Functional genomics of sporulation and microcompartment formation in *Clostridium beijerinckii*
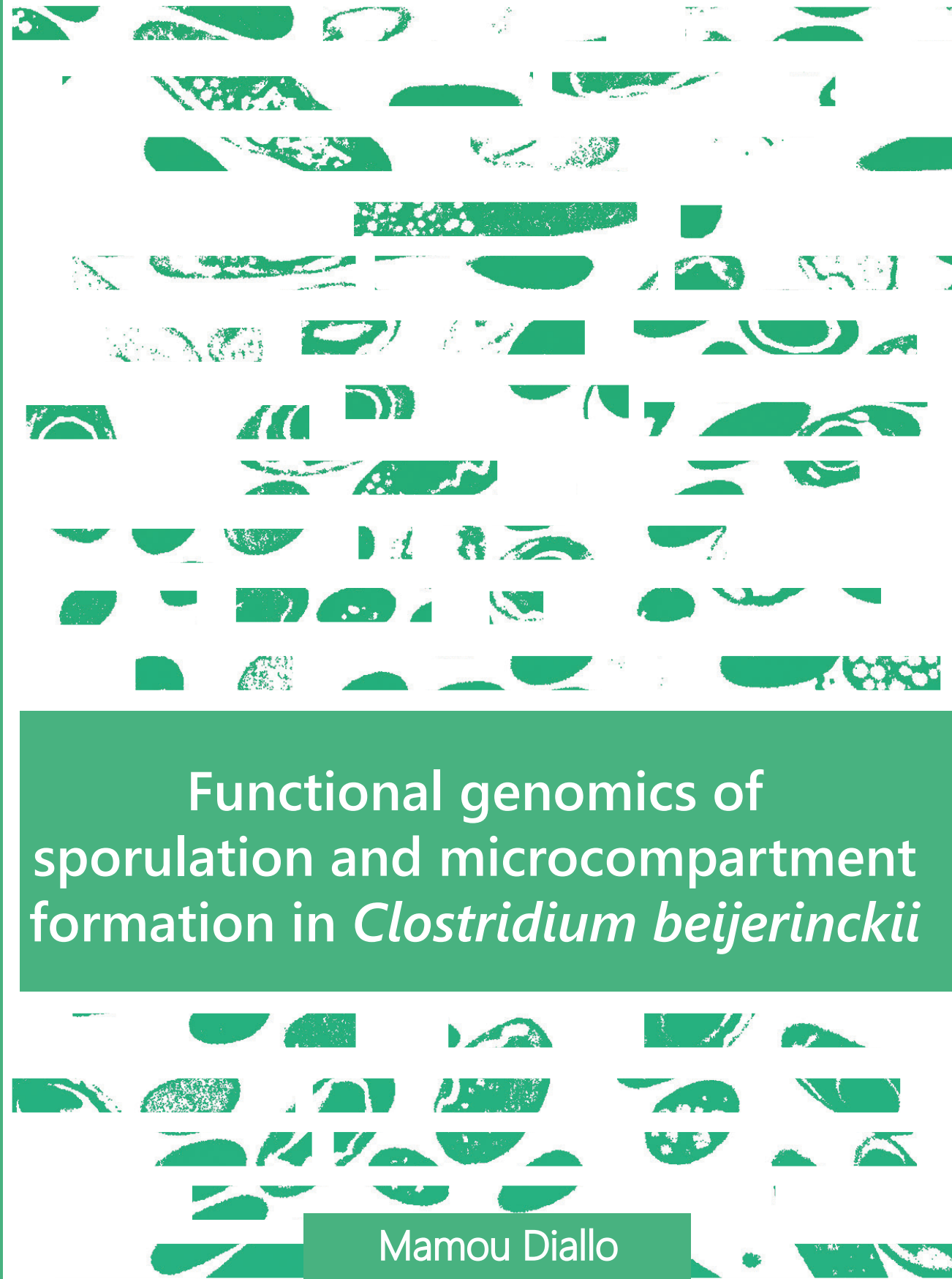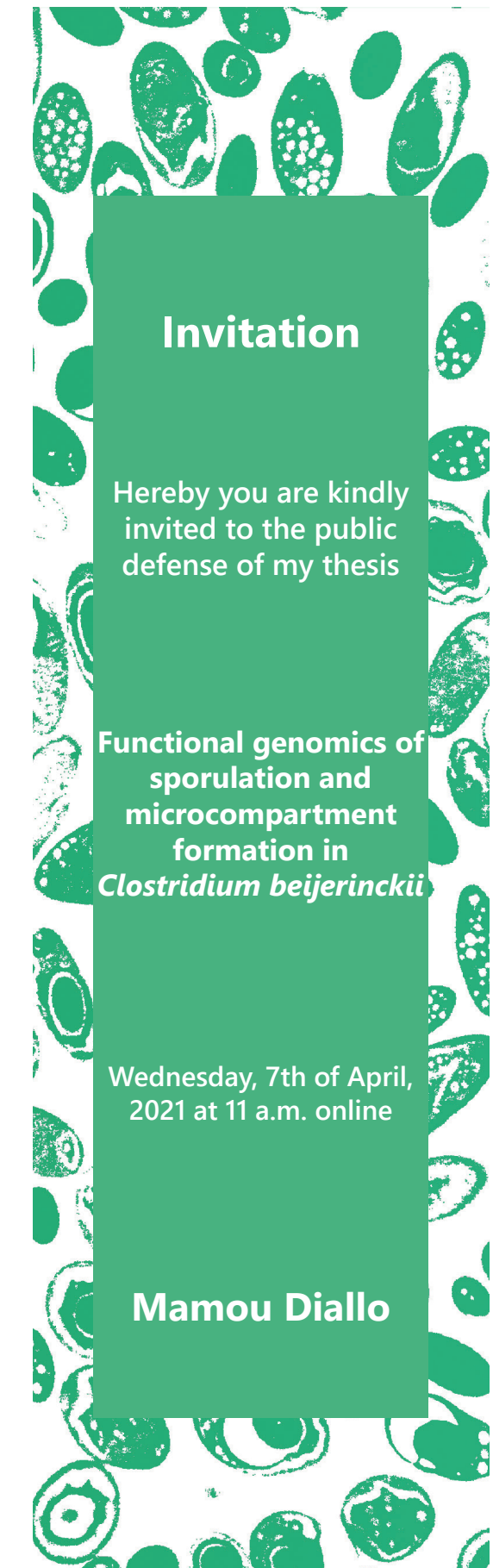
Mamou Diallo

Propositions

1. Sporulation is a complex cell differentiation process that involves more than 120 genes in *Clostridium beijerinckii*.
   (this thesis)

2. Analyzing omic data is like walking into a maze.
   (this thesis)

3. Failed experiments also deserve to be published.

4. Machine learning is not the only way to address information overload.

5. Like scientific publications, outreach should be a requirement for career advancement.

6. Waste does not exist since everything can be transformed.

7. The road towards a sustainable future requires daily sacrifices.

Propositions belonging to the thesis, entitled:

"Functional genomics of sporulation and microcompartment formation in *Clostridium beijerinckii*"

Mamou Diallo

Wageningen, 7 April 2021

Functional genomics of sporulation and microcompartment formation in *Clostridium beijerinckii*

Mamou Diallo

Thesis committee

Promotor

Prof. Dr J. van der Oost
Professor of Microbial Genetics
Wageningen University & Research

Co-promotors

Dr A. M. López-Contreras
Senior scientist at the Wageningen Food and Biobased Research Institute
Wageningen University & Research

Dr S. W. M. Kengen
Assistant professor at the Laboratory of Microbiology
Wageningen University & Research

Other members

Prof. Dr P. Patakova, University of Chemistry and Technology, Prague, CZ
Prof. Dr E. J. Smid, Wageningen University & Research
Prof. Dr S. Brul, University of Amsterdam
Dr W. K. Smits, Leiden University Medical Center

Functional genomics of sporulation and microcompartment formation in *Clostridium beijerinckii*

Mamou Diallo

Thesis

À mes fé mogos sur terre et dans l'au-delà

Contents

Chapter 1-

Sporulation in solventogenic clostridia

Mamou Diallo[1,2], Servé W. M. Kengen[2], Ana M. López-Contreras[1]

