	*0.97*	*1.40*	*1.55*	*0.06*	*12.20*	*25.50*	*2.14*	*12.28*	*6.19*	*1.67*	*0.83*
DKO t21 average	756.94	86.10	58.40	52.44	32.04	14.35	0.00	12.00	4.22	0.00	9.86	0.34	0.00	48.84	2.08	13.29	20.80	8.23	4.63
DKO t21 stdev	*202.45*	*4.68*	*12.54*	*9.05*	*4.79*	*1.10*	<	*3.47*	*1.01*	<	*5.80*	*0.02*	<	*3.99*	*1.08*	*0.99*	*3.66*	*0.63*	*1.12*
WT difference 21-2 h	6.23	-56.41	-65.79	-89.18	-52.30	-18.87	-54.36	-44.14	-6.46	-27.36	-28.32	-0.46	-50.82	-55.42	-23.47	-22.83	-33.29	-11.88	-5.59
DKO difference 21-2 h	592.39	21.08	-35.10	-69.97	-35.41	-7.32	-78.82	-47.97	-11.92	-84.09	-37.40	-0.45	-95.31	-8.72	-32.35	-13.80	-24.44	-6.30	-4.67

Values in italics are standard deviations, all values are averages of duplicate experiments and technical triplicates. Red values were above the lowest standard. A '<' indicates a value that was below the lowest standard and could not be measured.
Abbreviations: DKO: strain ET 138 Δ*ldhL*-Δ*sigF* ('double knockout'); WT: strain ET 138 wild-type.

Table S5. Extracellular metabolite measurements in *B. smithii* ET 138 fermentations.

	Extracellular metabolite concentration (µM)																					
name	Pyr	Fum	Mal	Glc	aKG	2PG	DHAP	3PG	E4P	Rib5P	Ribu5P	Xyl5P	F6P	G6P	S7P	Tre	FBP	PEP	Suc	G1P	T6P	G3P	CDW (g)
WT t2 average	67.57	0.86	2.11	>	0.91	<	0.58	<	<	<	<	<	<	5.51	<	39.99	0.02	0.05	42.72	0.02	0.31	0.23	0.46
WT t2 stdev	*5.56*	*0.15*	*0.36*	<	<	<	*0.21*	<	<	<	<	<	<	*0.59*	<	*9.01*	*0.03*	*0.01*	*6.13*	*0.01*	*0.00*	*0.04*	*0.04*
WT t21 average	209.59	14.30	29.65	>	<	<	0.93	<	<	<	<	<	<	8.86	<	<	0.41	0.55	259.18	0.14	0.38	2.35	0.51
WT t21 stdev	*63.21*	*2.44*	*2.42*	<	<	<	*0.13*	<	<	<	<	<	<	*1.16*	<	<	*0.17*	*0.12*	*14.32*	*0.02*	*0.09*	*0.30*	*0.03*
DKO t2 average	146.18	0.87	2.36	>	0.85	<	0.20	<	<	<	<	<	<	6.09	<	65.60	0.01	0.01	40.58	0.04	0.30	0.12	0.44
DKO t2 stdev	*191.42*	*0.29*	*0.60*	<	*0.40*	<	*0.22*	<	<	<	<	<	<	*1.22*	<	*18.57*	*0.02*	*0.01*	*2.99*	*0.03*	*0.03*	*0.03*	*0.02*
DKO t21 average	1510.93	1.34	1.96	>	<	<	0.59	<	<	<	<	<	<	4.60	<	34.11	0.51	0.04	40.81	0.15	0.17	0.18	0.33
DKO t21 stdev	*152.97*	*0.30*	*0.14*	<	<	<	*0.21*	<	<	<	<	<	<	*0.06*	<	*4.93*	*0.09*	*0.02*	*10.60*	*0.05*	*0.06*	*0.05*	*0.02*

Values in italics are standard deviations, all values are averages of duplicate experiments and technical triplicates. Orange values were close to or below the lowest standard; red values were above the highest standard. A '<' indicates a value that was below the lowest standard and could not be measured, a '>' means above highest standard. Glyceraldehyde-3P and Citrate measurements were compromised by high glucose concentrations.
Abbreviations: DKO: strain ET 138 Δ*ldhL*-Δ*sigF* ('double knockout'); WT: strain ET 138 wild-type.

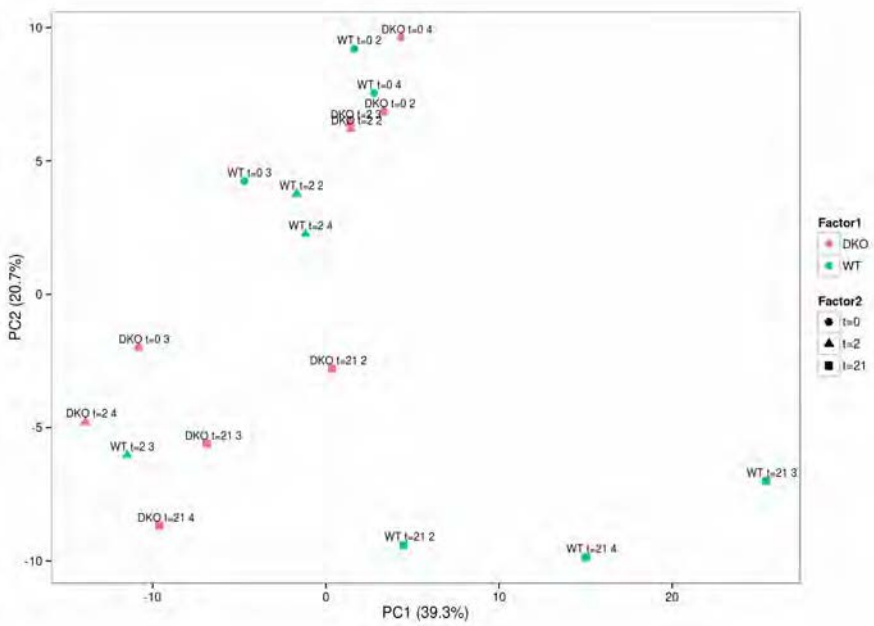

Figure S1. PCA plot of extracellular samples. PC1 represents the differences caused by the strain and PC2 represents differences caused by time (in hours, with time points being 0 h, 2 h and 21 h). The differences between the strains (PC1) are responsible for 39.3% of all variances and the difference in time (PC2) is responsible for 20.7% of all variances found in the samples. Abbreviations: DKO: strain ET 138 Δ*ldhL*-Δ*sigF* ('double knockout'); WT: strain ET 138 wild-type.

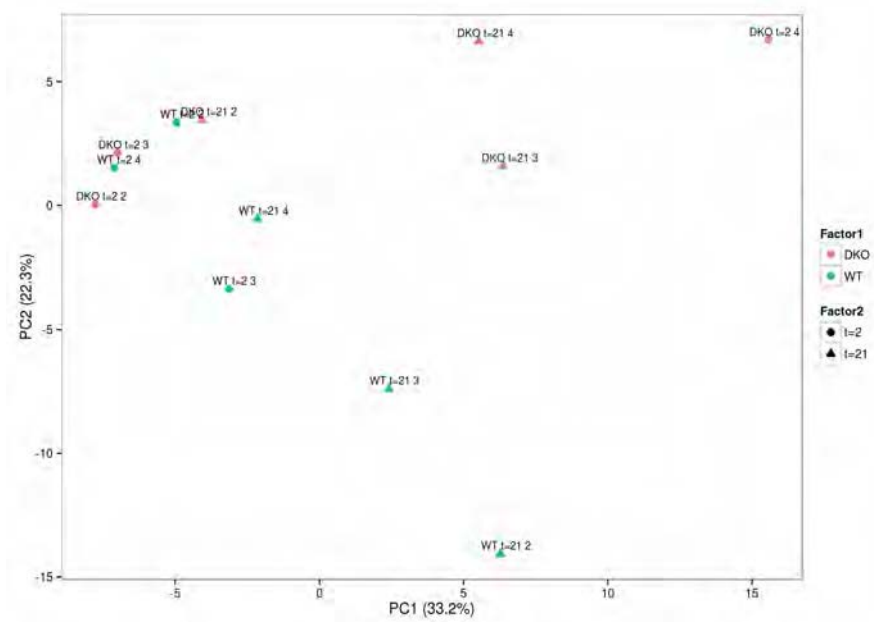

Figure S2. PCA plot of intracellular samples. PC1 represents the differences caused by the strain and PC2 represents differences caused by time (in hours, with time points being 0 h, 2 h and 21 h). The differences between the strains (PC1) are responsible for 33.2% of all variances and the difference in time (PC2) is responsible for 22.3% of all variances found in the samples. Abbreviations: DKO: strain ET 138 Δ*ldhL*-Δ*sigF* ('double knockout'); WT: strain ET 138 wild-type.

Chapter 7

Thesis summary and general discussion

THESIS SUMMARY

Due to the globally increasing demand for chemicals and fuels and the high environmental impact and limited amount of fossil resources, there is a growing interest in green chemicals and fuels derived from renewable resources. As described in **Chapter 1**, one of the most feasible alternatives on the short term is microbial conversion of the sugars in biomass to fuels and chemicals in a biorefinery. Especially the microbial production of chemicals from biomass is attractive as the biomass demand is lower than for fuels, thus requiring only a fraction of the biomass, while the profits on chemicals are significantly larger and markets for green chemicals are expected to grow 2-3 times faster than those for bio-fuels. One of the major bottlenecks in green chemical and fuel production is that most current production processes use sugars derived from food resources, which results in competition with the food and feed chain. To be economically and ethically feasible, non-food biomass should be used as a resource, which is often difficult with currently used production organisms. Also, to be economically feasible, the costs of green chemicals and fuels need to be further reduced to be below the costs of products based on fossil resources. To do so, other organisms than the currently most-used platform organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* should be used. Ideally, this alternative organism is genetically accessible, has high productivity, titre and yield, is flexible in carbon source, robust, moderately thermophilic, acidophilic, facultatively anaerobic and has little nutritional requirements. The organisms that come closest to these criteria are thermophilic bacilli, which form a diverse class of organisms in the family of *Bacillaceae*. This thesis describes the isolation, characterization and metabolic engineering of *Bacillus smithii*, a novel potential thermophilic platform organism.

Chapter 2 provides more detail on the use of thermophilic microorganisms as platform organisms for green chemical production in a biorefinery concept. Non-food biomass consists of tightly packed lignocellulose, which needs to be degraded to fermentable monomeric sugars by hydrolytic enzymes. In the biorefinery, hydrolytic enzyme mixtures can either be added in a separate reactor in the biorefinery after which the sugars are fermented by the microbes in another reactor (separated hydrolysis and fermentation, SHF), or they can be added to the same reactor containing the microorganisms (simultaneous saccharification and fermentation, SSF). Alternatively, the enzymes can be produced by the organism itself (consolidated bioprocessing, CBP), but CBP is still mostly being developed at laboratory scale. SSF requires less reactors than SHF and is therefore a potentially economically interesting setup if the optimum conditions for the enzymes and the production organism are similar. As commercially available enzyme mixtures have their optimum temperature around 50-60°C, using a moderately thermophilic organism would reduce the costs of the SSF process by reducing the amount of required enzyme. Also, thermophilic processes are

less prone to contaminations, and substrate and product solubility are increased. Ideally, the organism is not strictly anaerobic but facultatively anaerobic, as aeration is costly, but fully anaerobic conditions inhibit SSF-enzymes. Several successful examples of the application of facultatively anaerobic thermophiles for green chemical production from lignocellulose in an SSF setting are for example *Bacillus coagulans* for lactic acid production and *Bacillus licheniformis* for 2,3-butanediol production. However, whereas strongly developed genetic toolboxes are available for current mesophilic production organisms, these tools are still in their infancy for thermophilic organisms. Such tools are required to optimize production and to study metabolism. Thermophilic organisms show a wide variety in metabolism and in many cases the metabolism of these organisms is still poorly understood, hampering full optimization. Chapter 2 furthermore provides an overview of transformation, integration and counter-selection methods currently used for thermophiles, such as the Cre-*lox* system, *pyrF* counter selection and double homologous recombination systems based on *pta-ack* and *tdk-hpt*. Although several deletion mutants have been constructed using these methods, not all of them are entirely markerless and most are not suited as high-throughput engineering tools, stressing the need for further research in this area.

Despite several facultatively anaerobic thermophiles being described as genetically accessible, this feature is still one the major bottlenecks in developing these organisms into platform organisms. Therefore, in **Chapter 3**, we set out to isolate a facultatively anaerobic, moderately thermophilic bacterium that was genetically accessible and produced high titers of organic acids. A total of 267 strains of different thermophilic bacilli species were isolated from compost and screened for C_5 and C_6 sugar utilization and acid production. The 44 best strains were screened for genetic accessibility via electroporation. Only 3 strains tested positive for this, namely *Geobacillus thermodenitrificans* strains ET 144-2 and ET 251 and *B. smithii* strain ET 138. In subsequent evaluations in lab-scale bioreactors at 55°C and pH 6.5 on glucose, the two *G. thermodenitrificans* strains performed poorly whereas *B. smithii* performed well with high titers, yields and productivity of mainly lactate. In similar lab-scale reactors, this strain also performed well on xylose and at pH 5.5 and was still able to perform for 48 at pH 4.5. The electroporation protocol for this strain was optimized, resulting in a maximum efficiency of $5x10^3$ colonies per µg plasmid pNW33n. Two other *B. smithii* strains, among which the type strain DSM 4216^T, were also shown to be transformable with pNW33n. This is the first time that genetic accessibility is described for *B. smithii* and it is the first step towards developing it into a platform organism, for which it appears to be suitable based on its efficient C_5 and C_6 sugar utilization and acid production profile.

In order to perform metabolic engineering and create a platform organism, the genome sequence of the organism needs to be determined. Therefore, **Chapter 4** describes the complete genome sequence of *B. smithii* DSM 4216^T. The genome was sequenced using Illumina HiSeq2000 mate-pair and paired-end sequencing, followed by PacBio sequen-

cing to close remaining gaps. The genome consists of a 3,368,778 bp chromosome and a 12,514 bp plasmid, together encoding 3886 genes, and has a GC-content of 40.8%. The complete pathways for glycolysis, pentose phosphate pathway and TCA-cycle were identified, but not those for the phosphoketolase and Entner-Doudoroff pathways. After the initial RAST-annotation of the completed genome sequence, several canonical genes were not identified and a second round of annotation was performed via analysis of protein domains. This method identified candidate genes for the methylglyoxal pathway to both D- and L-lactate and confirmed the absence of the genes for the canonical acetate production pathway in bacteria formed by phosphotransacetylase *pta* and acetate kinase *ack*, as well as the absence of pyruvate-formate lyase *pfl*. Whereas the latter is in line with the absence of formate production and observed more commonly in bacilli, the absence of *pta* and *ack* was remarkable as acetate is the second product of *B. smithii* under micro-aerobic conditions and its main product under aerobic conditions, and the two genes are present in all bacilli genomes known so far. Both genes were also absent in the publically available genome of *B. smithii* strain 7_3_47FAA, which was sequenced during a metagenome study, and in the genome of our isolate strain ET 138. This suggests that the absence of these genes is typical for *B. smithii* and that its metabolism is different from closely related species such as *B. coagulans*.

In order to become a platform organism and to study its atypical metabolism, a genetic toolbox needs to be established for *B. smithii*. **Chapter 5** describes the development of a markerless gene deletion method for *B. smithii*. For strains ET 138 and DSM 4216^T, the *ldhL* gene was markerlessly removed via double homologous recombination using plasmid pNW33n. Despite the replicative nature of this plasmid at 55°C, single crossover integrants were readily obtained. After several transfers on a more defined medium containing acetate or lactate, mixtures were obtained of single crossovers via both upstream and downstream flanks and double crossover wild-type and mutant genotypes. A pure double crossover deletion mutant was obtained after several transfers and PCR-based screenings. As the presence of mixed genotypes made the process very laborious, a counter-selection method was established in strain ET 138 to select against plasmid presence and more easily obtain pure double crossovers. We used the *lacZ*-counter-selection system, which is based on the toxicity of high X-gal concentrations in the presence of the plasmid-encoded *lacZ* gene. Using this method, the sporulation-specific sigma factor *sigF* and pyruvate dehydrogenase complex E1-α *pdhA* were consecutively removed from the *B. smithii* ET 138 genome in a markerless way. An initial evaluation of the growth and production profiles of the mutant strains in tubes showed that removal of the *ldhL* gene eliminates L-lactate production and causes a severe decrease in anaerobic growth and production capacities. This is different from *ldhL* mutants of its close relative *B. coagulans*, which produces ethanol or 2,3-butanediol, suggesting that alternative NAD^+-regenerating pathways are not sufficiently

available in *B. smithii*. This is in line with the absence of genes encoding acetolactate decarboxylase *alsD* and a bifunctional alcohol dehydrogenase *adhE* from the *B. smithii* genome. *B. smithii* mutants lacking the *sigF* gene were unable to sporulate and removal of the *pdhA* gene eliminated acetate production and rendered the strains auxotrophic for acetate. The latter confirms the observation that a *pfl* gene is not encoded in the *B. smithii* genome and that pyruvate dehydrogenase is the only route from pyruvate to acetyl-CoA.

In **Chapter 6**, we continue the characterization of the metabolism of the mutant strains described in Chapter 5 in a more detailed way via lab-scale fermentations, transcriptomics, metabolomics, enzyme assays and overexpression of several genes. Transcriptomics and metabolomics suggest that a major bottleneck is created at glyceraldehyde-3-phosphate dehydrogenase, the NADH-producing step in glycolysis. Compared to the wild-type, the mutant accumulates metabolites upstream of this enzyme and shows a decrease in the concentrations of metabolites downstream of it. Both the genes encoding the glycolytic NADH-producing GapA and the gluconeogenic $NADP^+$-producing GapB were upregulated, as well as most other genes involved in glycolysis, gluconeogenesis and the pentose phosphate pathway. Further downstream in the metabolism, transcriptomics suggested a cycle around pyruvate via highly upregulated expression of pyruvate dehydrogenase, the glyoxylate shunt and phosphoenol pyruvate carboxykinase. Plasmid-based heterologous expression of an NAD^+-generating D-lactate dehydrogenase partially restored growth and production capacities in *B. smithii* Δ*ldhL*. To evaluate another NAD^+-regenerating route and direct flux towards succinate, native pyruvate carboxylase *pyc* and NAD^+-generating malate dehydrogenase *mdh* were overexpressed in *B. smithii* Δ*ldhL*. Only a minor increase in malate and succinate levels was observed and pyruvate production remained high. Subsequent enzyme assays showed an increased activity of malic enzyme in the strains overexpressing *pyc* compared to the empty vector strains. This is in line with the cycle around pyruvate suggested by the transcriptomics data. Altogether, these data suggest that NAD^+-regeneration is a major but possibly not the only bottleneck in the *B. smithii* strains lacking *ldhL*, and that metabolism is strongly directed to pyruvate. This provides insight into the metabolism of *B. smithii*, which appears to be different from closely related species, and provides directions for future engineering steps.

GENERAL DISCUSSION

Green alternatives for fossil resource-based fuels and chemicals are gaining increased attention. As described in Chapter 1 and 2, several alternatives are available and the feasibility differs per application, with microbial conversion of biomass being one of the most feasible alternatives for green chemical production. The choice of the production organism is important for many factors in the process and this is a rapidly developing research field. In this thesis, we described the development of *B. smithii* as a novel production host for green chemical production. After isolating *B. smithii* ET 138, and developing a genetic system for and performing metabolomics and transcriptomics studies on this strain, there are still many open questions about its metabolism and the genetic system needs further improvement. This final chapter will place the preceding chapters in perspective by discussing the relevance of screening for novel production organisms, the variety in metabolism and what this implies for engineering strategies in these new organisms in general and *B. smithii* in particular, as well as directions for future development of *B. smithii* as production host and its required genetic tools.

THE IDEAL PRODUCTION ORGANISM AND THE IMPORTANCE OF SCREENING NEW ORGANISMS

Current model organisms such as *E. coli* and *S. cerevisiae* have been engineered to achieve very high titres of a wide range of both natural and unnatural compounds. The metabolic flexibility exhibited by these organisms in combination with metabolic engineering is astonishing. Their long history of study has resulted in a degree of understanding of these organisms that makes it possible to achieve remarkable results. On one hand the development of a novel organism gives ample advantages as 'you get what you screen for': enabling the selection of an organism that meets desired criteria as listed in Chapter 1. On the other hand a novel organism poses a large challenge as the history and concomitant degree of understanding present for model organisms is lacking and all tools and knowledge has to be built from scratch, which is time-consuming and can be risky. In this regard, especially genetic accessibility is a crucial, tricky and often neglected feature. In many studies, an organism is isolated that has many advantageous properties, that ferments very fast and is well capable of producing a certain compound, but very often the subsequent attempts to develop a genetic system fail, disabling the use of the strain for anything else than its natural product. Only few studies take genetic accessibility into account when isolating novel organisms. Narumi *et al.* describe the screening of 52 isolates and 15 commercially available strains of *G. stearothermophilus* (at that time called *Bacillus stearothermophilus*) for electroporation (Narumi et al., 1992). The authors do not describe the total number of transforma-

ble strains but only the optimization of the protocol for the strain that was best transformable after the screening (Narumi et al., 1992). Patel *et al.*, on the other hand, do not mention how many strains they screened exactly, but only note that they found one *B. coagulans* isolated to be transformable (Patel et al., 2006). Kita *et al.* initially developed a genetic system for the thermophilic autotroph *Moorella thermoacetica* that required pre-methylation of the plasmid. Subsequently they started screening for strains that were more easily genetically accessible and did not require pre-methylation (Kita et al., 2013). The authors did not only screen for transformation, but immediately for integration of an antibiotic resistance gene in the genome via the *pyrF*-system (Kita et al., 2013). The genome sequence of the easier-transformable strain revealed that this strain lacks a restriction-modification (RM-)system that likely explains the better transformability (Tsukahara et al., 2014). We have made similar observations in *B. smithii*, in which the type strain is less transformable than our isolate and also encodes an extra RM-system (Chapter 4). This was further supported by the fact that about 10-fold more colonies were obtained for the *B. smithii* type strain when it was transformed with pNW33n plasmid DNA isolated from a transformed *B. smithii* DSM 4216^T strain instead of *E. coli* DH5α (unpublished results). The strong strain-specificity observed by us and the studies mentioned here was also suggested by Studholme *et al.*, who performed a phylogenetic analysis of transformable thermophilic bacilli and indicate that the feature is not limited to a certain species but rather more widely distributed and strain-specific (Studholme et al., 1999), further stressing the importance of screening multiple strains of a species.

In our experience, the two major bottlenecks when screening for the ideal novel platform organism were i) the combination of genetic accessibility and good fermentation performance, and ii) the reproducibility (or: robustness) of growth and fermentation performance of the strains. Regarding the first point, in general there often seems to be a trade-off between genetic accessibility and other key properties such as fermentation performance, sugar utilization and growth conditions (especially nutrient and oxygen requirements). The two genetically accessible strains of *G. thermodenitrificans* described in Chapter 3 were performing very poorly in fermentations compared to other strains, but these others were not genetically accessible. A similar remark was made by Patel *et al.*, who found only one *B. coagulans* isolate to be genetically accessible. This isolate was performing less well than other isolates in fermentations, but was nevertheless used for further studies (Patel et al., 2006). Regarding the second point, we found several strains that on some occasions outperformed our selected strain *B. smithii* in lab-scale fermentations (Chapter 3, data not shown) but on other occasions hardly grew under the exact same conditions. These issues have also been touched upon recently by Hussein *et al.* (Hussein et al., 2015) for *G. thermoglucosidans*. As reproducibility and reliability is a key feature of a robust production strain, this was a major criterion for our selection for both fermentation and genetic accessibility. The reproducibility might increase with increasing understanding of the metabolism,

as very little is known about the reasons for the irreproducibility (Hussein et al., 2015), which is a problem less frequently encountered when working with well-known model organisms.

The main advantage of screening for a novel organism is that criteria can be set such that it has characteristics that are crucial for an efficient process but that are lacking in currently used organisms. As the metabolic diversity in nature is very large, this gives ample opportunities to find a suitable organism, although being able to cultivate these organisms is already a challenge in itself and the subsequent understanding of its metabolism will be laborious. In order to make full use of the capacities of the organism, its metabolism needs to be fully understood, which is a time-consuming but exciting process.

Understanding metabolism, and consequences of the diversity in central carbon metabolism for metabolic engineering

During the course of the study described in this thesis, the major bottleneck in developing our new isolate into a platform organism and especially for succinate production, was the lack of understanding of its metabolism. Its metabolism shows some remarkable differences to closely related species, such as the lack of the canonical acetate production pathway (Chapter 4, 5 and 6). This is an observation made more frequently with novel organisms, as the metabolic diversity is large and most organisms differ at crucial points from well-known model organisms such as LAB, *B. subtilis* and *E. coli*. Also differences between Gram-positive and Gram-negative organisms should not be underestimated – as *E. coli* is still the most-used organism for metabolic engineering for green chemical production, most knowledge is on this organism. Regulation and pathways can differ significantly with Gram-positives, making it impossible to extrapolate *E. coli* results to predict the behaviour of bacilli. A well-known example is the different role of the PTS-system in catabolite repression (Saier et al., 1996). Also within a group of more closely related organisms there is a wide variety in metabolism, as briefly touched upon in Chapter 1. Besides differences in for example carbon source utilization capacities, also seemingly subtle differences such as the absence or presence of one gene can have a large impact on the outcome of metabolic engineering and carbon fluxes in general.

Especially around the phosphoenol pyruvate-pyruvate-oxaloacetate (PEP-pyr-OAA) node there is variety of single gene differences that have a large impact on carbon flows and engineering strategies, as this is a crucial branching point for many valuable metabolites and an important switch and regulation point between glycolysis, gluconeogenesis and the TCA-cycle (Sauer & Eikmanns, 2005) (Figure 1). For example, *C. thermocellum* lacks a pyruvate kinase gene to form pyruvate from PEP, while this gene is present in close relatives such as *T. saccharolyticum* (Deng et al., 2013; Zhou et al., 2013a). Two candidate genes present in the genome of *C. thermocellum* that could convert PEP to pyruvate were pyruva-

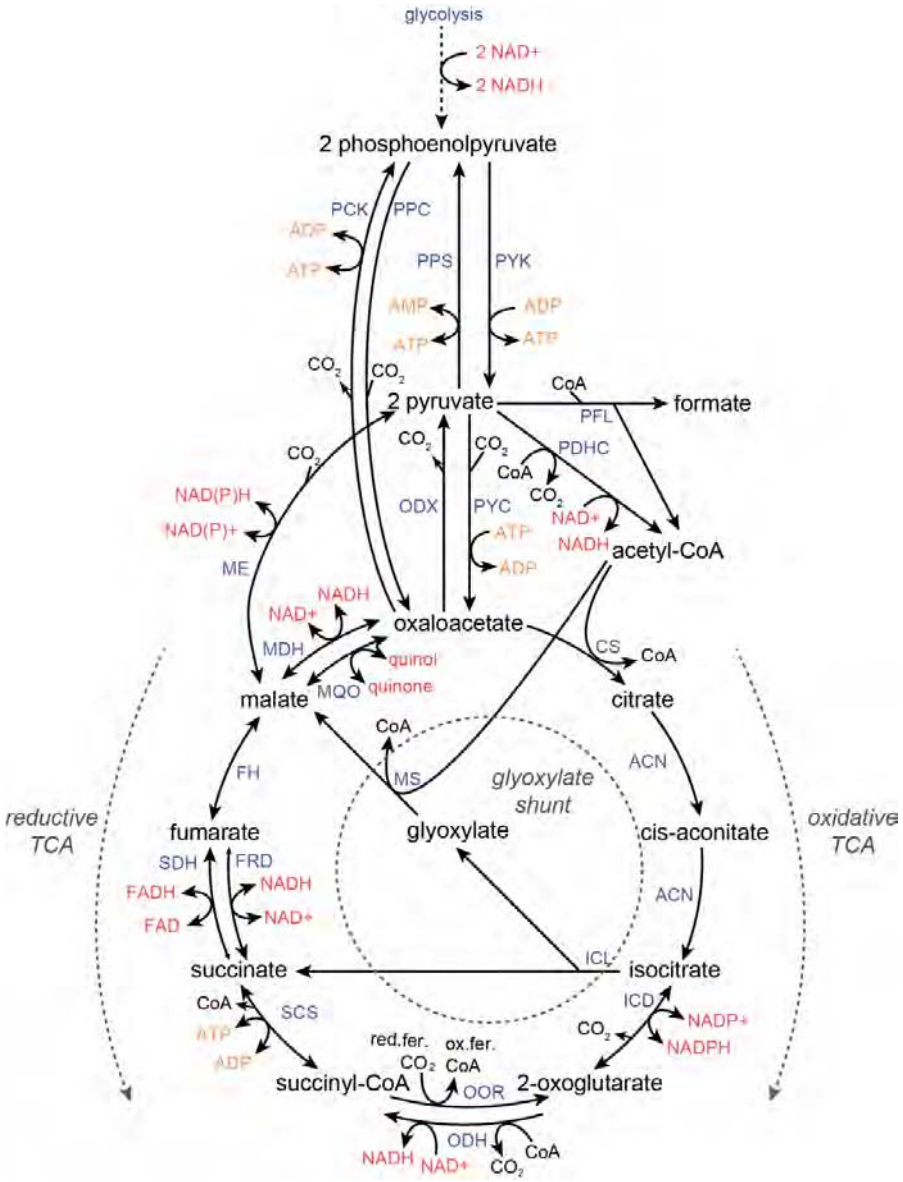

Figure 1. General overview (not organism-specific) of enzymes involved in the PEP-pyr-OAA node and in the different parts of the TCA-cycle and glyoxylate shunt. Based on (Sauer & Eikmanns, 2005). Abbreviations: PCK: phosphoenol pyruvate carboxykinase; PPC: phosphoenol pyruvate carboxylase; PPS: phosphoenol pyruvate synthase; PYK: pyruvate kinase; ODX; oxaloacetate decarboxylase; PYC: pyruvate carboxylase; PDHC: pyruvate dehydrogenase complex; ME: malic enzyme; MDH: malate dehydrogenase; MQO: malate:quinone oxidoreductase; CS: citrate synthase; ACN: aconitase; ICL: isocitrate lyase; MS: malate synthase; ICD: isocitrate dehydrogenase; OOR: 2-oxoglutarate reductase; ODH: 2-oxoglutratae dehydrogenase; SCS: succinyl-CoA synthetase; FRD: fumarate reductase; SDH: succinate dehydrogenase; FH: fumarate hydratase.

te-phosphate dikinase (*ppdk*) and PEP synthase (*peps*). The latter one was hardly expressed and therefore assumed not to play a critical role in glycolysis, while *ppdk* was also not critical as deletion of this gene did not affect growth (Zhou et al., 2013a). It was then suggested that a so-called 'malate shunt' was responsible for pyruvate formation via GDP-linked PEP carboxykinase to convert PEP to oxaloacetate, NADH-linked malate dehydrogenase to convert oxaloacetate to malate and $NADP^+$-linked malic enzyme to convert malate into pyruvate (Zhou et al., 2013a). The presence of the malate shunt instead of a pyruvate kinase (*pyk)* gene had important implications for the subsequent engineering of the strain towards ethanol production. Initial attempts to increase ethanol yields in *C. thermocellum* by deletion of by-product pathways such as acetate and lactate did not increase ethanol yields (Argyros et al., 2011) while this had been a successful strategy in *T. saccharolyticum* (Shaw et al., 2008). Only when NADH availability was increased by expressing the *pyk* gene from *T. saccharolyticum* in *C. thermocellum* in combination with either decreased PEP carboxykinase expression or deleted malic enzyme and partially deleted malate dehydrogenase (Figure 1), the ethanol yield increased 3-fold (Deng et al., 2013). Overall, glycolysis in *C. thermocellum* seems to be atypical with a GTP- instead of ATP-dependent glucokinase, a PPi- instead of an ATP-dependent phosphofructokinase and the replacement of pyruvate kinase by the malate shunt (Zhou et al., 2013a). All these findings were only confirmed after extensive characterization via enzyme assays, as in initial genome annotations several of the genes were predicted to be present but never experimentally verified (Zhou et al., 2013a), stressing the importance of experimental validation of genome annotation.

Another example of a major effect caused by the presence or absence of genes around PEP-pyr-OAA is with PEP carboxylase *(ppc*), PEP carboxykinase (*pck*) and pyruvate carboxylase (*pyc*) (Figure 1). Whereas *E. coli* encodes both glycolytic *ppc* and gluconeogenic *pck* but does not encode *pyc*, most bacilli do not encode a *ppc* but do have *pck* and *pyc*. This has implications for the effect of metabolic engineering strategies, especially those targeted at increasing flux into the TCA cycle, which can also be used to decrease the amount of acetate produced. If not enough oxaloacetate is present to form citrate together with acetyl-CoA to run the TCA-cycle, the acetyl-CoA is used for the production of acetate. In order to increase the oxaloacetate pool and with that the TCA-cycle flux in *E. coli*, *ppc* was overexpressed (March et al., 2002). Since this enzyme uses PEP which is also used by the PTS-system for glucose uptake, glycolysis rate was decreased by 30% in these cells. Heterologous expression of *pyc*, however, hardly affected glycolytic rates while decreasing acetate production and increasing heterologous protein production, suggesting a better flux through the TCA-cycle to support the extra demand for amino acids (March et al., 2002).

As described in Chapter 4 and 5, several genes involved in pyruvate metabolism commonly found in close relatives of *B. smithii* are not present in this organism, such as pyruvate-formate lyase (*pfl*) for formate production, acetolactate synthase (*alsS*) and dehydroge-

nase (*alsD*) for 2,3-butanediol/acetoin production, phosphotransacetylase (*pta*) and acetate kinase (*ack*) for acetate production and the bifunctional alcohol dehydrogenase (*adhE*) for ethanol production. Consequently, mutants lacking *ldhL* do not have sufficient other means to regenerate NAD^+ and a mutant lacking *pdhA* is acetate auxotroph because it cannot use *pfl* as an alternative to generate acetyl-CoA. On one hand, this limits the possibilities of making these products with *B. smithii* as the genes need to be heterologously inserted. On the other hand, it creates an ideal platform organism since only one gene needs to be deleted to eliminate production and other genes linked to NAD^+ and product formation can be inserted, as was shown for D-lactate production in Chapter 6. The above-mentioned differences in pyruvate metabolism, like for *C. thermocellum,* might also play an important role in *B. smithii*, as shown by the malate shunt-like behaviour of the *ldhL* mutants suggested by transcriptomics via *pdhc*-glyoxylate shunt-*mdh*-*pck* and *ldhL* mutants overexpressing *pyc* via MDH and ME. It is currently unclear what exactly triggers this behaviour and for example labelling experiments might be used to further investigate fluxes in these strains. In analogy to *C. thermocellum* for ethanol production, this kind of shuttling behaviour in *B. smithii* has implications for engineering the strain for succinate production, which will be discussed later in this Chapter.

Besides the transcriptomics and metabolomics analyses described in Chapter 6, a proteomics analysis would be a valuable addition to these datasets to reveal post-translational modification, as well as ^{13}C-labelling experiments to determine fluxes. The integration of different '-omics' data (genomics, transcriptomics, metabolomics, proteomics) and combining these datasets with metabolic modelling can provide very valuable insights and has been used as a powerful tool to determine engineering strategies (Patnaik, 2008; Toya & Shimizu, 2013). The case of *C. thermocellum* (Zhou et al., 2013a) illustrates that genome annotations and transcriptomics can be misleading and classical biochemical assays are also required. This is also emphasized by the data shown in Chapter 6, where enzyme assays indicated a reflux from malate to pyruvate via malic enzyme during overexpression of pyruvate carboxylase in *B. smithii.* Allosteric interactions and post-translational modifications cannot be observed via transcriptomics analysis but often have a large influence (Mehmeti et al., 2012). In *C. glutamicum*, overexpression of *pyc* increased flux to oxaloacetate, whereas overexpression of *ppc* did not have any effect. It was found that PPC was strongly allosterically inhibited by malate and aspartate while PYC was not. After engineering of PPC to decrease the allosteric inhibition, oxaloacetate pools were increase and subsequently the target product lysine increased by 37% (Chen et al., 2014).

Overall, although a start has been made with understanding the *B. smithii* metabolism, further research is needed to elucidate unknown pathways, to reveal regulation patterns and to determine fluxes such as the distribution between the oxidative pentose phosphate pathway (oPPP) and glycolysis. The accumulation of oPPP intermediates in *B. smithii* and

the high activity observed of these enzymes especially in the mutants (Chapter 6) suggests that this pathway plays an important role in this organism, but metabolic flux analysis and labelling experiments are required to further determine this. Pathway distribution through glycolysis and oPPP in *G. thermoglucosidans* was shown to be dependent on growth rate (Hussein et al., 2015), illustrating the ability of these organisms to adjust their metabolism according to circumstances. Similar to this, we found different product distributions under different aeration conditions.