[1] Wageningen Food and Biobased Research, Wageningen, The Netherlands

[2] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

1

Keywords: Butanol, Sporulation, *Clostridium*, ABE fermentation, strain engineering, quorum-sensing

Abbreviations: ABE, Acetone Butanol Ethanol; **ACE,** Allelic Coupled Exchange **CRISPR**, Clustered Regularly Interspaced Short Palindromic Repeats; **Cas**, CRISPR-associated protein; **DPA**, dipicolinic acid

Abstract

Solventogenic clostridia are compelling organisms for biotechnological production processes, as they can utilize a wide range of carbon sources to produce commercially valuable carboxylic acids, alcohols and ketones by fermentation. Despite their potential, their conversion of carbohydrates to solvents is not cost-effective enough to result in economically viable processes. Engineering solventogenic clostridia by impairing sporulation is one of the investigated approaches to improve solvent productivity. Sporulation is a cell differentiation process triggered in bacteria in response to exposure to environmental stressors. The generated spores are metabolically inactive but resistant to harsh conditions (UV, chemicals, heat, oxygen). In Firmicutes, sporulation has been mainly studied in bacilli and pathogenic clostridia, and our knowledge of sporulation in solvent-producing clostridia is limited. Still, sporulation is an integral part of the cellular physiology of clostridia; thus, understanding the regulation of sporulation and its connection to solvent production may give clues to improve the performance of the solventogenic clostridia. This review aims to provide an overview of research done on the triggers and the regulation mechanism of sporulation in solventogenic clostridia. The potential applications of spores for process improvement are discussed.

Potential of clostridia for biochemical production

As a growing part of the world's population is getting access to affordable energy, the global energy demand is increasing drastically. The energy consumed mainly originates from fossil resources, resulting in an acceleration of the depletion of natural resources and increased greenhouse gas (GHG) emissions. To inverse this trend, our societies are transitioning towards more sustainable economies, and countries worldwide are promoting the renewable energy sector. In that aspect, European countries set a common energy policy framework aiming for a carbon-neutral economy by 2055, which means that the European Union's greenhouse gas emissions have to decrease by 55% by 2030, and the share of renewable energy to rise from 20% in 2020 to 40% by 2030 (1) to reach this goal. While substitutions to fossil energy generation processes such as hydrothermal, geothermal, solar, or wind energy are promoted, few alternatives to oil for freight, aviation, or the petrochemical sector are cost-effective.

One of the most promising substitutes to oil is "advanced biofuels", which are defined by the International Energy Agency (IEA) as liquid or gaseous fuels derived from lignocellulosic (second-generation biofuel) or algal biomass (third-generation biofuel). These feedstocks have a relatively diverse composition, and few microorganisms can grow on such a wide range of carbon sources. Moreover, these carbohydrates are available as complex polymeric structures (lignocellulose) that require dedicated enzyme systems to release the fermentable sugars. Some bacterial species from the *Clostridium* genus can hydrolyze these polymers, and an even larger group can ferment and the carbohydrates to produce solvents. Clostridia are anaerobic and spore-forming Gram-positive bacteria, and the entire *Clostridium* genus comprises over 270 species, including pathogenic, probiotic, thermophilic and benign soil bacteria. Several non-pathogenic clostridia have been studied for the production of advanced fuels and other biochemicals (2). These species are commonly divided into acid-producing species, solvent-producing species, cellulolytic species and acetogens (3). These clostridia convert simple and complex carbon sources, from C1 compounds to cellulose, into a diverse range of metabolites, ranging from carboxylic acids such as acrylate to solvents like propanol.

The solventogenic clostridia cluster consists of ten clostridia species, with five species being the most studied: *C. acetobutylicum, C. beijerinckii, C. pasteurianum, C. saccharobutylicum, C. saccharoperbutylacetonicum*. These species have been used industrially during the 20th century for acetone production through the ABE fermentation process (4–6), and their strains have been well preserved in strain collections. Other solventogenic strains with high solvent productivity were isolated recently (7). Despite many efforts, bioprocesses relying on these bacteria are not cost-effective due to high feedstock costs, expensive pretreatment of the feedstock, poor substrate use, and low solvent productivity (8, 9). These issues need to be tackled in order to enable competitive

biofuel prices. According to estimates, the product titer in a bioprocess aiming for the biofuel market needs to reach at least 50 g.L^{-1}, and the productivity should amount to 3 g.L^{-1}.h^{-1} to be commercially viable (10). While solvent productivity up to 10 g.L^{-1}.h^{-1} was reached with clostridial fermentation, the clostridial butanol titer in batch fermentation and solvent tolerance did not exceed 20 g.L^{-1} (11) and 4 % (v/v) (12), respectively. It appears that both solvent toxicity and sporulation prevent the production of higher solvent titers (13, 14). During the fermentation, bacteria are grown under stringent conditions necessary for solvent production yet harsh for the cells. As a defense mechanism, the bacteria differentiate into highly resistant cells called endospores (henceforth designated as spores) while producing solvents. Spores are metabolically inactive, and their formation requires metabolic energy and impairs solvent productivity (2). Several efforts were made to engineer asporogenous solvent-producing strains, to prevent these undesirable effects (15–17). The regulation of sporulation and solventogenesis appears to be coupled, but the underlying regulatory networks remain unclear (18). Due to the lack of efficient engineering tools for *Clostridium*, studies on the regulation of sporulation have been scarce. The recent development of markerless tools for clostridia (19) has made the genetic engineering of clostridia much more attainable. As a result, the number of engineered clostridia increased substantially, and together with the rise of omics studies, the current knowledge on sporulation in solventogenic clostridia has expanded considerably.

Sporulation regulation in solventogenic clostridia

In response to changes in the environment, some bacteria produce spores to survive under unfavorable conditions. Depending on the formation mechanism and the structure, different spore types can be found in the environment (20). The spores, formed by Firmicutes, called endospores (21, 22), are the most resilient. Endospores can survive harsh treatments such as high temperatures, the presence of oxygen (for anaerobic bacteria), desiccation, lysozyme incubation, ionizing radiation, and chemical solvents. The most studied sporulating bacteria belong to the *Bacillus* and *Clostridium* genus. Sporulation was first described in *Bacillus subtilis,* the model organism amongst spore formers, and the main features of its sporulation process are conserved in the *Clostridium* genus. Still, substantial differences in the spore morphology and sporulation initiation have been demonstrated between the two genera and within the *Clostridium* genus (21, 23).

A sporulation model for the solventogenic clostridia (23) has been developed thanks to studies in *C. acetobutylicum* ATCC 824. Few studies were done on sporulation in other solventogenic clostridia strains to confirm the universality of this model. Although solventogenic clostridia are often presented as a homogenous group of bacteria, based on the first phylogenic studies on the *Clostridium* genus (24, 25), this is not the case. Several strains were renamed and reclassified since 2000 (26), and recent phylogenic studies (27,