Influence of process conditions on metabolism: aerobic and dual-phase production profiles of *B. smithii*

We were mainly interested in fermentations without any gas additions as the addition of oxygen in large-scale reactors is inefficient and costly, but also the sparging of such reactors with for example nitrogen to obtain fully anaerobic conditions is undesired as it inhibits SSF enzymes (Podkaminer et al., 2012). Generally, non-gassed fermentations are indicated as 'micro-aerobic' as some air is present in the headspace and at the very start of the fermentation also in the medium, but within maximally a few hours no measurable dissolved oxygen (dO_2) remains in the medium. As the *ldhL* mutants grew very poorly under such micro-aerobic conditions (Chapter 5 and 6), we also tested them under aerated conditions as well as in dual-phase fermentations, in which reactors are initially sparged with air or oxygen to create biomass, after which the gas supply is switched off to start fermentation. Such a dual-phase strategy is a proven method for many organisms for which anaerobic (or micro-aerobic) growth is too challenging to immediately start producing maximally under these conditions, or to produce products requiring pathways that need to be activated by oxygen after which they can also run anaerobically, such as the glyoxylate shunt in *E. coli* for succinate production (Vemuri et al., 2002).

Acetate

Growing the *ldhL-sigF* mutant in a reactor with a dO_2 of 1% restored growth to OD_{600} values that were even above those obtained by the wild-type strain under micro-aerobic conditions (data not shown). Also, the production profile became almost homoacetogenic. When the wild-type was grown under the same aerated conditions, it showed the same homoacetogenic profile as the mutant strain and OD_{600} values were about 3-fold higher than under micro-aerobic conditions. However, whereas the mutant strain converted all glucose to acetate, the wild-type was not able to do so and produced less. The same effect was observed in tubes, in which the effect of the amount of oxygen was further evaluated by growing the ET 138 *ldhL-sigF* mutant strain as well as wild-type *B. smithii* ET 138 and *B. coagulans* DSM 1^T in 50 mL tubes containing 10, 20, 30, 40 or 50 mL medium to create

different aeration conditions (Figure 2). Clearly, *B. smithii* shows an almost linear relation between oxygen presence and acetate/lactate ratio, while this is not the case for *B. coagulans*. *B. coagulans* does show a difference between fully aerobic (10 mL) and fully micro-aerobic (50 mL) conditions, but the values in between are all very similar. The standard route to acetate production via *pta* and *ack* is missing in *B. smithii*, but despite efforts to unravel the acetate production pathway we have no solid indications yet of an alternative pathway.

Succinate

Besides the relation between oxygen concentration and acetate production, Figure 2 furthermore shows that the amounts of succinate are comparable for all volumes for the two wild-type strains, whereas in the mutant strain there is a clear increase in succinate production under the most aerobic conditions. This would indicate that succinate is produced via the oxidative TCA-cycle, which is common for thermophilic bacilli (Hussein et al., 2015). On the other hand, in dual-phase fermentations using the double mutant containing either pNW33n or pNW33n with *pyc-mdh* as described in Chapter 6, production of malate was observed during the aerated phase, whereas succinate production started only after oxygen was depleted (data not shown). The apparent discrepancy between the tube and reactor experiments might be caused by the amount of oxygen present or by the pH of the cultures, which is not constant in tubes. Further research via enzyme assays, proteomics

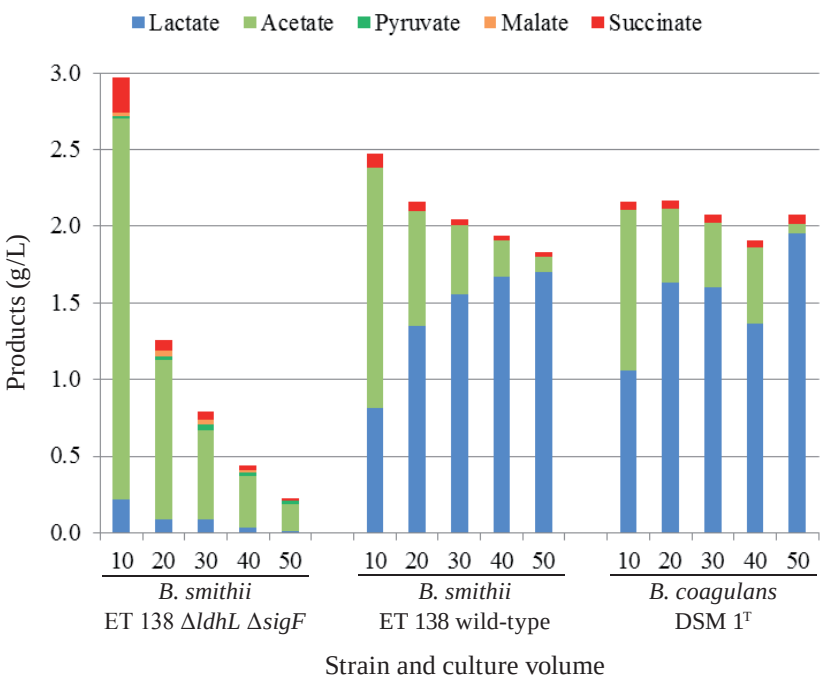

Figure 2. HPLC results of aerobic cultivation in tubes of *B. smithii* ET 138 and *B. coagulans* DSM 1^T. Strains were grown in 50 mL Greiner tubes in the indicated volumes.

and labelling experiments is required to determine the exact route of succinate production in *B. smithii* and the exact co-factors that are used by the involved enzymes. In order to do this, a metabolic model would be highly helpful. Nevertheless, some directions for optimization of succinate production can already be obtained from the genome (Chapter 4) and the data presented in Chapter 6, which will be discussed in the next section of this chapter.

Pyruvate

Triple mutant strain *B. smithii* ET 138 Δ*ldhL*-Δ*sigF*-Δ*pdhA* cannot produce acetate anymore and when this strain was grown aerobically, it produced high amounts of pyruvate. Pyruvate has applications as a green building block chemical (Xu et al., 2008) and is currently produced either via chemical synthesis or via fermentation by several yeast strains or *E. coli*. Also an engineered *C. glutamicum* strain lacking *pdhA* and *ldhL* has been shown to produce large amounts of pyruvate (Wieschalka et al., 2012). By optimizing the process conditions and using a triple-phase fermentation, this strain produced 500 mM pyruvate in a fed-batch fermentation with a maximum yield of 0.97 mol/mol glucose (Wieschalka et al., 2012). In the first phase, acetate and aeration were provided to allow fast growth. The second phase started after acetate was depleted and aeration was stopped, allowing the strain to consume the remaining oxygen. In the third, anaerobic phase, pyruvate production was observed. We observed similar rapid growth of our triple mutant strain during the first phase with a dO_2 of 20%, as well as strongly increased pyruvate accumulation after the depletion of acetate. However, as soon as oxygen was switched off and depleted, pyruvate production stopped and only some D-lactate was produced, indicating oxygen is required for pyruvate production by our strain. During the first aerobic 24 hours, 56 mM of pyruvate was produced together with 2 mM lactate and OD_{600} of 1.5. Process conditions were subsequently optimized to a dO_2 of 20% for 5 hours, after which this was set to 10% for the remainder of the fermentation. Again, growth and acetate consumption are coupled (Figure 3) and pyruvate production is increased after the depletion of acetate. Succinate and malate could not be measured due to technical issues. The source of the minor amounts (± 0.8 mM) of acetate produced at the end of the fermentation is unclear. After 47 h, 102 mM (8.98 g/L) pyruvate was produced with a total volumetric productivity of 2.15 mol/L/h (0.19 g/L/h) and a maximum of 5.67 mol/L/h (0.50 g/L/h), yielding a total of 1.04 mol pyruvate per mol glucose (0.51 g/g). Compared to *C. glutamicum* described above, this is a slightly higher yield. Even higher productivities, titres and yields were obtained using *S. cerevisiae* and *E. coli* (van Maris et al., 2004; Zelić et al., 2004; Zhu et al., 2008) and further process improvements might also improve results with the *B. smithii* triple mutant strain. Interestingly, the triple mutant strain could no longer be transformed, neither with pNW33n nor with pNW33n containing the deleted *pdhA* subunit, eliminating the possibility for further engineering steps to increase production.

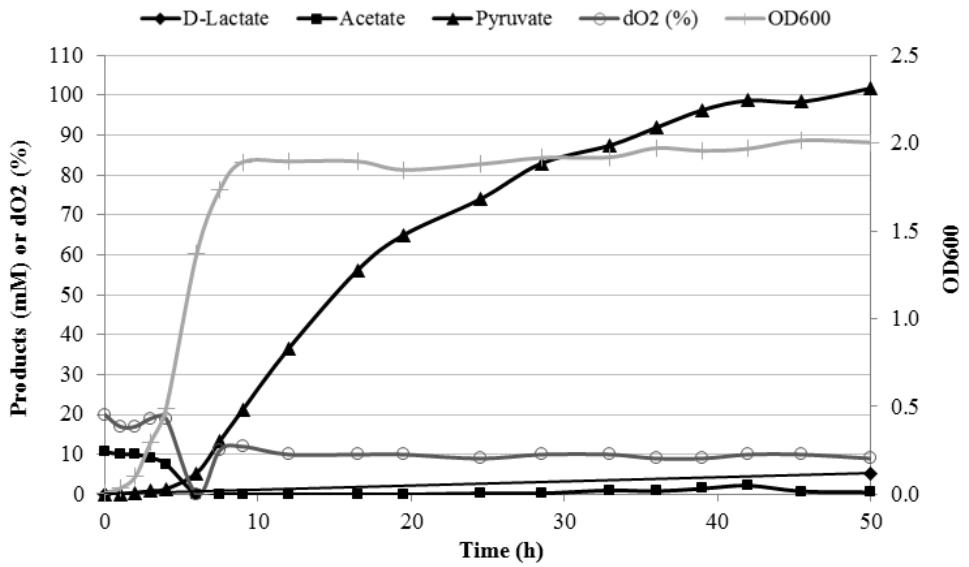

Figure 3. Aerated fermentation of *B. smithii* ET 138 Δ*ldhL*-Δ*sigF*-Δ*pdhA*. Fermentation was performed in a 2 L vessel with 1 L working volume containing TVMY medium with 30 g/L glucose and 10 mM acetate; pH was controlled at 6.5 and temperature at 55°C. Initially, oxygen levels were set at 20%, but during exponential growth all oxygen was consumed, after which levels were reset to 10%.

Further engineering steps towards succinate production in B. smithii

In Chapter 6, we attempted to direct metabolism towards the production of C_4-dicarboxylic acids malate and particularly succinate. Succinate is a highly interesting green building block chemical with a wide range of applications and a very large market potential (Bozell & Petersen, 2010; Werpy et al., 2004). Its production has been achieved near or at maximum theoretical yields with natural producers (Bretz & Kabasci, 2012; Jiang et al., 2012; Lee et al., 2002; Li et al., 2010), as well as with engineered model organisms (Agren et al., 2013; Tsuji et al., 2013; Vemuri et al., 2002; Zhang et al., 2009) or engineered natural producers (Becker et al., 2013; Lee et al., 2006). Only one thermophile has been described for succinate production, namely *C. thermosuccinogenes* (Drent et al., 1991). However, only few reports are available on this organism from the early 90's and production has never achieved very high yields. It would thus be interesting to achieve high succinate yields with a thermophilic organism to reduce production costs.

<u>Heterologous expression of a reductive TCA-cycle and modification of the PEP-pyr-OAA node</u>

Based on the *B. smithii* genome, the data in Chapter 6 and the data described above under aerobic conditions, it seems likely that succinate is mainly produced in an aerobic way via the respiratory oxidative TCA-cycle (Figure 1). The *B. smithii* genome does not

encode a fumarate reductase (*frd*), and succinate dehydrogenase (*sdh*) is not well capable of performing the reverse direction from fumarate to succinate. This makes heterologous expression of a gene cluster encoding *frd* a logical first engineering step for the production of succinate via the reductive TCA-cycle. As with all heterologous expressions in thermophiles, the source of the new enzymes should preferably also be thermophilic to ensure proper folding and activity of the enzymes at thermophilic temperatures. As mentioned, only one thermophilic succinate producer is known and this might be a good source of the *frd*. Besides *frd*, also fumarate hydratase *(fum*/FH) expression might be necessary as also no high concentrations of fumarate are detected in the medium. Malate dehydrogenase could also be heterologously expressed, but instead this could also be achieved with the native *mdh* in combination with removal of pathways creating a re-flux to pyruvate as shown in Chapter 6. This would include removal of malic enzyme as has been shown to be effective in *C. thermocellum* (Deng et al., 2013). Furthermore, whereas *pck* was also removed in *C. thermocellum* to increase TCA-cycle flux (Deng et al., 2013), in our case the direction of this enzyme could be reversed as has been shown in *E. coli* via point mutations in the promoter and the gene itself (Zhang et al., 2009). In this way, PCK contributes to both ATP-production and flux towards oxaloacetate (Figure 1). However, as PCK is the only gluconeogenic route after removal of malic enzyme, it is important that it is still able to function in the gluconeogenic direction and therefore another option is to also use heterologous expression of a *pck* known to function glycolytically, such as that from *C. thermocellum* (Zhou et al., 2013a), next to the native *pck*. Furthermore, as it is possible that the cells use malic enzyme as a transhydrogenase-like cycle to maintain redox balance, it is important to supply the cells with an alternative to do this. An interesting option would be the cofactor engineering of *gapdh* as mentioned in Chapter 6. By supplying the cells with a glycolytic *gapA* that is cofactor-engineered to use $NADP^+$ instead of NAD^+ next to the native NAD^+-dependent *gapA*, they can use these two enzymes to regulate NADH/NADPH ratios.

Making use of the native glyoxylate shunt to obtain maximum succinate yield

Apart from creating a reductive TCA-cycle, an interesting addition to this would be to make use of the native glyoxylate shunt (Figure 1). Activation of the glyoxylate shunt is a requirement to maintain redox balance and produce succinate at the maximum theoretical yield of 1.7 mol/mol glucose, as this is only possible when using both glyoxylate shunt and reductive TCA-cycle together (Sánchez et al., 2005) (Figure 1). Conversion of 1 mol glucose to 2 mol PEP via glycolysis yields 2 NADH. If only the reductive TCA-cycle is used, 2 NADH are required for 1 mol succinate, resulting in a maximum yield of 1 mol/mol glucose. Via the glyoxylate shunt, 1 mol succinate and 1 mol malate are formed, after which the malate can be converted to succinate by using only 1 NADH. If the reductive pathway, which uses 2 mol NADH, is used for 5/7 per mol glucose and the glyoxylate shunt, which uses 1 NADH,

for 2/7, the theoretical maximum yield of 1.7 mol/mol can be obtained (5/7 * 2 NADH + 2/7 * 1 NADH = 1.7) (Figure 1).

In *E. coli*, the glyoxylate shunt is typically active under aerobic conditions and repressed under anaerobic conditions by the repressor *iclR* (Lin et al., 2005; Sánchez et al., 2005). Remarkably, as shown in Chapter 6, the glyoxylate shunt is highly expressed in the *B. smithii* ET 138 *ldhL* and *ldhL-sigF* mutants. A homolog to the known *E. coli iclR* could not be identified, but 5 genes have been annotated as *iclR*-family repressors and those are all strongly downregulated in the *ldhL* and *ldhL-sigF* mutants. The glyoxylate shunt has also been shown to be activated in *E. coli* after removal of *pta* and *ack* in an *ldh* and *adhE* deficient mutant strain (Zhu et al., 2013). As all these genes are not present in *B. smithii*, this might be an alternative explanation for the activation of the glyoxylate shunt without removal of the repressor.

Requirements for an improved genetic system

Taken together, it will require a vast amount of engineering steps to produce succinate with *B. smithii*. Although the tools described in Chapter 5 enable clean gene deletions and work efficiently for some genes, we were unable to use to the same system for several other genes. This indicates that either these genes are essential, or our tools are not efficient enough. Based on homology with related organisms, the first option seems unlikely and therefore future research should focus on developing more high-throughput tools to further develop *B. smithii* into a platform organism. The main disadvantage of the *lacZ*-counter selection system is that it can still result in wild-type revertants. A high-throughput engineering system should enable only mutants to survive the screening and selection procedure. Preferably, this should be performed in a markerless way, but in order to screen for the effect of mutations, also an initial non-markerless system could be used. Several examples of counter-selection systems successfully applied in thermophiles have been discusses in Chapter 2. Of the systems depicted in Figure 4 (Chapter 2's Figure 2), only the one in 4B, C and D ensure that only mutants are obtained. The Cre-*lox* system in Figure 4D is not completely markerless and requires the availability of two different plasmids, which is not always the case (for example, *B. smithii* could so far only be transformed with pNW33n). The other two systems are markerless, but for the *pta-ack* system in Figure 4B the possibility to transform with linear DNA is required as otherwise the system could also insert via single crossover without pressure to perform the second crossover. Only the *tdk-hpt* system in Figure 4C can be used with a single plasmid, is markerless and only results in mutant colonies and prevents growth of wild-type revertants. The reason for this is the two-step or 'triple recombination' procedure. The initial double recombination can easily be screened for because there is a positive selection marker in between the homologous flanks as well as outside the flanks on the plasmid. The former one ensures selection for the double crossover, whereas the latter

one ensures plasmid curing, thereby preventing revertants and ensuring stable integration. The third crossover is initially prevented because antibiotics are added so the cells cannot survive without keeping the *abR-hpt* cassette. Then, the third crossover is initiated by removing the antibiotic pressure and actively selected for by adding AZH, which is toxic as long as the *abR-hpt* cassette is inserted. As only one homologous region is still present, the only outcome of this third recombination is the mutant genotype. The *pyrF*-system in Figure 4A could also be changed into a more solid triple recombination system by adding the *pyrF* and *abR* genes in between the flanks and adding for example the *lacZ* gene as we used in Chapter 5 outside the flanks on the plasmid backbone. By using one antibiotic marker and two markers that can be selected against in this way, several genes could be used to set up a triple recombination system like this. The advantages are that only one plasmid and one transformation are required. The disadvantage is that the described positive markers should be available, which often means that knockouts of the marker genes need to be constructed prior to using the system. This is usually rather easy, however, as pressure against these genes can easily be applied when making these initial knockouts. Also, efficient homologous recombination is a prerequisite for using such a system. In *B. smithii*, homologous recombination seems to occur very easily as single and double crossovers where readily obtained even when transforming with a replicating plasmid (Chapter 5). In most thermophiles described in literature, thermosensitive plasmids are used and it is unknown whether integration also occurs when the temperature is not increased. It would be an interesting line of future research to evaluate this in different thermophilic bacilli.