28) show that *C. beijerinckii* and *C. acetobutylicum* even belong to two different clades. Out of the seventeen clades dividing the *Clostridium* genus "sensu stricto", solventogenic clostridia can be found in two groups, one harboring *C. acetobutylicum* and *C. pasteurianum* and another consisting of *C. beijerinckii*, *C. saccharoperbutylacetonicum* and *C. saccharobutylicum*. *C. beijerinckii* is closer to the human pathogens *C. perfringens* and *C. botulinum* E than to the model solventogenic clostridia *C. acetobutylicum*, as depicted in Figure 1. In line with what has been suggested for toxin genes (28), solventogenesis genes might have been acquired by horizontal transfer. The localization of the sol operon on a megaplasmid in *C. acetobutylicum*, harboring the essential solvent genes, supports this hypothesis. Thus, the regulation mechanisms described for *C. acetobutylicum* might not be identical in *C. beijerinckii* or other solventogenic species (18).
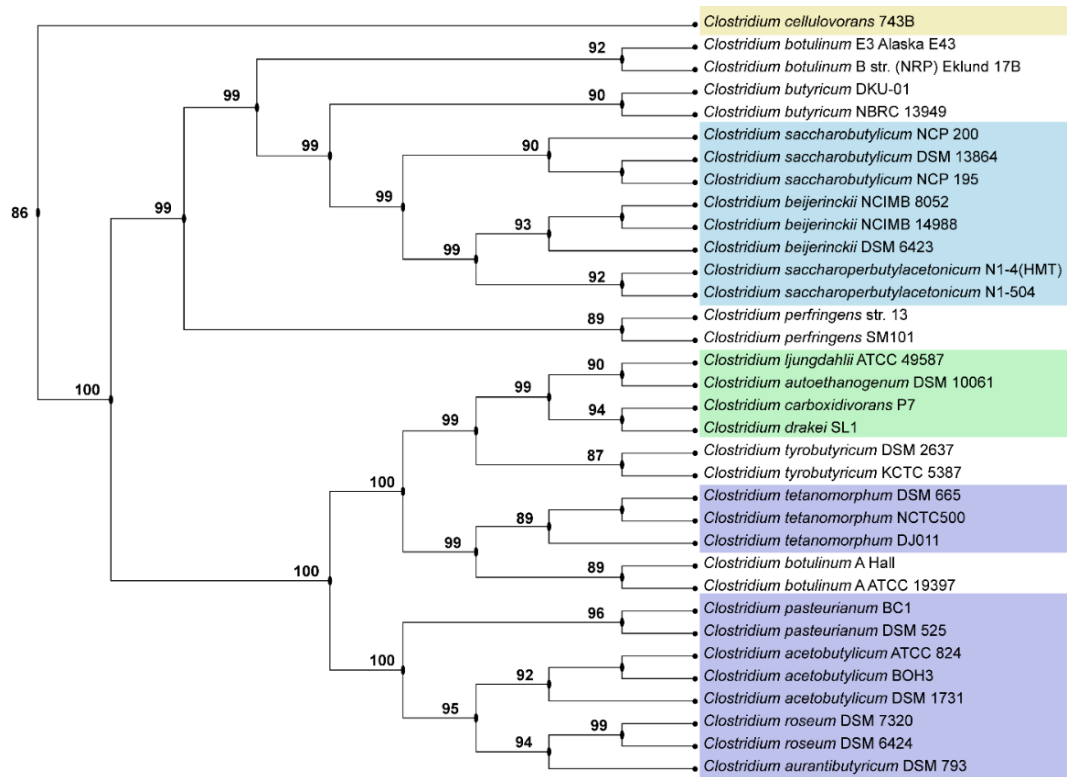

Figure 1: Cladogram of the *Clostridium* genus generated with the MicroScope platform (29) **with the Mash algorithm** (30) **relying on the neighbor-joining method.** The strains in the blue and purple boxes belong to the solventogenic group, the acetogenic clostridia are in the green box and the cellulolytic strains in the yellow box.

Sporulation triggers

The sporulation is known to be a response to stressful conditions. The primary triggers of sporulation in *Bacillus* species have been shown to be nutrient starvation and high cell density (31, 32). In contrast to bacilli, where the cessation of growth occurs due to a lack of nutrients, in *Clostridium*, an excess of a carbon or nitrogen source led to growth cessation (3, 33, 34). Furthermore, unlike bacilli, clostridia are anaerobic bacteria; thus, oxygen is a significant stress factor for the cells and triggers sporulation.

Several studies (32, 33, 35, 36) have reported the substantial impact of medium composition on sporulation initiation and sporulation efficiency (number of spore generated and the spores that can germinate) and suggested a link between carbon source, mineral content, external pH and the number of spores produced.

The carbon source

Solventogenic clostridia can ferment several carbohydrates (including C6 and C5 sugars and sugar polymers such as starch or xylan), yet, the solvent yield varies depending on the carbon source (32, 37). Similarly, sporulation efficiency depends on the carbon source, as described for *C. saccharobutylicum* NCP 262 (35). A complementary study with the same strain assessed the effect of 13 different carbon sources (carboxymethylcellulose, xylan, inulin, a starch/glucose mix, lactose, cellobiose, sucrose, maltose, glucose, mannose, fructose, galactose and xylose) on both sporulation and solvent formation (32). Depending on the carbohydrate, sporulation frequency varied up to 44%, with glucose utilization leading to the most spores. Likewise, when grown on rhamnose, a decrease in spore formation was observed in *C. beijerinckii* cultures (38). The substrate may also accelerate sporulation initiation. For instance, xylose-fed cultures of *C. acetobutylicum* BOH3 sporulated earlier than glucose-fed cultures (31). In addition to the nature of the carbohydrate, its concentration also affects sporulation. High glucose concentrations doubled the number of endospores generated by *C. saccharobutylicum* NCP 262 (39).

These observations have been confirmed by studies on carbon metabolite repression in clostridia. The carbon catabolite repression protein A (CcpA) regulates the carbon catabolite repression in Firmicutes and is involved in triggering sporulation in pathogenic clostridia. CcpA activates (40) or represses sporulation (41) depending on the species. In solventogenic clostridia, its role was studied only in *C. acetobutylicum* (42). In a CcpA deficient strain, sporulation was delayed, and the sporulation efficiency decreased (Table 1). This observation was complemented by a transcriptome analysis, which indicated a down-regulation of several sporulation-related genes in the mutant strain. According to these results, CcpA positively regulates sporulation in *C. acetobutylicum*.

Other media components

Solventogenic clostridia are increasingly investigated for their potential to produce solvents from complex feedstock such a lignocellulosic and algal feedstocks. Pretreatment of these feedstocks is necessary to the utilization by the bacteria of the carbohydrates present. During the pretreatment, di- and monosaccharides, as well as inhibitory chemicals (salts, furfurans and phenolic compounds), are formed. Studies have been conducted to evaluate the impact of these toxic compounds on cell growth and solvent formation. Hence, few reports on their effect on sporulation can be found; still, three transcriptomic studies of *C. beijerinckii* and *C. acetobutylicum* cultures exposed to phenolic compounds detected changes in the expression of sporulation genes. Exposure to ferulic acid (43) and syringaldehyde (44) caused an up-regulation of the late-stage sporulation genes *C. beijerinckii*. In *C.acetobutylicum*, a recent study showed through a gene coexpression network analysis (45) that exposure to vanillin and p-coumaric acid disturbed the transcription of early sporulation genes (*spo0A, spoIIE, spoIIP*) and sporulation specific sigma factors.

Acids, various metals and minerals, vitamins and amino acids also affect both solvent production and sporulation in clostridia. Acids, metal and mineral ions and amino acid content were shown to impact solventogenesis (46–50), but few studies mentioned their impact on sporulation. Long et al. (39) investigated butyrate and acetate's effect on sporulation in *C. saccharobutylicum* by adding them to the medium at the start of the fermentation in different concentrations. Even though the addition of acids was not necessary for sporulation, it increased the number of spores present in the culture by 40% to 100% for concentration below 4 $g.L^{-1}$.

As for the impact of other media components on sporulation, one study reports that the addition of adenine in the media caused a 20-h delay in the onset of sporulation in *C. saccharoperbutylacetonicum* cultures (51). It is worth noting that, depending on the species, a compound can have an opposite impact on sporulation. For example, in *C. perfringens*, iron is necessary for sporulation (52), but its addition in the medium impairs sporulation in *C. sporogenes* (53).

Solventogenic clostridia harbor sporulation proteins requiring metal-containing cofactors; thus, the media's metal content is expected to impact sporulation regulation. For instance, homologs of CsfB, an anti-sigma factor of σ^E and σ^G and SpoIIE, a phosphatase involved in sporulation regulation were identified in *Clostridium*. In *Bacillus*, their activity requires Zn^{2+} and Mn^{2+}, respectively (51, 52). Ca^{2+} is also involved in spore formation since it forms together with dipicolinic acid (DPA) Ca^{2+}-DPA, a crucial component for spore heat resistance (53, 56–58).

Studies of the transcriptome of wild-type and mutant *C. beijerinckii* cultures during fermentation indicated changes in the expression of genes involved in ion- and amino acid transport at sporulation initiation. In *C. beijerinckii* NRRL B598, sporulation initiation was concomitant with an up-regulation of the genes encoding a magnesium transporter, and an upregulation of genes encoding potassium, sodium and iron transporters was detected during stationary phase (59). Conversely, in cultures of the asporogenous *C. beijerinckii* Δ*spoIIE* strain, the genes encoding iron transporters were down-regulated during stationary phase (60), indicating a potential role of iron in sporulation.

Metabolite concentration

Metabolite stress has been suggested to trigger sporulation in solventogenic clostridia (61). When grown in chemostats, solventogenic clostridia ferment the available carbohydrates into carboxylic acids, mainly acetate and butyrate (Figure 2). Their accumulation in the culture causes a drop in pH.
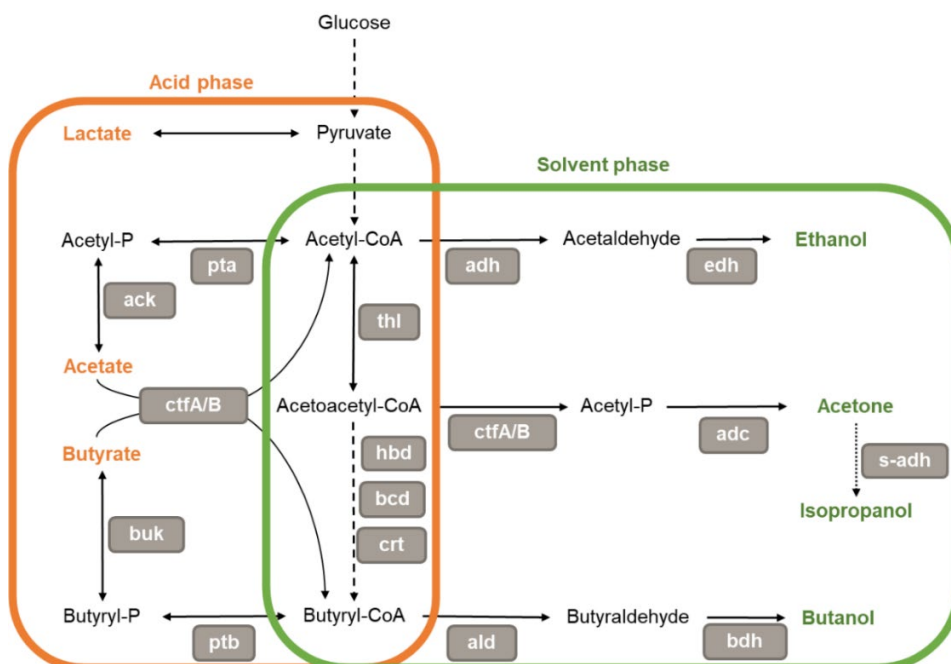
Figure 2: Acetone-Butanol-Ethanol metabolic pathway in solventogenic clostridia. Some strains harbor a secondary alcohol dehydrogenase (s-adh) that enables the formation of isopropanol. During exponential growth, the substrate is metabolized to form lactate, acetate and butyrate during the acid phase. Following the transition to the stationary phase, the acids are reassimilated, and the culture produces ethanol, acetone (or isopropanol) and butanol. The enzymes involved in the metabolic pathway are in grey boxes: pta, phosphotransacetylase; ack, acetate kinase; thl, thiolase; hbd, 3-hydroxybutyryl-CoA-dehydrogenase; crt, crotonase; bcd, butyryl-CoA-dehydrogenase; ctfA/B, CoA-transferase; buk, butyrate kinase; ptb, phosphotransbutyrylase; adh, aldehyde/alcohol dehydrogenase; edh, ethanol dehydrogenase; adc, acetoacetate decarboxylase; s-adh, secondary alcohol dehydrogenase; ald, butyraldehyde dehydrogenase; bdh, butanol dehydrogenase and -P stands for phosphate.

At the onset of solventogenesis and sporulation, acids are reassimilated and converted into solvents, resulting in a rise of pH in the culture. While solvent formation enables short-term relief from the pH stress, sporulation is regarded as a long-term stress response mechanism, protecting the cells from metabolic stress and interrupting sugar degradation (62). Acetate and butyrate accumulation during exponential growth is proposed to trigger both solventogenesis and sporulation (63). In *C. acetobutylicum*, a peak in the intracellular undissociated acid concentration was observed at the start of the solventogenesis (64, 65). Characterization of *C. acetobutylicum* recombinants deficient in phosphotransbutyrylase (Ptb), butyrate kinase (Buk) and acetate kinase (Ack) (66–69) suggested that instead of the concentration of undissociated acids, the intracellular concentration of butyryl phosphate (BuP) might trigger for both sporulation and solventogenesis. BuP is an intermediary metabolite in the ABE metabolic pathway (Figure 2), formed during acidogenesis during butyrate formation. Studies evaluating the intracellular concentration of BuP during batch cultivation showed that BuP was indeed accumulated in the cell (70, 71). Two peaks in the cytoplasmic BuP concentration were detected inside the cells, one at the beginning of the cultivation and a second one coinciding with solventogenesis and sporulation initiation (70). It was suggested that BuP acted as a phosphate donor enabling the activation of Spo0A, the master regulator of sporulation and solventogenesis (70, 72).

Recent data by Xu et al. (71) present another post-translational regulation mechanism involving protein acylation. Protein acylation is a post-translational modification that consists of adding an acyl group to a lysine residue. This reaction is reversible and does not need the intervention of an enzyme. Acylation neutralizes the negative charge of the lysine residue, altering the protein structure and its interaction with other proteins, cofactors, or substrates (73). The study conducted in *C. acetobutylicum* looked at the dynamics of lysine acetylation and butyrylation (71). Several key proteins involved in metabolism and life cycle, such as Buk and Spo0A, were butyrylated during cultivation. Two butyrylation sites were detected in the Spo0A sequence, one close to the phosphorylated domain and another on the DNA binding domain. In fact, replacing the lysine residue (K217) located in the DNA binding domain with glutamine decreased the DNA binding affinity of Spo0A. While the wild-type Spo0A could bind to its binding motif in the promoter region of Spo0A, the mutated Spo0A could not. This result suggests that Spo0A butyrylation is necessary for Spo0A activity and its autoregulation.

Butanol has also been suspected of triggering sporulation (61). Even though solventogenic clostridia naturally produce butanol, it affects cell growth when its concentration exceeds 0,5% v/v in the culture (74) and becomes lethal, around 1,5% v/v (75). Butanol concentration being a stress factor for the cells, researchers supposed that a rise in butanol concentration would initiate sporulation before it reaches toxic concentration. However, a decrease in granulose and spore number were observed in butanol stressed *C. beijerinckii*

cultures (75). Transcriptional studies on butanol stressed cultures of *C. acetobutylicum* and *C. beijerinckii* showed no notable changes in the expression of the genes encoding the sporulation-specific sigma factors (*sigF*, *sigE*, *sigG*, *sigK*) (76, 77). In *C. acetobutylicum* ATCC 824, no butanol-dependent impact on sporulation efficiency was described. Instead, a decrease in the expression of genes encoding small acid-soluble proteins was observed (76, 78). These proteins protect the DNA present in the spores and are crucial for their heat-resistance (79).

Secondary metabolites have also been reported to promote sporulation in solventogenic clostridia. Two categories of secondary metabolite biosynthesis gene clusters were identified in solventogenic clostridia (80), polyketide- and ranthipeptide biosynthesis clusters. Polyketides have been studied in *C. saccharoperbutylacetonicum* and *C. acetobutylicum*. In *C. saccharoperbutylacetonicum,* polyketides involved in sporulation initiation, solvent formation and tolerance were detected (81, 82). In *C. acetobutylicum*, three polyketides were detected, and the structures of two of them, clostrienose and clostrienoic acid, were solved (83). In both species, the disruption of polyketide clusters decreased sporulation (Table 1). In *C. beijerinckii*, polyketides might also intervene in the regulation of sporulation; the interruption of sporulation in *C. beijerinckii ΔspoIIE* affected the expression of the polyketide gene cluster (60). Recently the role of ranthipeptides, secondary metabolites belonging to the ribosomally synthesized and post-translationally modified peptide (RiPP) superfamily, was studied in *C. beijerinckii* and *C. ljungdahlii* (84). In *C. beijerinckii*, the genes encoding the precursor peptide and the radical SAM protein were disrupted, and the impact on the transcriptome was evaluated by RNA sequencing. In the mutant strain, sporulation genes were strongly down-regulated, and the *agr* locus encoding the Agr quorum-sensing mechanism was up-regulated. Secondary metabolites seem to play an important role in the initiation of sporulation, even so the interactions between the polyketides and ranthipeptides with sporulation regulators remains to be investigated.