Besides using such a 'triple recombination system', another feasible and recently emerged tool is CRISPR-mediated engineering. Such a system makes use of the exonuclease activity of a heterologously expressed bacterial CRISPR-system that is targeted to cut a very specific sequence in the target genome. In many eukaryotes, Cas9 from the *S. pyogenes* Type II CRISPR-system is now a very common engineering tool. After Cas9 has made a double-strand break in the target sequence of the genome, the eukaryotic non-homologous end-joining (NHEJ) system repairs the double strand break but does so in an error-prone way, creating defective genes and mutant genotypes. Alternatively, homologous flanks to the target site can be provided in order to repair the break in a specific way and make very controlled mutations. Although this CRISPR-Cas9-system is of bacterial origin, it has mainly been used in eukaryotes and hardly in bacteria. The main reason for this is that bacteria mostly do not have NHEJ, requiring the need for homologous recombination, which is also not always very effective and often requires the co-expression of a phage-derived recombineering system (Oh & van Pijkeren, 2014; Selle & Barrangou, 2015). As *B. smithii* seems to recombine very efficiently, it will be interesting to test the Cas9 engineering system in combination with homologous recombination as an efficient engineering tool. As long as Cas9 is expressed in the cells, it will be impossible for wild-type cells to survive: only

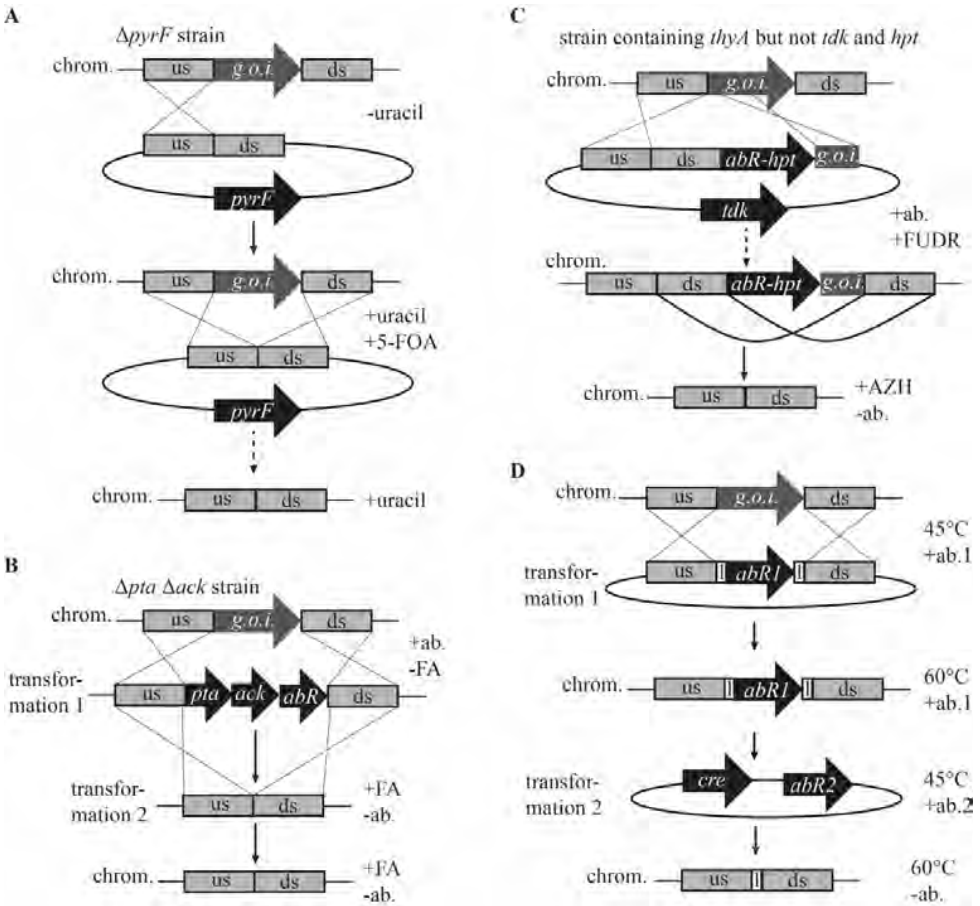

Figure 4 (same as Figure 2 in Chapter 2). Overview of counter-selection and antibiotic-cassette removal systems used in thermophiles for targeted double crossovers. Abbreviations: chrom.: chromosome; us: upstream flanking region; ds: downstream flanking region; g.o.i.: gene of interest; ab.: antibiotic; abR: antibiotic-resistance gene. **A:** *pyrF*-system. After insertion of the *pyrF*-containing plasmid into a Δ*pyrF* strain, single crossovers can be obtained by selecting for restoration of the uracil auxotrophy. Double crossovers are obtained by adding 5-fluoroorotic acid (5-FOA), which will be toxic for strains still harbouring the *pyrF*-plasmid, and by adding uracil to select for strains being uracil auxotrophs due to lack of *pyrF*. Based on (Kita et al., 2012; Suzuki et al., 2012) **B:** *pta-ack*-system. In this case, linear DNA fragments were used for transformation, which allows simple and direct screening for double crossovers, but the system can also be used with plasmid DNA. First, the *pta* and *ack* genes for acetate formation need to be deleted, resulting in a strain resistant against haloacetate compounds such as fluoroacetate (FA). This strain is then transformed with DNA containing *pta*, *ack* and an antibiotic-resistance gene, which replace the gene of interest via the flanking regions. Successful double crossover and thus insertion of the three genes is detected by antibiotic resistance. In a second transformation, the strain is transformed with fused flanking regions of the gene of interest and cultured on FA, thereby deleting the *pta*, *ack* and an antibiotic-resistance gene since FA is toxic when *pta* and *ack* are present. Based on (Shaw et al., 2011). **C:** *tdk-hpt* system. After selecting transformants based on antibiotic-resistance, double crossovers can be detected by growing the culture on antibiotics and fluoro-deoxyuracil (FUDR), to which the strains will be resistant due to loss of *tdk* (which together with the native *thyA* makes the strain sensitive to FUDR) after they have recombined and cured the plasmid. Subsequently, clean knockouts created by a second crossover event can be detected by culturing on 8-azahypoxannthine (AZH), to which strains will be resistant when they have recombined and lost the *hpt* and antibiotic-resistance gene. Based on (Argyros et al., 2011). **D:** Cre-*lox* system. Using a thermosensitive plasmid, in a first step the gene of interest is replaced by an antibiotic resistance gene flanked by lox sites, enabling selection for antibiotic resistance. In a second transformation, a second thermosensitive plasmid coding for the Cre-recombinase and conveying a different antibiotic resistance is inserted. The Cre-recombinase recombines the *lox66* and *lox71* sites to a *lox72* site, thereby cleaving out the first antibiotic-resistance gene. Based on (Kovacs et al., 2010).

mutants will survive that recombined with the homologous flanks and thereby deleted the wild-type sequence targeted by Cas9. However, as the currently known Cas9-engieerning system is derived from and only used in mesophiles, it needs to be determined whether the system will also be active under thermophilic conditions.

Besides the inability to revert to wild-type genotypes, a high-throughput genetic toolbox also requires that gene expression can be finely tuned. During the work described in this thesis, we have tested a total of three promoters. Preliminary tests with these promoters cloned to drive *lacZ* gene expression showed different expression under different conditions (data not shown). To date, no inducible promoters have been developed for *B. smithii* and existing constitutive promoters need to be characterized in more detail. In general, studies evaluating different promoters for thermophiles (both inducible and constitutive) are scarce and gaining more attention only recently, such as in *G. thermoglucosidans* (Bartosiak-Jentys et al., 2012) and *C. thermocellum* (Mearls et al., 2015). Especially when tools such as Cas9 are used, it is important that expression of such nucleases can be induced and switched off in order to prevent toxicity and off-target effects. Also, in order to tune metabolic fluxes, changing promoter activity is a feasible option. For this, also exchange of native promoters via integration into the genome could be used. To identify strong and weak promoters under different conditions, the transcriptome data from Chapter 6 should be further expanded by transcriptomics analysis on for example different substrates. Further research should also determine for example transcription start sites and ribosome binding site sequences.

B. SMITHII: A PROMISING NOVEL PLATFORM ORGANISM?

In this thesis, we set out to develop a novel thermophilic platform organism for the cost-efficient production of green chemicals. It is unlikely that the completely ideal organism as described in Chapter 1 exists or can be isolated, but we can compare our novel organism to other organisms. *B. smithii* has all the advantages of a thermophile and is capable of utilizing a wide range of carbon sources, giving it a major advantage over many currently used organisms. Although we always added 0.5 g/L yeast extract and vitamins, this is very little compared to many currently used organisms (Chapter 2) and as the organism was isolated on a minimal medium, medium optimization can likely reduce the nutritional requirements even further. Importantly, *B. smithii* shows reproducible growth under non-aerated conditions in this low-yeast extract medium, which is not always the case for for example *Geobacilli* (Hussein et al., 2015). The strain grows fast in a wide range of temperatures and pHs, making it robust against contamination not only with mesophiles but also with related organisms, and against fluctuations in process conditions which might occur on larger scale. Its transformation and homologous recombination are also reliable and reproducible, something that we found is often not the case in thermophilic bacilli (Chapter 3 and 7). Robustness and reproducibility are important features for a production strain. The fact that

the *ldhL* mutants hardly produce any products on one hand limits the possibilities of making other products with *B. smithii* as the genes need to be heterologously inserted, although it might be an efficient producer of pyruvate and acetate without metabolic engineering but via process optimization as was shown in this chapter. On the other hand, the *B. smithii* Δ*ldhL* strain might be an ideal platform organism as only one gene needs to be deleted to eliminate production and other genes to new products can be inserted, as was shown for D-lactate production in Chapter 6. Especially for for example succinate production, many engineering steps are required. This holds true for basically every platform organism and a very large number of mutations (insertions and deletions) are no exception for *E. coli* or *S. cerevisiae* production strains. In order to achieve this in *B. smithii*, however, the current genetic tools are not sufficient. *B. smithii* can only become a real platform organism when more efficient, high-throughput tools are developed and its metabolism is better understood. Five other thermophiles that are currently being promising potential platform organisms for green chemical or fuel production are *B. coagulans*, *B. licheniformis*, *G. thermoglucosidans*, *C. thermocellum* and *T. saccharolyticum*. The latter two have the disadvantage of being strict anaerobes, while they have the advantage of being able to degrade raw biomass (though not all, as *C. thermocellum* for example is only cellulolytic). Nevertheless, the conversion of biomass in a consolidated bioprocessing (CBP) process has so far been mainly shown with model substrates and has not been fully industrially implemented (Olson et al., 2012), making SSF thus far the more feasible strategy, for which a facultative anaerobe is preferred. Genetic tools seem to be in similar stages for all four organisms, with some more promoters being described for *Geobacillus* and *Clostridium* and better integration and transformation tools for *Thermoanaerobacter*, especially since the latter one is naturally transformable (Shaw et al., 2010). For all these organisms, understanding of the metabolism seems to be in a similar stage, with several examples described in this chapter.

Altogether, engineering tools as well as understanding of their metabolism seems to be in similar stages for all these thermophilic organisms and all of them have different metabolic capacities such as differences in carbon utilization, growth characteristics and end-product pathways. Therefore, it is hard to predict which organism in the end will be the best platform organism as different bottlenecks might be found for the different organisms when they are being further developed. *B. smithii* is certainly among the top candidates for the reasons described here and worth further research. For all the thermophilic organisms there is still a large gap in understanding of physiology and availability of genetic tools compared to *E. coli* or *S. cerevisiae*. Nevertheless, the start has been made and these organisms each have advantages that will make them the more preferred organisms in the long run once they are better understood and 'domesticated'. With the emerging approaches of high-throughput engineering and automation of strain development this might go faster than before.

References

Abdel-Rahman, M.A., Tashiro, Y., Sonomoto, K. 2011. Lactic acid production from lignocellulose-derived sugars using lactic acid bacteria: Overview and limits. *Journal of Biotechnology*, 156(4), 286-301.

Abriouel, H., Franz, C.M., Ben Omar, N., Galvez, A. 2011. Diversity and applications of *Bacillus* bacteriocins. *FEMS Microbiology Reviews*, 35(1), 201-32.

Agbor, V.B., Cicek, N., Sparling, R., Berlin, A., Levin, D.B. 2011. Biomass pretreatment: Fundamentals toward application. *Biotechnology Advances*, 29(6), 675-685.

Agren, R., Otero, J., Nielsen, J. 2013. Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production. *Journal of Industrial Microbiology & Biotechnology*, 1-13.

Angelov, A., Li, H., Geissler, A., Leis, B., Liebl, W. 2013. Toxicity of indoxyl derivative accumulation in bacteria and its use as a new counterselection principle. *Systematic and Applied Microbiology*, 36(8), 585-592.

Argyros, D.A., Tripathi, S.A., Barrett, T.F., Rogers, S.R., Feinberg, L.F., Olson, D.G., . . . Caiazza, N.C. 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Applied and Environmental Microbiology*, 77(23), 8288-8294.

Atkinson, A., Cripps, R., Eley, K., Rudd, B., Todd, M. 2008. Thermophilic micro-organisms for ethanol production. PCT WO2008/038019

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., . . . Kubal, M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9(1), 75.

Balk, M., Heilig, H.H.J., Eekert, M.A., Stams, A.M., Rijpstra, I., Sinninghe-Damsté, J., . . . Kengen, S.M. 2009. Isolation and characterization of a new CO-utilizing strain, *Thermoanaerobacter thermohydrosulfuricus* subsp. *carboxydovorans*, isolated from a geothermal spring in Turkey. *Extremophiles*, 13(6), 885-894.

Barakat, A., de Vries, H., Rouau, X. 2013. Dry fractionation process as an important step in current and future lignocellulose biorefineries: A review. *Bioresource Technology*, 134(0), 362-373.

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), 5945-7.

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

Becker, J., Reinefeld, J., Stellmacher, R., Schäfer, R., Lange, A., Meyer, H., . . . Wittmann, C. 2013. Systems-wide analysis and engineering of metabolic pathway fluxes in bio-succinate producing *Basfia succiniciproducens*. *Biotechnology and Bioengineering*, 110(11), 3013-3023.

Bhalla, A., Bansal, N., Kumar, S., Bischoff, K.M., Sani, R.K. 2013. Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresource Technology*, 128(0), 751-759.

Bhandiwad, A., Shaw, A.J., Guss, A., Guseva, A., Bahl, H., Lynd, L.R. 2014. Metabolic engineering of *Thermoanaerobacterium saccharolyticum* for n-butanol production. *Metabolic Engineering*, 21, 17-25.

Bielen, A.A.M., Verhaart, M.R.A., VanFossen, A.L., Blumer-Schuette, S.E., Stams, A.J.M., van der Oost, J., . . . Kengen, S.W.M. 2013. A thermophile under pressure: Transcriptional analysis of the response of *Caldicellulosiruptor saccharolyticus* to different H2 partial pressures. *International Journal of Hydrogen Energy*, 38(4), 1837-1849.

Bischoff, K., Liu, S., Hughes, S., Rich, J. 2010. Fermentation of corn fiber hydrolysate to lactic acid by the moderate thermophile *Bacillus coagulans*. *Biotechnology Letters*, 32(6), 823-828.

Biswas, I., Gruss, A., Ehrlich, S.D., Maguin, E. 1993. High-efficiency gene inactivation and replacement system for gram-positive bacteria. *Journal of Bacteriology*, 175(11), 3628-3635.

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), 6715-6723.

Blomqvist, K., Nikkola, M., Lehtovaara, P., Suihko, M.L., Airaksinen, U., Stråby, K.B., . . . Penttilä, M.E. 1993. Characterization of the genes of the 2,3-butanediol operons from *Klebsiella terrigena* and *Enterobacter aerogenes*. *Journal of Bacteriology*, 175(5), 1392-1404.

Blumer-Schuette, S.E., Brown, S.D., Sander, K.B., Bayer, E.A., Kataeva, I., Zurawski, J.V., . . . Kelly, R.M. 2013. Thermophilic lignocellulose deconstruction. *FEMS Microbiology Reviews*, 38(3), 393-448.

Blumer-Schuette, S.E., Ozdemir, I., Mistry, D., Lucas, S., Lapidus, A., Cheng, J.-F., . . . Kelly, R.M. 2011. Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*. *Journal of Bacteriology*, 193(6), 1483-1484.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., Pirovano, W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, 27(4), 578-579.

Boetzer, M., Pirovano, W. 2012. Toward almost closed genomes with GapFiller. *Genome biology*, 13(6), R56.

Bommareddy, R.R., Chen, Z., Rappert, S., Zeng, A.P. 2014. A de novo NADPH generation pathway for improving lysine production of *Corynebacterium glutamicum* by rational design of the coenzyme specificity of glyceraldehyde 3-phosphate dehydrogenase. *Metab Eng*, 25, 30-7.

Bosma, E.F., van de Weijer, A.H.P., Daas, M.J.A., van der Oost, J., de Vos, W.M., van Kranenburg, R. 2015a. 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), 1874-1883.

Bosma, E.F., van de Weijer, A.H.P., van der Vlist, L., de Vos, W.M., van der Oost, J., van Kranenburg, R. 2015b. Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains. *Microbial Cell Factories*, 14(1), 99.

Bosma, E.F., van der Oost, J., de Vos, W.M., van Kranenburg, R. 2013. Sustainable production of bio-based chemicals by extremophiles. *Current Biotechnology*, 2(4), 360-379.

Boyle, N.R., Reynolds, T.S., Evans, R., Lynch, M., Gill, R.T. 2013. Recombineering to homogeneity: extension of multiplex recombineering to large-scale genome editing. *Biotechnology Journal*, 8(5), 515-522.

Bozell, J.J., Petersen, G.R. 2010. Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's "Top 10" revisited. *Green Chemistry*, 12(4), 539-554.

Bretz, K., Kabasci, S. 2012. Feed-control development for succinic acid production with *Anaerobiospirillum succiniciproducens*. *Biotechnology and Bioengineering*, 109(5), 1187-1192.