Cell density

As described for *B. subtilis* (85, 86), cell density regulates sporulation in solventogenic clostridia. Similarly, in continuous cultures of *Clostridium*, where the cells are kept at a specific dilution rate, a decrease in the number of sporulating cells is observed (87, 88). Thanks to quorum-sensing mechanisms, cells can monitor environmental changes such as cell density and launch their adaptation response when required. Two quorum-sensing mechanism superfamilies were described in Gram-positive bacteria (89): the membrane receptor family (TCS) and the cytoplasmic receptor family (RRNPP). Systems belonging to both families were found in solventogenic clostridia (Figure 3). Two TCS, an Agr system and the BtrK/BtrR system, were described in *C. acetobutylicum* ATCC 824 (90, 91). The Agr system only regulates both granulose formation and sporulation, while the BrtK/BtrR

system regulates the growth rate, the start of solventogenesis and butanol tolerance. No role in sporulation regulation was described; still, the overexpression of the *BtrK/BtrR* operon changed the expression of genes involved in sporulation initiation (*spo0J*, *spoIIE*, *spoIIR*) and spore maturation (*sigK*, *spoIVA*).

Cytoplasmic receptor-based systems, RRNPP quorum-sensing systems positively regulating sporulation were recently identified in *C. acetobutylicum* (92) and *C. saccharoperbutylacetonicum* (93). These systems named Qss are composed of two proteins: the receptor Qsr, which harbors a helix-turn-helix region (HTH) common to DNA binding domains, and the signal peptide precursor Qsp. In the genome of *C. acetobutylicum,* eight putative RRNPP systems were detected, but only two seem to intervene in sporulation regulation (Table 1). In *C. saccharoperbutylacetonicum*, four out of the five identified systems affected sporulation. The deletion of *qssR1* and *qssR2* increased sporulation efficiency, while sporulation frequency decreased in Δ*qssR3* and Δ*qssR5* mutants. Moreover, *spo0A* expression decreased in Δ*qssR3* and Δ*qssR5* mutants and *qssR1/2* affected the expression of *spo0E-like* genes*,* putatively involved in the activation of Spo0A. However, no regulation mechanism was described in any of these species. In *Clostridioides* (former *Clostridium*) *difficile,* a similar RRNPP system (RstA) was described. RstA was reported to modulate the expression of sporulation genes through its DNA binding domain (94). A similar regulation mechanism might take place in solventogenic clostridia. Changes in process parameters such as temperature or dilution rate also appear to trigger sporulation (51, 88). Sporulation is initiated by a considerable number of triggers (Figure 3), and several quorum-sensing mechanisms and messengers interact with transcriptional regulators to integrate all the environmental cues and induce a cellular response. Similarly, the molecular regulation of sporulation is complex and composed of several layers of sporulation inhibitors and activators.

1
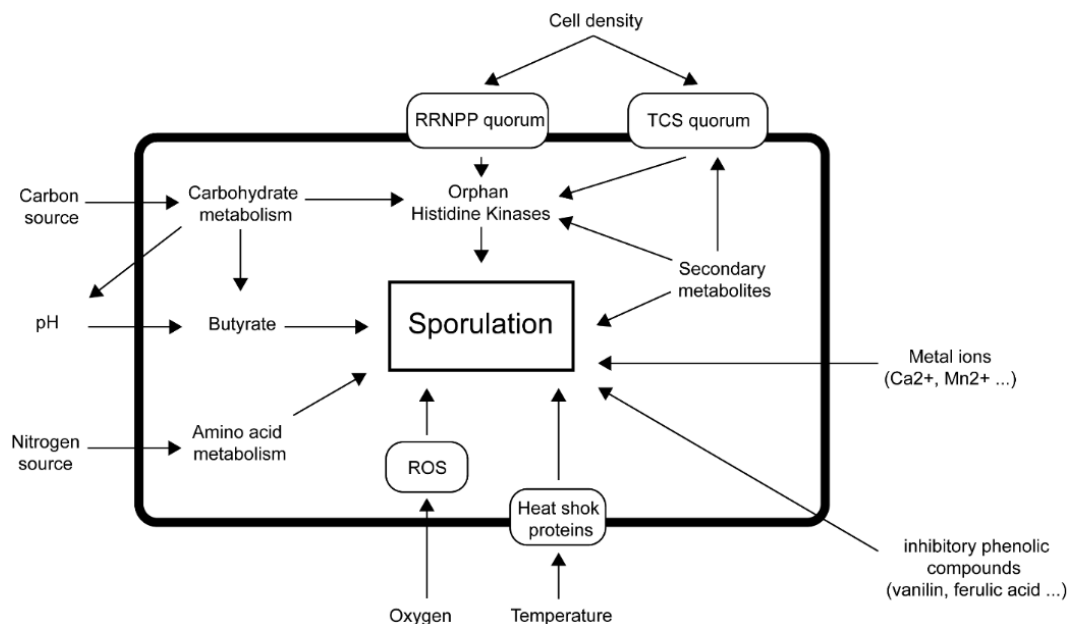
Figure 3: Triggers of sporulation in solventogenic clostridia and the signal transduction systems involved in sporulation regulation; TCS, two-component system family; RRNPP, Rgg/Rap/ NprR/PlcR/PrgX quorum system family; ROS, reactive oxygen species-dependent mechanisms.

The sporulation cycle

The initiation of sporulation

The sporulation regulation network described in *Bacillus* has been used as a template for understanding the sporulation cascade in *Clostridium* (95–97). However, several sporulation genes identified in *Bacillus* are absent in *Clostridium*, indicating a difference in the molecular regulation. Genome comparison studies were conducted to identify homologs to the sporulation genes studied in *B. subtilis* and the necessary set of genes required for sporulation (98–100). Fifty-two genes were identified as essential for sporulation as they were found both in sporulating clostridia and bacilli by comparing more than 217 genomes of sporulating and non-sporulating Firmicutes. Even when gene homologs are found in *Clostridium*, their functions are not always identical to their homologs in *Bacillus* (23, 101). Genome comparison, coupled with transcriptomic studies, enabled the identification of additional sporulation genes in solventogenic clostridia (43, 74, 87, 102). The role of the central sporulation regulators and sigma factors were studied in solventogenic clostridia through the generation of asporogenous mutants, mainly in *C. acetobutylicum* ATCC 824 and more recently in other solventogenic clostridia (Table 1).

Table 1: Mutant strains generated from solventogenic clostridia with changes in sporulation compared to the wild-type. When a gene is disrupted or its expression is reduced, a '–'is added; when the gene is overexpressed a '+' is added next to the gene abbreviation. NA stands for 'data not available'.

	Parent strain	Genotype	Sporulation	ABE production	Generation method	Reference
C. acetobutylicum	ATCC 824	*spo0A⁻*	No sporulation	Important decrease	double-crossover chromosomal integration	(103)
	ATCC 824	*spo0A⁻*	Reduced sporulation	Important decrease	allelic coupled exchange (ACE)	(104)
	ATCC 824 Δ*spo0A*	*spo0A⁺*; native promoter	Early sporulation	NA	allelic coupled exchange (ACE)	(104)
	ATCC 824 Δ*spo0A*	*spo0A⁺*; ferredoxin promoter	Early sporulation	NA	allelic coupled exchange (ACE)	(104)
	ATCC 824	*spo0A⁻*	Reduced sporulation	Important decrease	CRISPR-Cas9i	(105)
	ATCC 824	*spo0A⁺*; native promoter + 43 nt from the upstream *spoIVB*	Early and increased sporulation	Similar to WT	plasmid overexpression	(103, 106)
	ATCC 824	*spo0A*-G179S	Reduced sporulation and no heat resistance	No production	Double-crossover allelic exchange	(107)
	ATCC 824	*lys⁻*	Reduced sporulation	Similar to WT	Clostron	(108)
	ATCC 824	*spoIIE⁻*	Reduced sporulation	Increase	antisense RNA	(15)
	ATCC 824	*spoIIE⁻*	No sporulation	Slight decrease (30%)	single-crossover chromosomal integration	(16)
	ATCC 824	*sigE⁻*	No sporulation	Similar but inoculum dependent	single-crossover chromosomal integration	(109)
	ATCC 824	*sigF⁻*	No sporulation	Similar but inoculum dependent	single-crossover chromosomal integration	(110)
	ATCC 824	*sigG⁻*	No sporulation	Similar to WT	single-crossover chromosomal integration	(109)

Parent strain	Genotype	Sporulation	ABE production	Generation method	Reference
ATCC 824	*sigK*⁻	No sporulation	Important decrease	double-crossover chromosomal integration	(17)
ATCC 824 ΔsigK	*sigK*⁻, *spo0A*⁺; *ptb* promoter	No sporulation	Similar to WT	double-crossover chromosomal integration	(17)
ATCC 824	*cac0437*⁻	Increased sporulation	NA	Clostron	(111)
ATCC 824	*cac0437*⁺	Reduced sporulation	NA	Clostron	(111)
ATCC 824	*cac0323*⁻	Reduced sporulation	NA	Clostron	(111)
ATCC 824	*cac0903*⁻	Reduced sporulation	NA	Clostron	(111)
ATCC 824	*cac3319*⁻	Reduced sporulation	NA	Clostron	(111)
ATCC 824	*cac3319*⁻	No sporulation	Increased production	Clostron	(112)
ATCC 824	*cac0323*⁻; *cac0903*⁻	No sporulation	NA	Clostron	(111)
ATCC 824	*cac0323*⁻; *cac3319*⁻	No sporulation	NA	Clostron	(111)
ATCC 824	*cac3319*⁻; *cac0437*⁻	Increased sporulation	NA	Clostron	(111)
ATCC 824	*cac0903*⁻; *cac0437*⁻	Increased sporulation	NA	Clostron	(111)
ATCC 824	*abrB310*⁻	Reduced sporulation	Important decrease	antisense RNA	(113)
ATCC 824	*agrA*⁻	Reduced sporulation	Similar to WT	Clostron	(90)
ATCC 824	*agrB*⁻	Reduced sporulation	Similar to WT	Clostron	(90)
ATCC 824	*agrC*⁻	Reduced sporulation	Similar to WT	Clostron	(90)
ATCC 824	*pks*⁻	No sporulation	Increased production	Double-crossover allelic exchange	(83)
ATCC 824	*qsrB*⁺	Increased sporulation	Important decrease	plasmid overexpression	(92)
ATCC 824	*qsrG*⁻	Decrease in heat resistant spores	Decreased production (about 30%)	Clostron	(92)
ATCC 824	*ccpA*⁻	Delayed and reduced	Similar with pH control	Targetron	(42)

C. acetobutylicum (row label spanning the left side of the table)

	Parent strain	Genotype	Sporulation	ABE production	Generation method	Reference
C. acetobutylicum	ATCC 824	*ptb*	No sporulation	Important decrease	antisense RNA	(66)
	ATCC 824	SNPs in 8 sporulation genes + SNPs in the promoter of 2 sporulation genes	No sporulation	Increased production	random mutagenesis and directed evolution	(114)
	ATCC 4259	Unknown	No sporulation	Increased production	random mutagenesis	(115)
	CGMCC 5234	*spo0A⁻*	Reduced sporulation	Important decrease	Clostron	(116)
C. beijerinckii	NCIMB 8052	*spo0A⁻*	No sporulation	Important decrease	chromosomal integration	(117, 118)
	NCIMB 8052	*spo0A⁻*	Reduced sporulation	Decreased production (about 15%)	CRISPR-Cas9i	(105)
	NRRL B-598	*spo0A⁺*; ferredoxin promoter	No sporulation	Important decrease	plasmid overexpression	(36)
	NCIMB 8052	*spoIIE⁻*	No sporulation	Increased production	CRISPR-Cas9	(60, 119)
	NCIMB 8052	Unknown	Reduced and delayed sporulation	Important decrease	random mutagenesis and directed evolution selection	(120)
	NCIMB 8052	Unknown (the authors suggest SNPs in *cbei_0769* and *cbei_4400*)	Delayed sporulation	Increased production	random mutagenesis and directed evolution	(121)
	SA-1	2 SNPs (0A box of *abrB* homolog *cbei_4895*)	No sporulation	Important decrease	random mutagenesis and directed evolution	(122)
	NCIMB 8052	Unknown	No sporulation	Important decrease	random mutagenesis	(123, 124)
	CC101	*cbei_2073⁻*	Decrease in heat resistant spores	Increased production	CRISPR-Cas9n	(125)

Parent strain	Genotype	Sporulation	ABE production	Generation method	Reference
C. beijerinckii					
CC101	*cbei_4484* -	Decrease in heat resistant spores	Increased production	CRISPR-Cas9n	(125)
NCIMB 8052	*sigL* -	Decrease in heat resistant spores	No production	CRISPR-Cas9	(126)
NRRL B-593	*sigL* -; 14 additional SNPs	Decrease in heat resistant spores	No production	random mutagenesis and directed evolution	(126)
C. pasteurianum					
ATCC 6013	*spo0A* -; 66 additional SNPs	No sporulation	Increased production	random mutagenesis and directed evolution	(127)
ATCC 6013	*spo0A*-	No sporulation	Increased production	double-crossover allelic exchange	(127)
DSM 525-H1	*spo0A*-	No sporulation	Increased production	allelic coupled exchange (ACE)	(128)
C. saccharo-butylicum					
NCP 262	*spo0A*-G172S	Reduced sporulation no heat resistance	No production	double-crossover allelic exchange	(107)
C. saccharoper-butylacetonicum					
N1-4(HMT)	*spo0A*-I261T	No sporulation	Similar to WT	Endogenous CRISPR-Cas	(129)
N1-4(HMT)	*qss1* -	Increased sporulation	Similar to WT with pH control	CRISPR-Cas9	(93)
N1-4(HMT)	*qss2* -	Increased sporulation	Similar to WT with pH control	CRISPR-Cas9	(93)
N1-4(HMT)	*qss3* -	No sporulation	Similar to WT with pH control	CRISPR-Cas9	(93)
N1-4(HMT)	*qss5* -	No sporulation	Similar to WT with pH control	CRISPR-Cas9	(93)