Brumm, P., De Maayer, P., Cowan, D.A., MEAD, D.A. 2015. Genomic analysis of six new *Geobacillus* strains reveals highly conserved carbohydrate degradation architectures and strategies. *Frontiers in Microbiology*, 6(May), art. nr. 430.

Budhavaram, N.K., Fan, Z. 2009. Production of lactic acid from paper sludge using acid-tolerant, thermophilic *Bacillus coagulans* strains. *Bioresource Technology*, 100(23), 5966-5972.

Buschke, N., Schäfer, R., Becker, J., Wittmann, C. 2012. Metabolic engineering of industrial platform microorganisms for biorefinery applications - optimization of substrate spectrum and process robustness by rational and evolutive strategies. *Bioresource Technology*, 135, 544-554.

Cai, Y., Lai, C., Li, S., Liang, Z., Zhu, M., Liang, S., Wang, J. 2011. Disruption of lactate dehydrogenase through homologous recombination to improve bioethanol production in *Thermoanaerobacterium aotearoense*. *Enzyme and Microbial Technology*, 48(2), 155-161.

Cecchini, G., Schroder, I., Gunsalus, R.P., Maklashina, E. 2002. Succinate dehydrogenase and fumarate reductase from Escherichia coli. *Biochimica et Biophysica Acta*, 1553(1-2), 140-57.

Celińska, E., Grajek, W. 2009. Biotechnological production of 2,3-butanediol—Current state and prospects. *Biotechnology Advances*, 27(6), 715-725.

Chaisson, M., Tesler, G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*, 13(1), 238.

Chang, T., Yao, S. 2011. Thermophilic, lignocellulolytic bacteria for ethanol production: current state and perspectives. *Applied Microbiology and Biotechnology*, 92(1), 13-27.

Chang, V., Nagwani, M., Holtzapple, M. 1998. Lime pretreatment of crop residues bagasse and wheat straw. *Applied Biochemistry and Biotechnology*, 74(3), 135-159.

Chapman, A.G., Fall, L., Atkinson, D.E. 1971. Adenylate energy charge in *Escherichia coli* during growth and starvation. *Journal of Bacteriology*, 108(3), 1072-1086.

Chen, G.Q., Patel, M.K. 2012. Plastics derived from biological sources: Present and future: A technical and environmental review. *Chemical Reviews*, 112(4), 2082-2099.

Chen, X., Zhou, L., Kangming, T., Kumar, A., Singh, S., Prior, B.A., Wang, Z. 2013. Metabolic engineering of *Escherichia coli:* A sustainable industrial platform for bio-based chemical production. *Biotechnology Advances*, 31(8), 1200-1223.

Chen, Z., Bommareddy, R.R., Frank, D., Rappert, S., Zeng, A.-P. 2014. Deregulation of feedback inhibition of phosphoenolpyruvate carboxylase for improved lysine production in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology*, 80(4), 1388-1393.

Chung, D., Cha, M., Guss, A.M., Westpheling, J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proceedings of the National Academy of Sciences*, 111(24), 8931-8936.

Chung, D., Farkas, J., Huddleston, J.R., Olivar, E., Westpheling, J. 2012. Methylation by a unique α-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. *PLoS ONE*, 7(8), e43844.

Cohen, S.N., Chang, A.C.Y. 1973. Recircularization and autonomous replication of a sheared R-Factor DNA segment in *Escherichia coli* transformants. *Proceedings of the National Academy of Sciences*, 70(5), 1293-1297.

Cohn, F. 1872. Untersuchungen tiber Bakterien. *Beitrage zur Biologie der Pflanzen Heft 2*, 1, 127-224.

Coorevits, A., Dinsdale, A.E., Halket, G., Lebbe, L., de Vos, P., Van Landschoot, A., Logan, N.A. 2012. Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus, G. jurassicus, G. toebii, G. thermodenitrificans and G. thermoglucosidans* (nom. corrig., formerly '*thermoglucosidasius*'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus*; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 62(7), 1470-1485.

Cotter, P.D., Hill, C. 2003. Surviving the acid test: Responses of gram-positive bacteria to low pH. *Microbiology and Molecular Biology Reviews*, 67(3), 429-453.

Cripps, R.E., Eley, K., Leak, D.J., Rudd, B., Taylor, M., Todd, M., . . . Atkinson, T. 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering*, 11(6), 398-408.

Datta, R., Henry, M. 2006. Lactic acid: recent advances in products, processes and technologies — a review. *Journal of Chemical Technology and Biotechnology*, 81(7), 1119-1129.

De Maayer, P., Brumm, P.J., Mead, D.A., Cowan, D.A. 2014. Comparative analysis of the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC Genomics*, 15(1), 836.

Deloitte. 2014. Opportunities for the fermentation-based chemical industry - An analysis of the market potential and competitiveness of North-West Europe. *Deloitte Netherlands report*.

Demain, A.L., Newcomb, M., Wu, J.H.D. 2005. Cellulase, *Clostridia*, and ethanol. *Microbiology and Molecular Biology Reviews*, 69(1), 124-154.

Deng, Y., Olson, D.G., Zhou, J., Herring, C.D., Shaw, A.J., Lynd, L.R. 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. *Metabolic Engineering*, 15, 151-158.

Desai, S., Guerinot, M., Lynd, L. 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), 600-605.

Don, T.M., Chen, C.W., Chan, T.H. 2006. Preparation and characterization of poly(hydroxyalkanoate) from the fermentation of *Haloferax mediterranei*. *Journal of Biomaterials Science, Polymer Edition*, 17(12), 1425-1438.

Drent, W.J., Lahpor, G.A., Wiegant, W.M., Gottschal, J.C. 1991. Fermentation of inulin by *Clostridium thermosuccinogenes* sp. nov., a thermophilic anaerobic bacterium isolated from various habitats. *Applied and Environmental Microbiology*, 57(2), 455-462.

Elkins, J.G., Lochner, A., Hamilton-Brehm, S.D., Davenport, K.W., Podar, M., Brown, S.D., . . . Graham, D.E. 2010. Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47T. *Journal of Bacteriology*, 192(22), 6099-6100.

Euzéby, J. 2010. List of new names and new combinations previously effectively, but not validly, published. *International Journal of Systematic and Evolutionary Microbiology*, 60(3), 469-472.

Feinberg, L., Foden, J., Barrett, T., Davenport, K.W., Bruce, D., Detter, C., . . . Caiazza, N. 2011. Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313. *Journal of Bacteriology*, 193(11), 2906-2907.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 783-791.

Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., . . . Wang, L. 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), 5602-5607.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., . . . Wipat, A. 2008. The minimum information about a genome sequence (MIGS) specification. 26(5), 541-547.

Fillinger, S., Boschi-Muller, S., Azza, S.d., Dervyn, E., Branlant, G., Aymerich, S. 2000. Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. *Journal of Biological Chemistry*, 275(19), 14031-14037.

Fischer, A. 1894. *Untersuchungen über bakterien*. Borntraeger, Berlin.

FitzPatrick, M., Champagne, P., Cunningham, M.F., Whitney, R.A. 2010. A biorefinery processing perspective: Treatment of lignocellulosic materials for the production of value-added products. *Bioresource Technology*, 101(23), 8915-8922.

Fleming, A.B., Tangney, M., Jørgensen, P.L., Diderichsen, B., Priest, F.G. 1995. Extracellular enzyme synthesis in a sporulation-deficient strain of *Bacillus licheniformis*. *Applied and Environmental Microbiology*, 61(11), 3775-80.

Fong, J., Svenson, C., Nakasugi, K., Leong, C., Bowman, J., Chen, B., . . . Rogers, P. 2006. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles*, 10(5), 363-372.

Gao, C., Xue, Y., Ma, Y. 2011. Protoplast transformation of recalcitrant alkaliphilic *Bacillus* sp. with methylated plasmid DNA and a developed hard agar regeneration medium. *PLoS ONE*, 6(11), e28148.

Gao, W., Bao, Y., Liu, Y., Zhang, X., Wang, J., An, L. 2009. Characterization of thermo-stable endoinulinase from a new strain *Bacillus smithii* T7. *Applied Biochemistry and Biotechnology*, 157(3), 498-506.

Garrity, G.M., Holt, J.G. 2001. The road map to the manual. in: *Bergey's Manual® of Systematic Bacteriology*, Springer, pp. 119-166.

Ge, X.Y., Yuan, J., Qin, H., Zhang, W.G. 2011. Improvement of L-lactic acid production by osmotic-tolerant mutant of *Lactobacillus casei* at high temperature. *Applied Microbiology and Biotechnology*, 89(1), 73-78.

Gibbons, N.E., Murray, R.G.E. 1978. Proposals Concerning the Higher Taxa of Bacteria. *International Journal of Systematic Bacteriology*, 28(1), 1-6.

Goel, A., Santos, F., de Vos, W.M., Teusink, B., Molenaar, D. 2012. Standardized assay medium to measure *Lactococcus lactis* enzyme activities while mimicking intracellular conditions. *Applied and Environmental Microbiology*, 78(1), 134-143.

Gokarn, R.R., Eiteman, M.A., Altman, E. 2000. Metabolic analysis of *Escherichia coli* in the presence and absence of the carboxylating enzymes phosphoenolpyruvate carboxylase and pyruvate carboxylase. *Applied and Environmental Microbiology*, 66(5), 1844-1850.

Gordon, R.E., Smith, N.R. 1949. Aerobic sporeforming bacteria capable of growth at high temperatures. *Journal of Bacteriology*, 58(3), 327-41.

Guettler, M.V., Rumler, D., Jain, M.K. 1999. *Actinobacillus succinogenes* sp. nov., a novel succinic-acid-producing strain from the bovine rumen. *International Journal of Systematic Bacteriology*, 49(1), 207-216.

Gwinn, D.D., Thorne, C.B. 1964. Transformation of *Bacillus licheniformis*. *Journal of Bacteriology*, 87(3), 519-526.

Hamilton-Brehm, S.D., Vishnivetskaya, T.A., Allman, S.L., Mielenz, J.R., Elkins, J.G. 2012. Anaerobic high-throughput cultivation method for isolation of thermophiles using biomass-derived substrates. *Methods in Molecular Biology*, 908, 153-168.

Hammer, B. 1915. Bacteriological studies on the coagulation of evaporated milk. *Iowa Agricultural Experimental Station Research Bulletin*, 19, 119-131.

Han, L., Yang, G., Zhou, X., Yang, D., Hu, P., Lu, Q., Zhou, S. 2013. *Bacillus thermocopriae* sp. nov., isolated from a compost. *International Journal of Systematic and Evolutionary Microbiology*, 63(Pt 8), 3024-3029.

Hanahan, D., Jessee, J., Bloom, F.R., Jeffrey, H.M. 1991. 4: Plasmid transformation of *Escherichia coli* and other bacteria. in: *Methods in Enzymology*, Vol. Volume 204, Academic Press, pp. 63-113.

Hasunuma, T., Kondo, A. 2012. Development of yeast cell factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. *Biotechnology Advances*, 30(6), 1207-1218.

Heipieper, H.J., Keweloh, H., Rehm, H.J. 1991. Influence of phenols on growth and membrane permeability of free and immobilized *Escherichia coli*. *Applied and Environmental Microbiology*, 57(4), 1213-1217.

Hemme, C.L., Fields, M.W., He, Q., Deng, Y., Lin, L., Tu, Q., . . . Zhou, J. 2011. Correlation of genomic and physiological traits of *Thermoanaerobacter* species with biofuel yields. *Applied and Environmental Microbiology*, 77(22), 7998-8008.

Hoffmann, K., Wollherr, A., Larsen, M., Rachinger, M., Liesegang, H., Ehrenreich, A., Meinhardt, F. 2010. Facilitation of direct conditional knockout of essential genes in *Bacillus licheniformis* DSM13 by comparative genetic analysis and manipulation of genetic competence. *Applied and Environmental Microbiology*, 76(15), 5046-5057.

Hofvendahl, K., Hahn–Hägerdal, B. 2000. Factors affecting the fermentative lactic acid production from renewable resources. *Enzyme and Microbial Technology*, 26(2–4), 87-107.

Hols, P., Ramos, A., Hugenholtz, J., Delcour, J., De Vos, W.M., Santos, H., Kleerebezem, M. 1999. Acetate utilization in Lactococcus lactis deficient in lactate dehydrogenase: A rescue pathway for maintaining redox balance. *Journal of Bacteriology*, 181(17), 5521-5526.

Hong, K.-K., Nielsen, J. 2012. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cellular and Molecular Life Sciences*, 69(16), 2671-2690.

Hussein, A.H., Lisowska, B.K., Leak, D.J., Sima, S., Geoffrey Michael, G. 2015. Chapter One - The genus *Geobacillus* and their biotechnological potential. in: *Advances in Applied Microbiology*, Vol. Volume 92, Academic Press, pp. 1-48.

Ibrahim, M., Willems, A., Steinbüchel, A. 2010. Isolation and characterization of new poly(3HB)-accumulating star-shaped cell-aggregates-forming thermophilic bacteria. *Journal of Applied Microbiology*, 109(5), 1579-1590.

Ibrahim, M.H.A., Steinbüchel, A. 2010. High-cell-density cyclic fed-batch fermentation of a poly(3-Hydroxybutyrate)-accumulating thermophile, *Chelatococcus* sp. strain MW10. *Applied and Environmental Microbiology*, 76(23), 7890-7895.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., . . . Kyrpides, N. 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature*, 423(6935), 87-91.

Izquierdo, J.A., Sizova, M.V., Lynd, L.R. 2010. Diversity of bacteria and glycosyl hydrolase family 48 genes in cellulolytic consortia enriched from thermophilic biocompost. *Applied and Environmental Microbiology*, 76(11), 3545-3553.

Jang, Y.-S., Kim, B., Shin, J.H., Choi, Y.J., Choi, S., Song, C.W., . . . Lee, S.Y. 2012. Bio-based production of C2–C6 platform chemicals. *Biotechnology and Bioengineering*, 109(10), 2437-2459.

Ji, X.J., Huang, H., Zhu, J.G., Ren, L.J., Nie, Z.K., Du, J., Li, S. 2010. Engineering *Klebsiella oxytoca* for efficient 2, 3-butanediol production through insertional inactivation of acetaldehyde dehydrogenase gene. *Applied Microbiology and Biotechnology*, 85(6), 1751-1758.

Jiang, M., Xu, R., Xi, Y.-L., Zhang, J.-H., Dai, W.-Y., Wan, Y.-J., . . . Wei, P. 2012. Succinic acid production from cellobiose by *Actinobacillus succinogenes*. *Bioresource Technology*, 135, 469-474.

Jones, S.W., Tracy, B.P., Gaida, S.M., Papoutsakis, E.T. 2011. Inactivation of σF in *Clostridium acetobutylicum* ATCC 824 blocks sporulation prior to asymmetric division and abolishes σE and σG protein expression but does not block solvent formation. *Journal of Bacteriology*, 193(10), 2429-2440.

Jung, M.Y., Kim, J.-S., Chang, Y.-H. 2009. *Bacillus acidiproducens* sp. nov., vineyard soil isolates that produce lactic acid. *International Journal of Systematic and Evolutionary Microbiology*, 59(9), 2226-2231.

Jurado, M., López, M.J., Suárez-Estrella, F., Vargas-García, M.C., López-González, J.A., Moreno, J. 2014. Exploiting composting biodiversity: study of the persistent and biotechnologically relevant microorganisms from lignocellulose-based composting. *Bioresource Technology*, 162, Pages 283-293.

Kahru, A., Liiders, M., Vanatalu, K., Vilu, R. 1982. Adenylate energy charge during batch culture of *Thermoactinomyces vulgaris* 42. *Archives of Microbiology*, 133(2), 142-144.

Kambam, P.K.R., Henson, M.A. 2010. Engineering bacterial processes for cellulosic ethanol production. *Biofuels*, 1(5), 729-743.

Kataeva, I.A., Yang, S.-J., Dam, P., Poole, F.L., Yin, Y., Zhou, F., . . . Sims, D.R. 2009. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. *Journal of Bacteriology*, 191(11), 3760-3761.

Kengen, S.M., Stams, A.M. 1994. Formation of L-alanine as a reduced end product in carbohydrate fermentation by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Archives of Microbiology*, 161(2), 168-175.

Kent, W.J. 2002. BLAT—the BLAST-like alignment tool. *Genome Research*, 12(4), 656-664.

Kim, H.J., Hou, B.K., Lee, S.G., Kim, J.S., Lee, D.-W., Lee, S.J. 2013. Genome-wide analysis of redox reactions reveals metabolic engineering targets for d-lactate overproduction in *Escherichia coli*. *Metabolic Engineering*(0).

Kim, S., Dale, B.E. 2004. Global potential bioethanol production from wasted crops and crop residues. *Biomass and Bioenergy*, 26(4), 361-375.

Kita, A., Iwasaki, Y., Sakai, S., Okuto, S., Takaoka, K., Suzuki, T., . . . Nakashimada, Y. 2012. Development of genetic transformation and heterologous expression system in carboxydotrophic thermophilic acetogen *Moorella thermoacetica*. *Journal of Bioscience and Bioengineering*, 115(4), 347-352.

Kita, A., Iwasaki, Y., Yano, S., Nakashimada, Y., Hoshino, T., Murakami, K. 2013. Isolation of thermophilic acetogens and transformation of them with the *pyrF* and *kanr* genes. *Bioscience, Biotechnology and Biochemistry*, 77(2), 301-306.

Kovács, Á.T., Eckhardt, T.H., van Hartskamp, M., van Kranenburg, R., Kuipers, O.P. 2013. Functional analysis of the ComK protein of *Bacillus coagulans*. *PLoS ONE*, 8(1), e53471.

Kovács, Á.T., Smits, W.K., Mirończuk, A.M., Kuipers, O.P. 2009. Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries. *Environmental Microbiology*, 11(8), 1911-1922.

Kovacs, A.T., van Hartskamp, M., Kuipers, O.P., van Kranenburg, R. 2010. Genetic tool development for a new host for biotechnology, the thermotolerant bacterium *Bacillus coagulans*. *Applied and Environment Microbiology*, 76(12), 4085-4088.

Kroutil, W., Mang, H., Edegger, K., Faber, K. 2004. Recent advances in the biocatalytic reduction of ketones and oxidation of sec-alcohols. *Current Opinion in Chemical Biology*, 8(2), 120-126.