The regulation model of sporulation in *Clostridium* is divided, like in *Bacillus*, into seven stages associated with morphological changes of the cell. In most clostridia, the sporulation process coincides with granulose accumulation and ends with the lysis of the mother cell and the release of the spores in the environment (Figure 4).
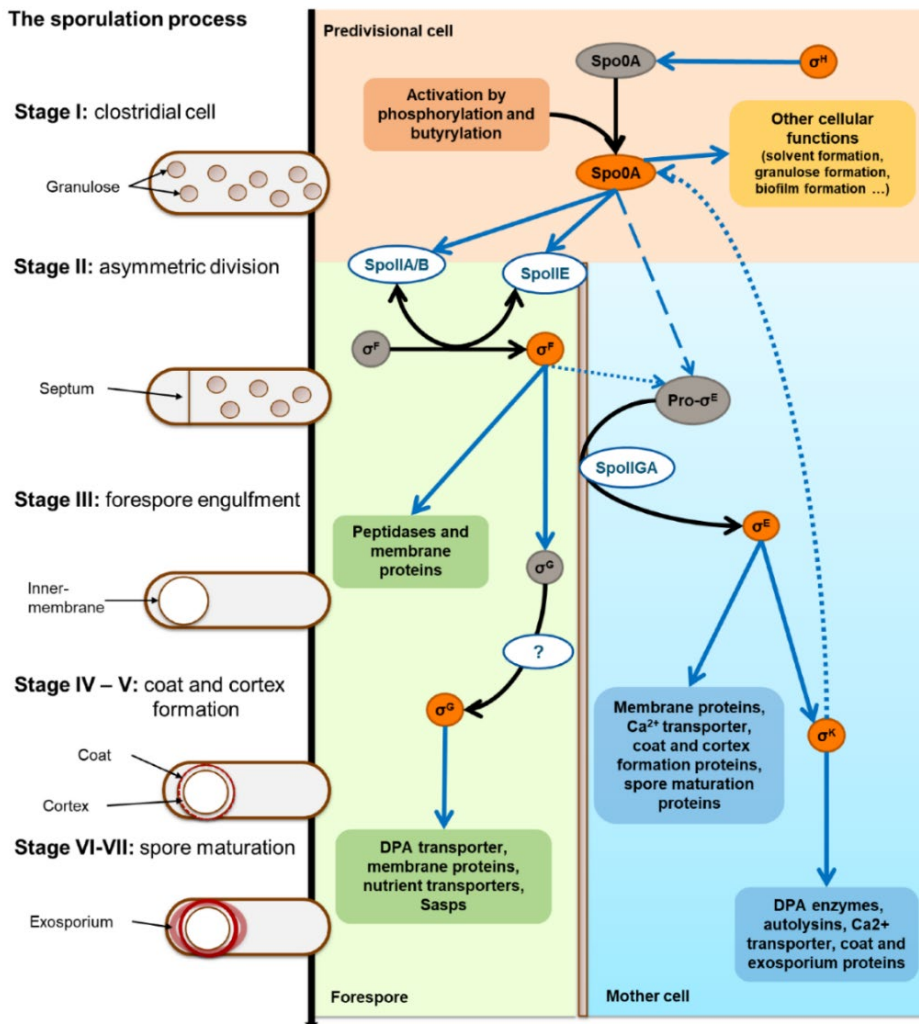
Figure 4: Morphological changes and molecular regulation of sporulation in *C. acetobutylicum* modified from (23)**.** The regulation of the sporulation process is mainly realized by the modulation of the transcription in each compartment. Post-translational regulation enables the activation of Spo0A and sporulation-specific transcription factors (σ^F, σ^G, σ^E). The activation mechanism of σ^G has not been investigated in *C. acetobutylicum*. Inactive transcriptional regulators are in grey, and active transcriptional regulators are in orange. DPA, dipicolinic acid; Sasps, small acid-soluble proteins. Black arrows indicate post-translational regulations, blue arrows transcriptional regulation. Arrows with short dashes indicate interactions described only in *C. acetobutylicum*, arrows with long dashes observed only in *C. beijerinckii*, and full arrows indicate interactions described in *C. acetobutylicum* and other clostridia.

Sporulation is initiated at the end of vegetative growth and is reflected at the transcriptomic level by an increase in *spo0A* expression. Spo0A, the general regulator of the transition from vegetative to stationary growth, is conserved in all Firmicutes and has a central role in sporulation, toxin and solvent production (21, 23, 118, 130).

In *Bacillus*, at the end of vegetative growth, AbrB lifts its repression on σ^H and σ^H promotes then the transcription of *spo0A* (95). In solventogenic clostridia, the role of σ^H in *spo0A* regulation is unknown. The gene encoding σ^H is constitutively expressed in clostridia, which implies that the mechanism described in *Bacillus* differs from the one occurring in *Clostridium*. Several AbrB homologs were identified in solventogenic clostridia, and in *C. acetobutylicum* ATCC 824, three homologs (*cac0310*, *cac1941* and *cac3647*) were identified, and each gene was disrupted to study its role in the regulation of cellular events. The disruption of the most expressed AbrB homolog, *cac0310*, delayed sporulation and impaired solvent production (113). The disruption of *cac3647* increased solvent production, while the solvent production of $\Delta cac1941$ cultures decreased by 6% compared to the wild-type (131). No change in the sporulation of the $\Delta cac1941$ or $\Delta cac3647$ mutant was reported. According to these results, AbrB homologs belong to the sporulation and solvent regulation network in solventogenic clostridia. Unlike in *Bacillus*, these results indicate that AbrB (Cac0310) may promote sporulation in *C. acetobutylicum*. The Spo0A DNA binding motif, called 0A box, was found upstream of *cac0310* and its homolog in *C. beijerinckii* NCIMB 8052, *cbei4885*, indicating that these *abrB* homologs belong to the Spo0A regulon. In fact, the disruption of this 0A box might impair solvent production and sporulation; one of the SNPs detected in a degenerated offspring of *C. beijerinckii* SA-1, was an SNP in the 0A box of *cbei4885* (122).

Mutations in the *spo0A* coding sequence of solventogenic clostridia have been shown to affect cell physiology considerably. Changes in growth, colony morphology, sporulation and solvent productivity have been reported. Numerous *spo0A* mutants have been characterized (Table 1), and all have impaired sporulation with sometimes a change in solvent productivity (103, 118, 129, 132). The consequences for the phenotype seem to depend on the mutation location and the studied species. The Spo0A sequence harbors various domains that are conserved among Firmicutes and which are putatively involved in sporulation regulation (Figure 5). Insertional mutations were shown to impair the protein function, blocking the central role of Spo0A in solventogenesis and sporulation regulation (103, 118, 133, 134). However, due to gene engineering techniques used, the impacts of polar effects on the phenotype cannot be excluded (135), but with the development of markerless gene engineering methods, precise mutations of *spo0A* were generated.
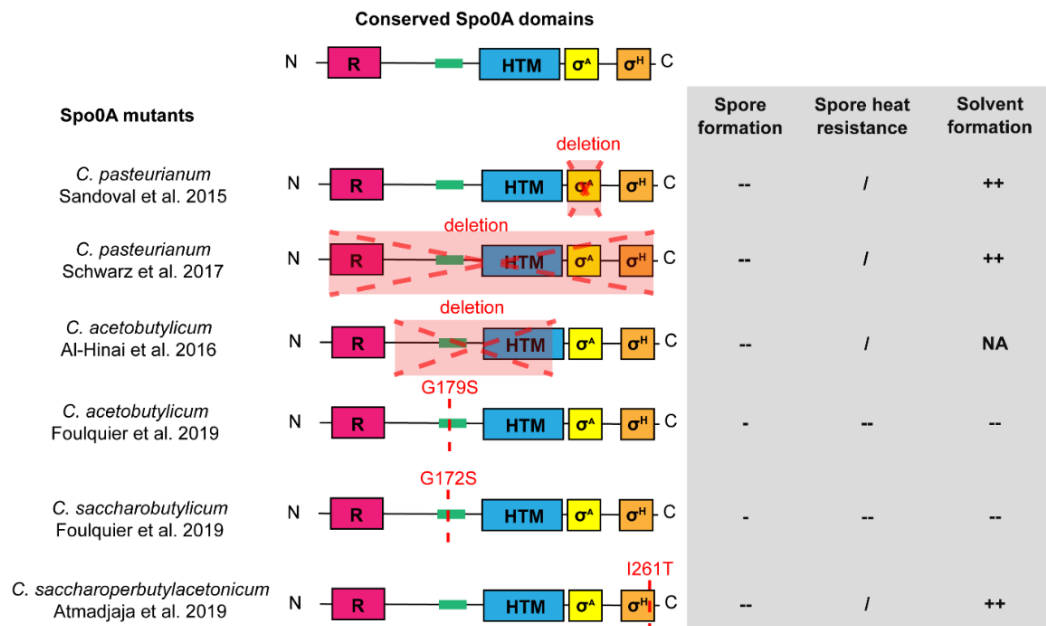
Figure 5: Mutations and associated phenotypes of markerless *spo0A-* strains. Five conserved regions are indicated by colored boxes; R: signal receiver domain, Green box: conserved region with no known function, HTM : helix turn motif, σ^A: putative σ^A activator region, σ^H: putative σ^H activator region. The ability to sporulate, to form heat resistant spores and to produce solvent is indicated next to the scheme of the *spo0A* mutation present in each mutant strain; - : indicate a decrease compared to the wild-type phenotype; -- : indicate the abolition of the feature in the mutant compared to the wild-type phenotype; / : indicate that this characteristic was not evaluated ; ++ indicate an increase compared to the wild-type phenotype and NA stands for no data available.