Kuhnert, P., Scholten, E., Haefner, S., Mayor, D., Frey, J. 2010. *Basfia succiniciproducens* gen. nov., sp. nov., a new member of the family *Pasteurellaceae* isolated from bovine rumen. *International Journal of Systematic and Evolutionary Microbiology*, 60(1), 44-50.

Kumar, L., Awasthi, G., Singh, B. 2011. Extremophiles: A novel source of industrially important enzymes. *Biotechnology*, 10(2), 121-135.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., . . . Danchin, A. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657), 249-256.

Lailaja, V.P., Chandrasekaran, M. 2013. Detergent compatible alkaline lipase produced by marine *Bacillus smithii* BTMS 11. *World Journal of Microbiology and Biotechnology*, 29(8), 1349-1360.

Lan, E.I., Liao, J.C. 2013. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. *Bioresource Technology*, 135, 339-349.

Laouami, S., Messaoudi, K., Alberto, F., Clavel, T., Duport, C. 2011. Lactate dehydrogenase A promotes communication between carbohydrate catabolism and virulence in *Bacillus cereus*. *Journal of Bacteriology*, 193(7), 1757-1766.

Lasko, D.R., Zamboni, N., Sauer, U. 2000. Bacterial response to acetate challenge: a comparison of tolerance among species. *Applied Microbiology and Biotechnology*, 54(2), 243-247.

Lee, P., Lee, S., Hong, S., Chang, H. 2002. Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen. *Applied Microbiology and Biotechnology*, 58(5), 663-668.

Lee, S.J., Song, H., Lee, S.Y. 2006. Genome-based metabolic engineering of *Mannheimia succiniproducens* for succinic acid production. *Applied and Environmental Microbiology*, 72(3), 1939-1948.

Lee, Z.M.-P., Bussema, C., Schmidt, T.M. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Research*, 37(suppl 1), D489-D493.

Lerondel, G., Doan, T., Zamboni, N., Sauer, U., Aymerich, S. 2006. YtsJ has the major physiological role of the four paralogous malic enzyme isoforms in *Bacillus subtilis*. *Journal of Bacteriology*, 188(13), 4727-4736.

Leu, S.Y., Zhu, J.Y. 2013. Substrate-related factors affecting enzymatic saccharification of lignocelluloses: Our recent understanding. *Bioenergy Research*, 6(2), 405-415.

Li, L., Su, F., Wang, Y., Zhang, L., Liu, C., Li, J., . . . Xu, P. 2012. Genome sequences of two thermophilic *Bacillus licheniformis* strains, efficient producers of platform chemical 2,3-Butanediol. *Journal of Bacteriology*, 194(15), 4133-4134.

Li, L., Zhang, L., Li, K., Wang, Y., Gao, C., Han, B., . . . Xu, P. 2013. A newly isolated *Bacillus licheniformis* strain thermophilically produces 2,3-butanediol, a platform and fuel bio-chemical. *Biotechnology for Biofuels*, 6(1), Art. nr. 123.

Li, Q., Siles, J.A., Thompson, I.P. 2010. Succinic acid production from orange peel and wheat straw by batch fermentations of *Fibrobacter succinogenes* S85. *Applied Microbiology and Biotechnology*, 1-8.

Li, R., Li, L., Huang, R., Sun, Y., Mei, X., Shen, B., Shen, Q. 2014. Variations of culturable thermophilic microbe numbers and bacterial communities during the thermophilic phase of composting. *World Journal of Microbiology and Biotechnology*, 30(6), 1737-1746.

Liang, L.-y., Liu, R.-m., Ma, J.-f., Chen, K.-q., Jiang, M., Wei, P. 2011. Increased production of succinic acid in *Escherichia coli* by overexpression of malate dehydrogenase. *Biotechnology Letters*, 33(12), 2439-2444.

Lin, H., Bennett, G.N., San, K.Y. 2005. Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield. *Metabolic Engineering*, 7(2), 116-127.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., Fernie, A.R. 2006. Gas chromatography mass spectrometry-based metabolite profiling in plants. 1(1), 387-396.

Liu, B., Wang, C., Yang, H., Tan, H. 2012. Establishment of a genetic transformation system and its application in *Thermoanaerobacter tengcongensis*. *Journal of Genetics and Genomics*, 39(10), 561-570.

Lopilato, J., Bortner, S., Beckwith, J. 1986. Mutations in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives. *Molecular and General Genetics MGG*, 205(2), 285-290.

Lücking, G., Stoeckel, M., Atamer, Z., Hinrichs, J., Ehling-Schulz, M. 2013. Characterization of aerobic spore-forming bacteria associated with industrial dairy processing environments and product spoilage. *International Journal of Food Microbiology*, 166(2), 270-279.

Ludwig, W., Schleifer, K., Whitman, W. 2009. Class I. Bacilli class nov. in: *Bergey's Manual of Systematic Bacteriology*, Vol. 3, pp. 19-20.

Lynd, L.R., Weimer, P.J., Van Zyl, W.H., Pretorius, I.S. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, 66(3), 506-577.

Ma, C., Wang, A., Qin, J., Li, L., Ai, X., Jiang, T., . . . Xu, P. 2009. Enhanced 2,3-butanediol production by *Klebsiella pneumoniae* SDM. *Applied Microbiology and Biotechnology*, 82(1), 49-57.

Ma, K., Maeda, T., You, H., Shirai, Y. 2014. Open fermentative production of L-lactic acid with high optical purity by thermophilic *Bacillus coagulans* using excess sludge as nutrient. *Bioresource Technology*, 151(0), 28-35.

Maas, R., Bakker, R., Jansen, M., Visser, D., de Jong, E., Eggink, G., Weusthuis, R. 2008. Lactic acid production from lime-treated wheat straw by *Bacillus coagulans*: neutralization of acid by fed-batch addition of alkaline substrate. *Applied Microbiology and Biotechnology*, 78(5), 751-758.

Mai, V., Lorenz, W.W., Wiegel, J. 1997. Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance. *FEMS Microbiology Letters*, 148(2), 163-167.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., . . . Koonin, E.V. 2011. Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, 9(6), 467-477.

March, J., Eiteman, M., Altman, E. 2002. Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in *Escherichia coli*. *Applied and Environmental Microbiology*, 68(11), 5620-5624.

McDonald, J.E., Rooks, D.J., McCarthy, A.J. 2012. Methods for the isolation of cellulose-degrading microorganisms. *Methods in Enzymology*, 510, 349-374.

Mearls, E.B., Olson, D.G., Herring, C.D., Lynd, L.R. 2015. Development of a regulatable plasmid-based gene expression system for *Clostridium thermocellum*. *Applied Microbiology and Biotechnology*.

Mehmeti, I., Faergestad, E.M., Bekker, M., Snipen, L., Nes, I.F., Holo, H. 2012. Growth rate-dependent control in *Enterococcus faecalis*: Effects on the transcriptome and proteome, and strong regulation of lactate dehydrogenase. *Applied and Environmental Microbiology*, 78(1), 170-176.

Mehmeti, I., Jönsson, M., Fergestad, E.M., Mathiesen, G., Nes, I.F., Holo, H. 2011. Transcriptome, proteome, and metabolite analyses of a lactate dehydrogenase-negative mutant of *Enterococcus faecalis* V583. *Applied and Environmental Microbiology*, 77(7), 2406-2413.

Meißner, T., Eckelt, E., Basler, T., Meens, J., Heinzmann, J., Suwandi, A., . . . Goethe, R. 2014. The *Mycobacterium avium* ssp. *paratuberculosis* specific *mptD* gene is required for maintenance of the metabolic homeostasis necessary for full virulence in mouse infections. *Frontiers in cellular and infection microbiology*, 2014(4), 110.

Meng, Y., Xue, Y., Yu, B., Gao, C., Ma, Y. 2012. Efficient production of L-lactic acid with high optical purity by alkaliphilic *Bacillus* sp. WL-S20. *Bioresource Technology*, 116(0), 334-339.

Messner, P., Scheberl, A., Schweigkofler, W., Hollaus, F., Rainey, F.A., Burghardt, J., Prillinger, H. 1997. Taxonomic comparison of different thermophilic sugar beet isolates with glycosylated surface layer (S-Layer) proteins and their affiliation to *Bacillus smithii*. *Systematic and Applied Microbiology*, 20(4), 559-565.

Mirończuk, A.M., Kovács, Á.T., Kuipers, O.P. 2008. Induction of natural competence in *Bacillus cereus* ATCC14579. *Microbial Biotechnology*, 1(3), 226-235.

Mojović, L., Pejin, D., Rakin, M., Pejin, J., Nikolić, S., Djukić-Vuković, A. 2012. How to improve the economy of bioethanol production in Serbia. *Renewable and Sustainable Energy Reviews*, 16(8), 6040-6047.

Morozkina, E., Slutskaya, E., Fedorova, T., Tugay, T., Golubeva, L., Koroleva, O. 2010. Extremophilic microorganisms: Biochemical adaptation and biotechnological application. *Applied Biochemistry and Microbiology*, 46(1), 1-14.

Morrow, J.F., Cohen, S.N., Chang, A.C.Y., Boyer, H.W., Goodman, H.M., Helling, R.B. 1974. Replication and transcription of eukaryotic DNA in *Esherichia coli*. *Proceedings of the National Academy of Sciences*, 71(5), 1743-1747.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5(7), 621-8.

Murray, R. 1984. The higher taxa, or, a place for everything. *Bergey's Manual of Systematic Bacteriology*, 1, 31-34.

Nair, L.S., Laurencin, C.T. 2007. Biodegradable polymers as biomaterials. *Progress in Polymer Science*, 32(8–9), 762-798.

Nakamura, L.K., Blumenstock, I., Claus, D. 1988. Taxonomic study of *Bacillus coagulans* Hammer 1915 with a proposal for *Bacillus smithii* sp. nov. *International Journal of Systematic Bacteriology*, 38(1), 63-73.

Narumi, I., Sawakami, K., Nakamoto, S., Nakayama, N., Yanagisawa, T., Takahashi, N., Kihara, H. 1992. A newly isolated *Bacillus stearotheromophilus* K1041 and its transformation by electroporation. *Biotechnology Techniques*, 6(1), 83-86.

Nghiem, N., Davison, B., Suttle, B., Richardson, G. 1997. Production of succinic acid by *Anaerobiospirillum succiniciproducens*. *Applied Biochemistry and Biotechnology*, 63-65(1), 565-576.

O'Donnell, A.G., Norris, J.R., Berkeley, R.C.W. 1980. Characterization of *Bacillus subtilis, Bacillus pumilus, Bacillus licheniformis,* and *Bacillus amyloliquefaciens* by pyrolysis gas-liquid chromatography, deoxyribonucleic acid-deoxyribonucleic acid hybridization, biochemical tests, and API systems. *International Journal of Systematic Bacteriology*, 30(2), 448-459.

Oh, J.-H., van Pijkeren, J.-P. 2014. CRISPR–Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Research*, 42(17), e131.

Okano, K., Yoshida, S., Yamada, R., Tanaka, T., Ogino, C., Fukuda, H., Kondo, A. 2009. Improved production of homo-D-lactic acid via xylose fermentation by introduction of xylose assimilation genes and redirection of the phosphoketolase pathway to the pentose phosphate pathway in L-lactate dehydrogenase gene-deficient *Lactobacillus plantarum*. *Applied and Environmental Microbiology*, 75(24), 7858-7861.

Oliver, J.W.K., Machado, I.M.P., Yoneda, H., Atsumi, S. 2013. Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proceedings of the National Academy of Sciences*, 110(4), 1249-1254.

Olofsson, K., Bertilsson, M., Lidén, G. 2008. A short review on SSF - An interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnology for Biofuels*, 1, art.nr. 7.

Olson, D., Lynd, L. 2012. Computational design and characterization of a temperature-sensitive plasmid replicon for gram positive thermophiles. *Journal of Biological Engineering*, 6(1), art.nr.5.

Olson, D.G., McBride, J.E., Joe Shaw, A., Lynd, L.R. 2012. Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology*, 23(3), 396-405.

Oren, A. 2010. Industrial and environmental applications of halophilic microorganisms. *Environmental Technology*, 31(8-9), 825-834.

Otagiri, M., Ui, S., Takusagawa, Y., Ohtsuki, T., Kurisu, G., Kusunoki, M. 2010. Structural basis for chiral substrate recognition by two 2,3-butanediol dehydrogenases. *FEBS Letters*, 584(1), 219-223.

Ou, M., Ingram, L., Shanmugam, K. 2011. L(+)-Lactic acid production from non-food carbohydrates by thermotolerant *Bacillus coagulans*. *Journal of Industrial Microbiology & Biotechnology*, 38(5), 599-605.

Ou, M.S., Mohammed, N., Ingram, L.O., Shanmugam, K.T. 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), 379-385.

Ouyang, J., Ma, R., Zheng, Z., Cai, C., Zhang, M., Jiang, T. 2013. Open fermentative production of L-lactic acid by *Bacillus* sp. strain NL01 using lignocellulosic hydrolyzates as low-cost raw material. *Bioresource Technology*, 135, 475-480.

Ozkan, M., Desai, S.G., Zhang, Y., Stevenson, D.M., Beane, J., White, E.A., . . . Lynd, L.R. 2001. Characterization of 13 newly isolated strains of anaerobic, cellulolytic, thermophilic bacteria. *Journal of Industrial Microbiology and Biotechnology*, 27(5), 275-280.

Patel, M., Ou, M., Ingram, L.O., Shanmugam, K.T. 2004. Fermentation of sugar cane bagasse hemicellulose hydrolysate to L(+)-lactic acid by a thermotolerant acidophilic *Bacillus* sp. *Biotechnology Letters*, 26(11), 865-868.

Patel, M.A., Ou, M.S., Harbrucker, R., Aldrich, H.C., Buszko, M.L., Ingram, L.O., Shanmugam, K.T. 2006. Isolation and characterization of acid-tolerant, thermophilic bacteria for effective fermentation of biomass-derived sugars to lactic acid. *Applied and Environmental Microbiology*, 72(5), 3228-3235.

Patel, M.A., Ou, M.S., Ingram, L.O., Shanmugam, K.T. 2005. Simultaneous saccharification and co-fermentation of crystalline cellulose and sugar cane bagasse hemicellulose hydrolysate to lactate by a thermotolerant acidophilic *Bacillus* sp. *Biotechnology Progress*, 21(5), 1453-1460.

Patnaik, R. 2008. Engineering complex phenotypes in industrial strains. *Biotechnology Progress*, 24(1), 38-47.

Peng, H., Fu, B., Mao, Z., Shao, W. 2006. Electrotransformation of *Thermoanaerobacter ethanolicus* JW200. *Biotechnology Letters*, 28(23), 1913-1917.

Peng, H., Li, H., Luo, H., Xu, J. 2013. A novel combined pretreatment of ball milling and microwave irradiation for enhancing enzymatic hydrolysis of microcrystalline cellulose. *Bioresource Technology*, 130, 81-87.

Playne, M.J. 1984. Increased digestibility of bagasse by pretreatment with alkalis and steam explosion. *Biotechnology and Bioengineering*, 26(5), 426-433.

Podkaminer, K.K., Kenealy, W.R., Herring, C.D., Hogsett, D.A., Lynd, L.R. 2012. Ethanol and anaerobic conditions reversibly inhibit commercial cellulase activity in thermophilic simultaneous saccharification and fermentation (tSSF). *Biotechnology for Biofuels*, 5, art. nr. 43.

Poli, A., Di Donato, P., Abbamondi, G.R., Nicolaus, B. 2011. Synthesis, production, and biotechnological applications of exopolysaccharides and polyhydroxyalkanoates by archaea. *Archaea*, 2011, art.nr.693253.

Portnoy, V.A., Bezdan, D., Zengler, K. 2011. Adaptive laboratory evolution-harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology*, 22(4), 590-594.

Prevot, A. 1953. Dictionnaire des bactéries pathogens. *Hauduroy P, Ehringer G, Guillot G, Magrou J, Prevot AR, Rosset, Urbain A*, 1-692.

Qi, X., Zhang, Y., Tu, R., Lin, Y., Li, X., Wang, Q. 2011. High-throughput screening and characterization of xylose-utilizing, ethanol-tolerant thermophilic bacteria for bioethanol production. *Journal of Applied Microbiology*, 110(6), 1584-1591.

Qin, J., Wang, X., Zheng, Z., Ma, C., Tang, H., Xu, P. 2010. Production of L-lactic acid by a thermophilic *Bacillus* mutant using sodium hydroxide as neutralizing agent. *Bioresource Technology*, 101(19), 7570-7576.

Qin, J., Zhao, B., Wang, X., Wang, L., Yu, B., Ma, Y., . . . Xu, P. 2009. Non-sterilized fermentative production of polymer-grade L-lactic acid by a newly isolated thermophilic strain *Bacillus* sp. 2–6. *PLoS ONE*, 4(2), e4359.

Quillaguamán, J., Delgado, O., Mattiasson, B., Hatti-Kaul, R. 2006. Poly(β-hydroxybutyrate) production by a moderate halophile, *Halomonas boliviensis* LC1. *Enzyme and Microbial Technology*, 38(1–2), 148-154.

Quillaguamán, J., Guzmán, H., Van-Thuoc, D., Hatti-Kaul, R. 2010. Synthesis and production of polyhydroxyalkanoates by halophiles: current potential and future prospects. *Applied Microbiology and Biotechnology*, 85(6), 1687-1696.

Ramos, A., Neves, A.R., Ventura, R., Maycock, C., López, P., Santos, H. 2004. Effect of pyruvate kinase overproduction on glucose metabolism of *Lactococcus lactis*. *Microbiology*, 150(4), 1103-1111.

Rathi, D.N., Amir, H.G., Abed, R.M.M., Kosugi, A., Arai, T., Sulaiman, O., . . . Sudesh, K. 2013. Polyhydroxyalkanoate biosynthesis and simplified polymer recovery by a novel moderately halophilic bacterium isolated from hypersaline microbial mats. *Journal of Applied Microbiology*, 114(2), 384-395.

Rhee, M.S., Kim, J.w., Qian, Y., Ingram, L.O., Shanmugam, K.T. 2007. Development of plasmid vector and electroporation condition for gene transfer in sporogenic lactic acid bacterium, *Bacillus coagulans*. *Plasmid*, 58(1), 13-22.

Rhee, M.S., Moritz, B.E., Xie, G., Glavina Del Rio, T., Dalin, E., Tice, H., . . . Shanmugam, K.T. 2011. Complete genome sequence of a thermotolerant sporogenic lactic acid bacterium, *Bacillus coagulans* strain 36D1. *Standards in Genomic Sciences*, 5(3), 331-340.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., . . . Keasling, J.D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. 440(7086), 940-943.

Robbins, M.P., Evans, G., Valentine, J., Donnison, I.S., Allison, G.G. 2012. New opportunities for the exploitation of energy crops by thermochemical conversion in Northern Europe and the UK. *Progress in Energy and Combustion Science*, 38(2), 138-155.

Rodriguez-Valera, F., Juez, G., Kushner, D.J. 1983. *Halobacterium mediterranei* spec, nov., a new carbohydrate-utilizing extreme halophile. *Systematic and Applied Microbiology*, 4(3), 369-381.

Saier, M.H., Chauvaux, S., Cook, G.M., Deutscher, J., Paulsen, I.T., Reizer, J., Ye, J.J. 1996. Catabolite repression and inducer control in Gram-positive bacteria. *Microbiology (Reading, England)*, 142 (Pt 2), 217-230.

Saitou, N., Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.

Sakaff, M.K.L.M., Rahman, A.Y.A., Saito, J.A., Hou, S., Alam, M. 2012. Complete genome sequence of the thermophilic bacterium *Geobacillus thermoleovorans* CCB_US3_UF5. *Journal of Bacteriology*, 194(5), 1239-1239.

Sakai, K., Yamanami, T. 2006. Thermotolerant *Bacillus licheniformis* TY7 produces optically active L-lactic acid from kitchen refuse under open condition. *Journal of Bioscience and Bioengineering*, 102(2), 132-134.

Sambrook, J., Fritsch, E.F., Maniatis, T. 1989. *Molecular cloning: a laboratory manual. 2 ed.* Cold Spring Harbor Press, Cold Spring Harbor, NY

Sánchez, A.M., Bennett, G.N., San, K.-Y. 2005. Novel pathway engineering design of the anaerobic central metabolic pathway in *Escherichia coli* to increase succinate yield and productivity. *Metabolic Engineering*, 7(3), 229-239.

Sarethy, I., Saxena, Y., Kapoor, A., Sharma, M., Sharma, S., Gupta, V., Gupta, S. 2011. Alkaliphilic bacteria: applications in industrial biotechnology. *Journal of Industrial Microbiology and Biotechnology*, 38(7), 769-790.

Saripan, A.F., Reungsang, A. 2013. Biohydrogen production by *Thermoanaerobacterium thermosaccharolyticum* KKU-ED1: Culture conditions optimization using xylan as the substrate. *International Journal of Hydrogen Energy*, 38(14), 6167-6173.

Sauer, U., Eikmanns, B.J. 2005. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology Reviews*, 29(4), 765-794.

Saw, J.H., Mountain, B.W., Feng, L., Omelchenko, M.V., Hou, S., Saito, J.A., . . . Alam, M. 2008. Encapsulated in silica: Genome, proteome and physiology of the thermophilic bacterium *Anoxybacillus flavithermus* WK1. *Genome Biology*, 9(11), art.nr. R161.

Schaeffer, A.B., Fulton, M.D. 1933. A Simplified Method of Staining Endospores. *Science*, 77(1990), 194.

Selle, K., Barrangou, R. 2015. Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends in Microbiology*, 23(4), 225-232.

Shao, X., Jin, M., Guseva, A., Liu, C., Balan, V., Hogsett, D., . . . Lynd, L. 2011. Conversion for Avicel and AFEX pretreated corn stover by *Clostridium thermocellum* and simultaneous saccharification and fermentation: Insights into microbial conversion of pretreated cellulosic biomass. *Bioresource Technology*, 102(17), 8040-8045.

Shaw, A.J., Covalla, S.F., Hogsett, D.A., Herring, C.D. 2011. Marker removal system for *Thermoanaerobacterium saccharolyticum* and development of a markerless ethanologen. *Applied and Environmental Microbiology*, 77(7), 2534-2536.

Shaw, A.J., Hogsett, D.A., Lynd, L.R. 2010. Natural competence in *Thermoanaerobacter* and *Thermoanaerobacterium* species. *Applied and Environmental Microbiology*, 76(14), 4713-4719.

Shaw, A.J., Podkaminer, K.K., Desai, S.G., Bardsley, J.S., Rogers, S.R., Thorne, P.G., . . . Lynd, L.R. 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America*, 105(37), 13769-13774.

Sheldon, R.A. 2011. Utilisation of biomass for sustainable fuels and chemicals: Molecules, methods and metrics. *Catalysis Today*, 167(1), 3-13.

Shimizu, M., Fujii, T., Masuo, S., Takaya, N. 2010. Mechanism of *de novo* branched-chain amino acid synthesis as an alternative electron sink in hypoxic *Aspergillus nidulans* cells. *Applied and Environmental Microbiology*, 76(5), 1507-1515.

Sizova, M.V., Izquierdo, J.A., Panikov, N.S., Lynd, L.R. 2011. Cellulose- and xylan-degrading thermophilic anaerobic bacteria from biocompost. *Applied and Environment Microbiology*, 77(7), 2282-2291.

Skerman, V.B.D., McGowan, V., Sneath, P.H.A. 1980. Approved lists of bacterial names. *International Journal of Systematic Bacteriology*, 30(1), 225-420.

Sridhar, J., Eiteman, M. 2001. Metabolic flux analysis of *Clostridium thermosuccinogenes*. *Applied Biochemistry and Biotechnology*, 94(1), 51-69.

Sridhar, J., Eiteman, M.A., Wiegel, J.W. 2000. Elucidation of enzymes in fermentation pathways used by *Clostridium thermosuccinogenes* growing on inulin. *Applied and Environmental Microbiology*, 66(1), 246-251.

Stoeckel, M., Abduh, S.B.M., Atamer, Z., Hinrichs, J. 2014. Inactivation of bacillus spores in batch vs continuous heating systems at sterilisation temperatures. *International Journal of Dairy Technology*, 67(3), 334-341.

Stols, L., Donnelly, M.I. 1997. Production of succinic acid through overexpression of NAD(+)-dependent malic enzyme in an *Escherichia coli* mutant. *Applied and Environmental Microbiology*, 63(7), 2695-701.

Studholme, D.J. 2014. Some (bacilli) like it hot: genomics of *Geobacillus* species. *Microbial Biotechnology*, 8(1), 40-48.

Studholme, D.J., Jackson, R.A., Leak, D.J. 1999. Phylogenetic analysis of transformable strains of thermophilic Bacillus species. *FEMS Microbiology Letters*, 172(1), 85-90.

Su, F., Tao, F., Tang, H., Xu, P. 2012. Genome sequence of the thermophile *Bacillus coagulans* Hammer, the type strain of the species. *Journal of Bacteriology*, 194(22), 6294-6295.

Su, F., Xu, P. 2014. Genomic analysis of thermophilic *Bacillus coagulans* strains: efficient producers for platform bio-chemicals. *Scientific Reports*, 4, art.nr. 3926.

Su, F., Yu, B., Sun, J., Ou, H.-Y., Zhao, B., Wang, L., . . . Xu, P. 2011a. Genome sequence of the thermophilic strain *Bacillus coagulans* 2-6, an efficient producer of high-optical-purity L-lactic acid. *Journal of Bacteriology*, 193(17), 4563-4564.

Su, Y., Rhee, M., Ingram, L., Shanmugam, K. 2011b. Physiological and fermentation properties of *Bacillus coagulans* and a mutant lacking fermentative lactate dehydrogenase activity. *Journal of Industrial Microbiology & Biotechnology*, 38(3), 441-450.

Suitso, I., Jõgi, E., Orro, T., Kavak, A., Kalmus, K., Viltrop, A., Nurk, A. 2014. *Bacillus Smithii* TBMI12 endospores as a potential component of probiotic feed additive for pigs. *Veterinarija ir Zootechnika*, 66(88), 64-68.

Suzuki, H., Murakami, A., Yoshida, K.-i. 2012. Counterselection system for *Geobacillus kaustophilus* HTA426 through *pyrF* and *pyrR* disruptions. *Applied and Environmental Microbiology*, 78(20), 7376-7383.

Suzuki, H., Yoshida, K.I. 2012. Genetic transformation of *Geobacillus kaustophilus* hta426 by conjugative transfer of host-mimicking plasmids. *Journal of Microbiology and Biotechnology*, 22(9), 1279-1287.

Suzuki, Y., Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M., Abe, S. 1983. *Bacillus thermoglucosidasius* sp. nov., a new species of obligately thermophilic bacilli. *Systematic and Applied Microbiology*, 4(4), 487-495.

Svetlitchnyi, V., Kensch, O., Falkenhan, D., Korseska, S., Lippert, N., Prinz, M., . . . Curvers, S. 2013. Single-step ethanol production from lignocellulose using novel extremely thermophilic bacteria. *Biotechnology for Biofuels*, 6(1), art.nr.31.

Taillefer, M., Rydzak, T., Levin, D.B., Oresnik, I.J., Sparling, R. 2015. Reassessment of the transhydrogenase/malate shunt pathway in *Clostridium thermocellum* ATCC 27405 through kinetic characterization of malic enzyme and malate dehydrogenase. *Applied and Environmental Microbiology*, 81(7), 2423-2432.

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., . . . Horikoshi, K. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Research*, 28(21), 4317-4331.

Takami, H., Nishi, S., Lu, J., Shimamura, S., Takaki, Y. 2004a. Genomic characterization of thermophilic *Geobacillus* species isolated from the deepest sea mud of the Mariana Trench. *Extremophiles*, 8(5), 351-356.

Takami, H., Takaki, Y., Chee, G.-J., Nishi, S., Shimamura, S., Suzuki, H., . . . Uchiyama, I. 2004b. Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Research*, 32(21), 6292-6303.

Takashima, Y., Kawabe, T., Mitsuda, S. 2000. Factors affecting the production of nitrile hydratase by thermophilic *Bacillus smithii* SC-J05-1. *Journal of Bioscience and Bioengineering*, 89(3), 282-284.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725-2729.

Tan, D., Xue, Y.-S., Aibaidula, G., Chen, G.-Q. 2011. Unsterile and continuous production of polyhydroxybutyrate by *Halomonas* TD01. *Bioresource Technology*, 102(17), 8130-8136.

Tanaka, K., Komiyama, A., Sonomoto, K., Ishizaki, A., Hall, S., Stanbury, P. 2002. Two different pathways for D-xylose metabolism and the effect of xylose concentration on the yield coefficient of L-lactate in mixed-acid fermentation by the lactic acid bacterium *Lactococcus lactis* IO-1. *Applied Microbiology and Biotechnology*, 60(1-2), 160-167.

Taylor, M., van Zyl, L., Tuffin, I., Leak, D., Cowan, D. 2011. Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microbial Biotechnology*, 4(4), 438-448.

Taylor, M.P., Eley, K.L., Martin, S., Tuffin, M.I., Burton, S.G., Cowan, D.A. 2009. Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. *Trends in Biotechnology*, 27(7), 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), 45-52.

Taylor, M.P., Mulako, I., Tuffin, M., Cowan, D. 2012. Understanding physiological responses to pre-treatment inhibitors in ethanologenic fermentations. *Biotechnology Journal*, 7(9), 1169-1181.

Team, R.C. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0.

Thauer, R.K. 1988. Citric-acid cycle, 50 years on. Modifications and an alternative pathway in anaerobic bacteria. *European Journal of Biochemistry*, 176(3), 497-508.

Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673-4680.

Tock, M.R., Dryden, D.T.F. 2005. The biology of restriction and anti-restriction. *Current Opinion in Microbiology*, 8(4), 466-472.

Toya, Y., Shimizu, H. 2013. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnology Advances*, 31(6), 818-826.

Tripathi, S.A., Olson, D.G., Argyros, D.A., Miller, B.B., Barrett, T.F., Murphy, D.M., . . . Caiazza, N.C. 2010. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. *Applied and Environmental Microbiology*, 76(19), 6591-6599.

Tsuji, A., Okada, S., Hols, P., Satoh, E. 2013. Metabolic engineering of *Lactobacillus plantarum* for succinic acid production through activation of the reductive branch of the tricarboxylic acid cycle. *Enzyme and Microbial Technology*, 53(2), 97-103.

Tsukahara, K., Kita, A., Nakashimada, Y., Hoshino, T., Murakami, K. 2014. Genome-guided analysis of transformation efficiency and carbon dioxide assimilation by *Moorella thermoacetica* Y72. *Gene*, 535(2), 150-155.

Tyurin, M.V., Desai, S.G., Lynd, L.R. 2004. Electrotransformation of *Clostridium thermocellum*. *Applied and Environment Microbiology*, 70(2), 883-890.

Valladares Juárez, A., Dreyer, J., Göpel, P., Koschke, N., Frank, D., Märkl, H., Müller, R. 2009. Characterisation of a new thermoalkaliphilic bacterium for the production of high-quality hemp fibres, *Geobacillus thermoglucosidasius* strain PB94A. *Applied Microbiology and Biotechnology*, 83(3), 521-527.

van Dam, J.C., Eman, M.R., Frank, J., Lange, H.C., van Dedem, G.W.K., Heijnen, S.J. 2002. Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. *Analytica Chimica Acta*, 460(2), 209-218.

van de Werken, H.J.G., Verhaart, M.R.A., VanFossen, A.L., Willquist, K., Lewis, D.L., Nichols, J.D., . . . Kengen, S.W.M. 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Applied and Environment Microbiology*, 74(21), 6720-6729.

van Hijum, S., de Jong, A., Baerends, R., Karsens, H., Kramer, N., Larsen, R., . . . Kuipers, O. 2005. A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics*, 6(1), 77.

van Kranenburg, R., van Hartskamp, M., Heintz, E., Anthonius, J., Van Mullekom, E., Snelders, J. 2007. Genetic modification of homolactic thermophilic Bacilli. WO Patent WO/2007/085,443

van Maris, A.J.A., Geertman, J.-M.A., Vermeulen, A., Groothuizen, M.K., Winkler, A.A., Piper, M.D.W., . . . Pronk, J.T. 2004. Directed evolution of pyruvate decarboxylase-negative *Saccharomyces cerevisiae*, yielding a C2-Independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Applied and Environmental Microbiology*, 70(1), 159-166.

van Spanning, R.J., Wansell, C.W., Reijnders, W.N., Harms, N., Ras, J., Oltmann, L.F., Stouthamer, A.H. 1991. A method for introduction of unmarked mutations in the genome of *Paracoccus denitrificans*: construction of strains with multiple mutations in the genes encoding periplasmic cytochromes c550, c551i, and c553i. *Journal of Bacteriology*, 173(21), 6962-6970.

Veen, D., Lo, J., Brown, S., Johnson, C., Tschaplinski, T., Martin, M., . . . Lynd, L. 2013. Characterization of *Clostridium thermocellum* strains with disrupted fermentation end-product pathways. *Journal of Industrial Microbiology & Biotechnology*, 40(7), 725-734.

Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K.H., Ehrenreich, P., . . . Gottschalk, G. 2004. The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. 7(4), 204-211.

Vemuri, G.N., Eiteman, M.A., Altman, E. 2002. Effects of growth mode and pyruvate carboxylase on succinic acid production by metabolically engineered strains of *Escherichia coli. Applied and Environmental Microbiology*, 68(4), 1715-1727.

Wahl, A.S., Seifar, R.M., ten Pierick, A., Ras, C., van Dam, J.C., Heijnen, J.J., van Gulik, W.M. 2014. Quantitative metabolomics using ID-MS. in: *Methods in Molecular Biology*, Vol. 1191, pp. 91-105.

Walton, S., Bischoff, K., van Heiningen, A., van Walsum, G. 2010. Production of lactic acid from hemicellulose extracts by *Bacillus coagulans* MXL-9. *Journal of Industrial Microbiology and Biotechnology*, 37(8), 823-830.

Wang, J.J., Greenhut, W.B., Shih, J.C.H. 2005. Development of an asporogenic *Bacillus licheniformis* for the production of keratinase. *Journal of Applied Microbiology*, 98(3), 761-767.

Wang, L., Xue, Z., Zhao, B., Yu, B., Xu, P., Ma, Y. 2012a. Jerusalem artichoke powder: a useful material in producing high-optical-purity L-lactate using an efficient sugar-utilizing thermophilic *Bacillus coagulans* strain. *Bioresource Technology*, 130, 174-180.

Wang, L., Zhao, B., Liu, B., Yu, B., Ma, C., Su, F., . . . Xu, P. 2010. Efficient production of L-lactic acid from corncob molasses, a waste by-product in xylitol production, by a newly isolated xylose utilizing *Bacillus* sp. strain. *Bioresource Technology*, 101(20), 7908-7915.

Wang, Q., Chen, T., Zhao, X., Chamu, J. 2012b. Metabolic engineering of thermophilic *Bacillus licheniformis* for chiral pure D-2,3-butanediol production. *Biotechnology and Bioengineering*, 109(7), 1610-1621.

Wang, Q., Ingram, L.O., Shanmugam, K.T. 2011a. Evolution of D-lactate dehydrogenase activity from glycerol dehydrogenase and its utility for D-lactate production from lignocellulose. *Proceedings of the National Academy of Sciences*, 108(47), 18920-18925.

Wang, Q., Zhao, X., Chamu, J., Shanmugam, K.T. 2011b. Isolation, characterization and evolution of a new thermophilic *Bacillus licheniformis* for lactic acid production in mineral salts medium. *Bioresource Technology*, 102(17), 8152-8158.

Watson, J.D., Crick, F.H. 1953. Molecular structure of nucleic acids. *Nature*, 171(4356), 737-738.

Weber, C., Farwick, A., Benisch, F., Brat, D., Dietz, H., Subtil, T., Boles, E. 2010. Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. *Applied Microbiology and Biotechnology*, 87(4), 1303-1315.

Wee, Y., Kim, J., Ryu, H. 2006. Biotechnological production of lactic acid and its recent applications. *Food Technology and Biotechnology*, 44(2), 163-172.

Wellisch, M., Jungmeier, G., Karbowski, A., Patel, M.K., Rogulska, M. 2010. Biorefinery systems-potential contributors to sustainable innovation. *Biofuels, Bioproducts and Biorefining*, 4(3), 275-286.

Werpy, T., Petersen, G., Aden, A., Bozell, J., Holladay, J., White, J., . . . Jones, S. 2004. Top value added chemicals from biomass. Volume 1 - Results of screening for potential candidates from sugars and synthesis gas. DTIC Document. .

Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J., van der Oost, J. 2012. The CRISPRs, they are A-Changin': how prokaryotes generate adaptive immunity. *Genetics*, 46(1), 311-339.

Weusthuis, R.A., Lamot, I., van der Oost, J., Sanders, J.P.M. 2011. Microbial production of bulk chemicals: development of anaerobic processes. *Trends in Biotechnology*, 29(4), 153-158.

Wieschalka, S., Blombach, B., Eikmanns, B. 2012. Engineering *Corynebacterium glutamicum* for the production of pyruvate. *Applied Microbiology and Biotechnology*, 94(2), 449-459.

Willquist, K., van Niel, E.W.J. 2012. Growth and hydrogen production characteristics of *Caldicellulosiruptor saccharolyticus* on chemically defined minimal media. *International Journal of Hydrogen Energy*, 37(6), 4925-4929.

Woese, C.R., Kandler, O., Wheelis, M.L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12), 4576-4579.

Wongwilaiwalin, S., Rattanachomsri, U., Laothanachareon, T., Eurwilaichitr, L., Igarashi, Y., Champreda, V. 2010. Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system. *Enzyme and Microbial Technology*, 47(6), 283-290.

Worm, P., Koehorst, J.J., Visser, M., Sedano-Núñez, V.T., Schaap, P.J., Plugge, C.M., . . . Stams, A.J.M. 2014. A genomic view on syntrophic versus non-syntrophic lifestyle in anaerobic fatty acid degrading communities. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1837(12), 2004-2016.

Wright, E. 2013. DECIPHER: database enabled code for ideal probe hybridization employing R. *R package version*, 1(0).

Wu, L., Mashego, M.R., van Dam, J.C., Proell, A.M., Vinke, J.L., Ras, C., . . . Heijnen, J.J. 2005. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards. *Analytical Biochemistry*, 336(2), 164-171.

Xiao, Z., Wang, X., Huang, Y., Huo, F., Zhu, X., Xi, L., Lu, J. 2012. Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain. *Biotechnology for Biofuels*, 5(1), art.nr.88.

Xu, P., Qiu, J., Gao, C., Ma, C. 2008. Biotechnological routes to pyruvate production. *Journal of Bioscience and Bioengineering*, 105(3), 169-175.

Xue, G.-P., Johnson, J.S., Dalrymple, B.P. 1999. High osmolarity improves the electro-transformation efficiency of the gram-positive bacteria *Bacillus subtilis* and *Bacillus licheniformis*. *Journal of Microbiological Methods*, 34(3), 183-191.

Yang, X., Lai, Z., Lai, C., Zhu, M., Li, S., Wang, J., Wang, X. 2013. Efficient production of L-lactic acid by an engineered *Thermoanaerobacterium aotearoense* with broad substrate specificity. *Biotechnology for Biofuels*, 6(1), art.nr. 124.

Yao, S., Mikkelsen, M.J. 2010a. Identification and overexpression of a bifunctional aldehyde/alcohol dehydrogenase responsible for ethanol production in *Thermoanaerobacter mathranii*. *Journal of Molecular Microbiology and Biotechnology*, 19(3), 123-133.

Yao, S., Mikkelsen, M.J. 2010b. Metabolic engineering to improve ethanol production in *Thermoanaerobacter mathranii*. *Applied Microbiology and Biotechnology*, 88(1), 199-208.

Ye, L., Zhou, X., Hudari, M.S.B., Li, Z., Wu, J.C. 2013. Highly efficient production of L-lactic acid from xylose by newly isolated *Bacillus coagulans* C106. *Bioresource Technology*, 132, 38-44.

Yee, K., Rodriguez Jr, M., Tschaplinski, T., Engle, N., Martin, M., Fu, C., . . . Mielenz, J. 2012. Evaluation of the bioconversion of genetically modified switchgrass using simultaneous saccharification and fermentation and a consolidated bioprocessing approach. *Biotechnology for Biofuels*, 5(1), art.nr.81.

Yudkin, M.D. 1987. Structure and function in a *Bacillus subtilis* sporulation-specific sigma factor: Molecular nature of mutations in *spoIIAC*. *Journal of General Microbiology*, 133(3), 475-481.

Zeigler, D.R. 2014. The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology*, 160(Pt 1), 1-11.

Zelić, B., Gostović, S., Vuorilehto, K., Vasić-Rac̆ki, Đ., Takors, R. 2004. Process strategies to enhance pyruvate production with recombinant *Escherichia coli:* From repetitive fed-batch to in situ product recovery with fully integrated electrodialysis. *Biotechnology and Bioengineering*, 85(6), 638-646.

Zeng, A.-P., Sabra, W. 2011. Microbial production of diols as platform chemicals: Recent progresses. *Current Opinion in Biotechnology*, 22(6), 749-757.

Zhang, X., Jantama, K., Moore, J.C., Jarboe, L.R., Shanmugam, K.T., Ingram, L.O. 2009. Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20180-20185.

Zhao, B., Wang, L., Ma, C., Yang, C., Xu, P., Ma, Y. 2010. Repeated open fermentative production of optically pure L-lactic acid using a thermophilic *Bacillus* sp. strain. *Bioresource Technology*, 101(16), 6494-6498.

Zhao, X., Zhang, L., Liu, D. 2012a. Biomass recalcitrance. Part II: Fundamentals of different pre-treatments to increase the enzymatic digestibility of lignocellulose. *Biofuels, Bioproducts and Biorefining*, 6(5), 561-579.

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

Zhou, J., Olson, D.G., Argyros, D.A., Deng, Y., van Gulik, W.M., van Dijken, J.P., Lynd, L.R. 2013a. Atypical glycolysis in *Clostridium thermocellum*. *Applied and Environmental Microbiology*, 79(9), 3000-3008.

Zhou, X., Ye, L., Wu, J. 2013b. Efficient production of L-lactic acid by newly isolated thermophilic *Bacillus coagulans* WCP10-4 with high glucose tolerance. *Applied Microbiology and Biotechnology*, 97(10), 4309-4314.

Zhu, L.-W., Li, X.-H., Zhang, L., Li, H.-M., Liu, J.-H., Yuan, Z.-P., . . . Tang, Y.-J. 2013. Activation of glyoxylate pathway without the activation of its related gene in succinate-producing engineered *Escherichia coli*. *Metabolic Engineering*, 20(0), 9-19.

Zhu, Y., Eiteman, M.A., Altman, R., Altman, E. 2008. High glycolytic flux improves pyruvate production by a metabolically engineered *Escherichia coli* strain. *Applied and Environmental Microbiology*, 74(21), 6649-6655.

Co-author affiliations

Martinus J.A. Daas[1]
Martijn Diender[1]
Sacha A. F. T. van Hijum[2]
Richard van Kranenburg[1,3]
Aleksander J. Kruis[1]
Jasper J. Koehorst[4]
John van der Oost[1]
Bernadet Renckens[2]
Peter J. Schaap[4]
Laurens van der Vlist[1]
Willem M. de Vos[1]
Bastienne Vriesendorp[3]
S. Aljoscha Wahl[5]
Antonius H.P. van de Weijer[1]

[1]Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands.

[2]CMBI, NCMLS, Geert-Grooteplein Zuid 26-28, 6525 GA, The Netherlands.

[3]Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands.

[4]Laboratory of Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands.

[5]Delft University of Technology, Julianalaan 67, 2628 BC Delft, Netherlands.

Acknowledgements / dankwoord

All the ups and downs included, I have enjoyed every moment of this project and learned so much; not only about research but also about many other aspects surrounding it. All that could not have been possible without the help, collaboration, support and care of many people, to whom this section is dedicated. I will try to keep it short, but for many of these people the simple 'thank you' that can be included here is only a fraction of the gratitude I feel and would like to express.

Tom, it has been a great pleasure working together with you. Thanks a lot for all your efforts for the project, for your no-nonsense attitude, for organizing so many nice events like the boerenkool dinner, Veluweloop and the 'BacGen beer stübe' and for sharing all the ups and downs during the four years of this project. I am very happy for both you and Microbiology that you can continue working there, I wish you all the best both professionally and personally and hope we will stay in touch and work together again in the future.

Richard, many thanks for all your support, involvement and patience. Whenever I had a question or was once again in doubt about something, I could always drop by your office for a talk on Mondays and if you weren't there you were always quick in replying via mail or phone. I am very grateful that you gave me the opportunity to do my PhD on this project by asking me if I wanted to apply for it at the time I was a student at Corbion. It's been an honour to be your first PhD student and I am very happy that you are/will be supervising more people, which I am sure will be very successful.

John, thanks a lot for the great BacGen times and for having me in the group! Your enthusiasm, drive and never-ending flow of ideas are great, as is the atmosphere in your group. The fact that I (and others) can always drop by your office to ask advice on anything also greatly contributed to the good feeling about my time at BacGen.

Willem, thank you for your great input in the project and for being such an inspirational leader and scientist. I always very much enjoyed our work discussions and project meetings, where you always took care that discussions stayed to the point and at high speed. Also for you it holds true that I could always drop by, mail, or call whenever necessary, which I highly appreciate and really contributes to a great collaboration and atmosphere.

Yannis, I am really happy to work together with you in the follow-up project of this thesis. Thanks a lot for all the good times with working, travelling and having fun together and sharing all the good and bad moments. I am sure you will make the new project a big success and am very glad to still be working together, and I'm honoured and happy that you and Tom are my paranymphs!

Jacques and **Aljoscha**, thank you for all the help, patience and pleasant collaboration with the metabolomics parts of the project. **Peter, Jasper, Bastienne, Bernadette** and **Sacha**, thanks for all your input and the nice collaboration with the genomics and transcriptomics parts. I am happy to be able to continue many of our collaborations in the current project.

To all my current and former office mates, thanks so much for all the great times both in the lab/office and outside. **Aleks**, it was great to have you first as a student and now as an office mate, travel companion and running buddy - thanks for pulling me through many tough kilometres! **Tijn, Tom, Yannis, Bram, Mark, Amos, Slav, John R.**, thanks for all the good moments with both work and nonsense discussions, office-beers, office-dinners, and for keeping me sharp and making me laugh via your inventive office-humour, whether that was an inverted computer screen or kruidnoten flying through the air...

I want to thank all the other current and former BacGen and other group members for the great atmosphere in the group. Whether there was something to celebrate or something the opposite of that, there was always support and you all always want to help. **Edze, Matthijs, Daan, Stan, Marnix, Tim, Yifan, Nico, Melvin, Vincent, Tessa, Bas, Teunke, Pradna, Hanne, Servé, Raymond, Franklin, Sjoerd** and **Thomas**, thanks a lot! **Kiira** and **Nico**, thanks for the nice times we had in Vancouver and afterwards with organizing the Metabolic Engineering Tools meetings. Thanks to the whole MIB-group for the great atmosphere, the nice breaks we had together and all the other events like We-Day, Veluweloop, barbeques, lab trips, drinks and everything else that makes both the work-part and the non-work part at Micro so enjoyable.

Wim, Anja, Carolien, Sjon and **Philippe**, thanks a lot for all your support and for always helping me out with all sorts of questions on administration, ordering, computers, etc.!

I also want to thank all my students **Tijn, Aleks, Sjuul, Laurens, Ward, Martijn,** and **Mike**. I always thought it was out of politeness that people said in their acknowledgements that they learned a lot from their students, but that is not the case, as I now know that it is true. I truly learned a lot from all of you! Thanks for your inputs in the project and for being such a nice part of our group. Many of you now are professionals in microbiology yourselves, which for sure you will do great.

Besides all the people involved professionally, I am also deeply grateful to all my friends and family who supported me outside the lab, for giving me both the opportunity to relax and talk about other things than work, and for never complaining about me being so absorbed

by the work that I didn't always have time to catch up…

Nadine, Annemarie, Iris, Susanne, **Sven & Gerliene, Jacob, Martijn, Barbara & Jan, Lasse & Kwang & Trihn,** heel erg bedankt voor alle fijne momenten met wandelingen, films, concerten, pizza nights, goed gezelschap, eten, muziek en alle andere leuke dingen. Ik hoop met jullie allemaal nog vele jaren vriendschap te mogen delen!

Papa en mama, heel erg bedankt voor al jullie support, vertrouwen, liefde en samenzijn en dat jullie ons altijd steunen om de dingen te doen en mogelijk te maken die we het allerleukste en belangrijkste vinden. Van jullie heb ik geleerd onafhankelijk te zijn, hard te werken en om anderen te geven. Het is fijn om te weten dat wat er ook gebeurt, ik altijd bij jullie kan komen. Lieve **Marjoleine**, datzelfde geldt voor jou. Dank je voor het zijn van mijn geweldige zusje, ik ben trots op jou en alles wat je hebt bereikt. En laten we vooral onze eet- en drinkavonden erin blijven houden tot we verstokte kibbelende oudjes zijn :)

Opa en Oma, heel erg bedankt voor al jullie betrokkenheid bij mijn PhD project en alles daarbuiten. Het is altijd leuk om met jullie bij te kunnen praten en jullie interesse in alles wat jullie (klein)kinderen doen is fantastisch. Dank jullie wel voor al jullie liefde, interesse en support door alle jaren heen, ik hoop dat we nog vele jaren van elkaar kunnen genieten!

Mary, ook jij heel erg bedankt voor al je interesse en support, en voor alle mooie momenten met wandelingen, eten, musea, boeken, concerten, en samenzijn. Het is fijn om over zoveel dingen te kunnen praten en discussiëren, juist ook over alle niet-werkzaken die er zo ontzettend veel toe doen.

Zeno, de laatste woorden in dit dankwoord zijn voor jou. Het is vele malen korter dan je verdient, want er zijn simpelweg geen woorden om te beschrijven wat jij voor mij betekent en om je te bedanken voor alles, dat je er altijd voor me bent, en dat we samen zijn.

About the author, publications and training activities

ABOUT THE AUTHOR

Elleke Fenna Bosma was born on April 13th, 1987, in Breukelen, The Netherlands. After completing secondary school (VWO) at RSG Broklede in Breukelen in 2005, she started her BSc studies in Biomedical Sciences at Utrecht University, with a minor in History of Newer Philosophy. After completion of her BSc in 2008, she continued with her MSc in Biomolecular Sciences at the same University. During these studies, she performed her major research thesis at the Membrane Enzymology department of Utrecht University, in the group of Joost Holthuis, under supervision of Fikadu Geta Tafesse and Ana Vacaru. Here she worked on SphingoMyelin Synthase-related protein (SMSr) and the unravelling of its function as an ER-resident ceramide sensor controlling mitochondria-mediated apoptosis, which resulted in co-authorship of a publication on this topic. Subsequently, she performed an internship at Purac BV (now called Corbion) in Gorinchem, The Netherlands, where she worked on the isolation, characterization and transformation of thermophilic bacilli under supervision of Richard van Kranenburg, Mariska van Hartskamp and Connie van Wijngaarden. After receiving her MSc degree *cum laude* in 2010, she continued with her PhD at the Laboratory of Microbiology at Wageningen University under supervision of Richard van Kranenburg, John van der Oost and Willem de Vos. During this PhD thesis, she worked on the development of a novel thermophilic platform organism for the production of green chemicals, which resulted in this thesis. She is currently working as a post-doctoral researcher on a follow-up project in the same lab, focussing on the development of high-throughput genome engineering and screening tools for thermophilic production organisms.

LIST OF PUBLICATIONS

Bosma, E.F., Koehorst, J.J., van Hijum, S.A.F.T., Renckens, B. Vriesendorp, B., van de Weijer, A.H.P., Schaap, P.J., de Vos, W.M., van der Oost, J., van Kranenburg, R. 2015. Complete genome sequence of thermophilic *Bacillus smithii* type strain DSM 4216^{T}. *Submitted.*

Bosma, E.F., van de Weijer, A.H.P., van der Vlist, L., de Vos, W.M., van der Oost, J., van Kranenburg, R. 2015. Establishment of markerless gene deletion tools in thermophilic *Bacillus smithii* and construction of multiple mutant strains. *Microbial Cell Factories*, 14(1), 99.

Bosma, E.F., van de Weijer, A.H.P., Daas, M.J.A., van der Oost, J., de Vos, W.M., van Kranenburg, R. 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), 1874-1883.

Tafesse, F.G., Vacaru, A.M., Bosma, E.F., Hermansson, M., Jain, A., Hilderink, A., Somerharju, P., Holthuis, J.C. 2014. Sphingomyelin synthase-related protein SMSr is a suppressor of ceramide-induced mitochondrial apoptosis. *Journal of Cell Science,* 127(Pt 2), 445-54.

Bosma, E.F., van der Oost, J., de Vos, W.M., van Kranenburg, R. 2013. Sustainable production of bio-based chemicals by extremophiles. *Current Biotechnology,* 2(4), 360-379.

OVERVIEW OF COMPLETED TRAINING ACTIVITIES

Discipline-specific activities

Meetings & conferences

- Metabolic Engineering X conference, Vancouver (CA), 2014
- Netherlands Biotechnology Congress (NBC), Ede (NL), 2014
- STW annual congress, Nieuwegein (NL), 2013
- NWO/ALW Molecular Genetics meeting, Lunteren (NL), 2013
- Gram Positives conference, Montecatini Terme (IT), 2013
- STW annual congress, Nieuwegein (NL), 2012
- Metabolic Engineering IX conference, Biarritz (FR), 2012
- Netherlands Biotechnology Congress (NBC), Ede (Nl), 2012
- NWO/ALW Molecular Genetics meeting, Lunteren (NL), 2011
- STW annual congress, Nieuwegein (NL), 2010
- NWO/ALW Molecular Genetics meeting, Lunteren (NL), 2010

Courses

- European Bioinformatics Institute (EBI) Roadshow, Amsterdam (NL), 2013
- Microbial Physiology and Fermentation Technology, Delft (NL), 2011
- Basic Perl scripting tutorial (CMBI), Wageningen (NL), 2010
- Systems biology course: Statistics of ~omics data analysis, Wageningen (NL), 2010

General courses

- Techniques for Writing and Presenting a Scientific Paper, Wageningen (NL), 2013
- Information Literacy including EndNote Introduction, Wageningen (NL), 2012
- Philosophy and Ethics of Food Science &Technology, Wageningen (NL), 2012
- Radiation Safety Course level 5B, Wageningen (NL), 2010
- Competence assessment, Wageningen (NL), 2011
- VLAG PhD week, Baarlo (NL), 2011

OPTIONALS

- Microbiology PhD study trip to China and Japan, 2011
- Preparation of PhD research proposal
- STW user committee meetings
- Bacterial Genetics group meetings
- Microbiology PhD-PostDoc meetings

COLOPHON

The research in this thesis was financially supported by Corbion (NL).

Cover design: Zeno van den Broek, based on Figure 2 Chapter 6.

Printing: Gildeprint.