Mutations (deletions and single nucleotide modifications) in the putative σ factor activator domains did not cause a decrease in solvent production, as shown in Figure 5. Like in *Bacillus subtilis* (136), a mutation in the putative σ^H activator region disrupted the sporulation of *C. saccharoperbutylacetonicum* (I261T). In *Bacillus*, sporulation impairment was explained by a loss in the binding affinity of Spo0A for *spoIIE* and *spoIIA* promoters. Mutations in the region upstream from the DNA binding domain impaired solvent production and decreased sporulation efficiency in *C.acetobutylicum* and *C. saccharobutylicum* (107). In *C. pasteurianum*, a deletion in the σ^A activator region disrupted sporulation and increased solvent production (127). Surprisingly, in *C. pasteurianum*, the deletion of most of the gene (816 bp out of 822) led to the same phenotype (128). Spo0A might play a different role in sporulation regulation in *C. pasteurianum*.

Next to the integrity of the *spo0A* gene, also intracellular Spo0A levels appear to play a role in sporulation regulation. In the non-sporulating strain, *Clostridium* sp. MF28, relatively low *spo0A* expression levels were detected, compared to other solventogenic clostridia, even though high solvent titers could be reached (137). In *C. acetobutylicum*,

independently from the promoter region used for *spo0A* overexpression, increased sporulation was observed (67, 104, 138). In contrast, in *C. beijerinckii*, *spo0A* overexpression led to a decrease in both sporulation and solvent production (36). A slight change in Spo0A homeostasis might lead to a different regulation of sporulation and solventogenesis. No detailed study on the variation of active Spo0A and its impact on sporulation has been done in solventogenic clostridia to confirm this hypothesis. In *Bacillus,* depending on the intracellular concentration of phosphorylated Spo0A, differences in the expression of the Spo0A regulon were described (139, 140).

Once *spo0A* is transcribed and translated, Spo0A is activated by at least two post-translational modifications: butyrylation (see section on metabolite concentration) and phosphorylation, as illustrated in Figure 6. In contrast to the phosphorelay system described in *B. subtilis*, Spo0A activation seems to be done by orphan histidine kinases in clostridia (Figure 6). Orphan histidine kinases are part of two-component quorum-sensing systems (TCS). TCS usually consists of a sensor histidine kinase and a response regulator located in the same operon. However, isolated genes encoding sensor histidine kinases or response regulators were identified in the genome of numerous bacteria and especially in Firmicutes. Since these genes are not co-located with a gene encoding a specific sensor protein or response regulator, they are labeled as orphans (97, 141). Unlike regular TCS kinases, orphan histidine kinases do not only control one partner but can modulate several response regulators. In clostridia, orphan histidine kinases are integrated for signal transduction to the sporulation regulation network (111, 125, 142, 143). In *C. acetobutylicum*, the impact of four histidine kinases (Cac0323, Cac0437, Cac0903 and Cac3319) on sporulation has been studied (111). In vitro studies proved that two of these kinases, Cac0903 and Cac3319, are able to activate Spo0A. While the disruption of *cac0323*, *cac0903* and *cac3319* resulted in a decreased sporulation, *cac0437* deficient strains hyper-sporulated (earlier sporulation and a fifteen fold increase in heat resistant spores). In *C. beijerinckii* NCIM 8052, six homologs of these kinases were disrupted in a recent study (125). A significant decrease in sporulation efficiency (between 70 and 90%) coupled with an increase in solvent production of 38% and 14% were reported only for the Δ*cbei2073* and Δ*cbei4484* mutants, respectively. Interactions between these histidine kinases and Spo0A remain to be studied.
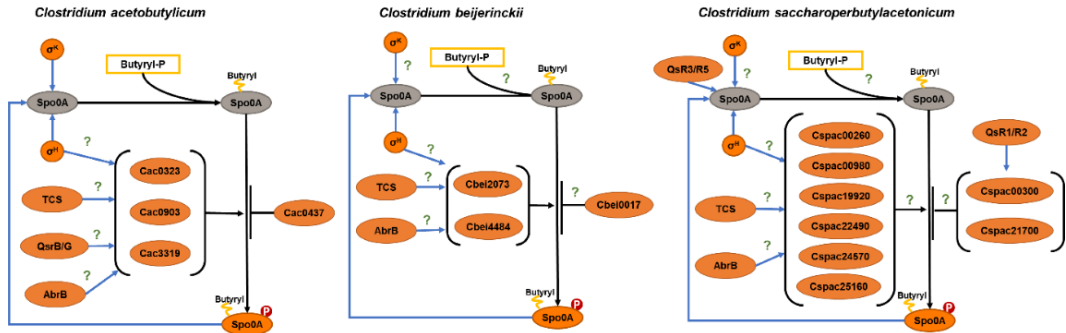
Figure 6: Transcriptional and post-translational regulation of Spo0A in *C. acetobutylicum*, *C. beijerinckii*, *C. saccharoperbutylacetonicum*. This figure was adapted from (23) with results from (71, 91–93, 125) Blue arrows transcriptional regulation and black arrows post-translational regulation. The interrogation marks indicate interaction that has to be proven experimentally.

Once phosphorylated, Spo0A binds to the 0A box to regulate the expression of sporulation- and solvent genes (118, 144). Differences in the motif sequence (118), the number (145) and the location of these 0A boxes (18) between solventogenic clostridia were detected. These discrepancies further illustrate the variations in the role of Spo0A in solventogenic clostridia.

Stages of sporulation

Stage I of the sporulation process starts with the DNA replication and the positioning of Z-rings, close to the poles, to prepare for asymmetric division (146). Solvent production is initiated, and in most clostridia, Stage I coincides with a morphological change. The cell swells due to the accumulation of a starch-like polymer called granulose. In *C. acetobutylicum*, granulose and sporulation are regulated by an Agr quorum system (90). No study investigating the link between granulose production and sporulation was done in other solventogenic clostridia. Moreover, granulose accumulation was not described in all solventogenic clostridia; *C. tetanomorphum*, for example, does not produce any granulose during the sporulation process (147).

During Stage II of the sporulation process, a septum forms on one pole of the cell dividing it into two compartments, the forespore and the mother cell. In *Bacillus*, this asymmetric division is orchestrated by SpoIIE and the cell division proteins (involved in binary fission) (146). SpoIIE simultaneously activates σ^F, the first sporulation-specific sigma factor, which is kept inactive by SpoIIAB. SpoIIE phosphorylates SpoIIAA, which binds the anti-sigma factor SpoIIAB which then releases σ^F in the forespore. In solventogenic clostridia, the role of SpoIIE was studied in *C. acetobutylicum* and *C. beijerinckii* through the generation of SpoIIE deficient mutants (15, 16, 60). Both mutants could no longer sporulate, but phenotypical differences between *C. beijerinckii ΔspoIIE* and *C. acetobutylicum ΔspoIIE*

were noted. In *C. acetobutylicum*, SpoIIE disruption prevented the formation of an asymmetric septum, while in *C. beijerinckii,* misplaced septa were observed. The morphology of the *C. beijerinckii* mutant corresponded to the *spoIIE* mutants described for *Bacillus*. This discordance in the mutants' morphology indicates differences in the asymmetric septation mechanism of solventogenic clostridia. Another difference was observed; no critical change in the expression of *sigF* and *sigE* was detected in *C. beijerinckii ΔspoIIE,* contrasting with *C. acetobutylicum ΔspoIIE,* where *sigF* and *sigE* were down-regulated.

The completion of asymmetric division marks the entrance in Stage III, during which the septum is prolonged and surrounds the whole forespore. This phenomenon, called engulfment, yields an isolated compartment surrounded by two membranes within the mother cell. Engulfment is coordinated by proteins belonging to σ^F (in the forespore) and σ^E regulons (in the mother cell). The roles of σ^F and σ^E were studied only in *C. acetobutylicum* ATCC 824, where *sigF* and *sigE* mutants were generated. Once again, these mutants' cell morphology did not correspond to the morphology of *sigF* and *sigE* mutants generated in *Bacillus* or other clostridia (23). In both *C. acetobutylicum* mutants, the sporulation process was interrupted before the formation of the asymmetric septum, as observed for the *spoIIE* mutant, which would indicate an earlier function of σ^F and σ^E in sporulation regulation (109, 110). The disruption of either *sigF* or *sigE* affected solventogenesis. Indeed, when mid to late exponential cells were inoculated for fermentation, solvent production decreased significantly.

After engulfment, the coat and the spore cortex are formed during Stage IV and V. Calcium dipicolinic acid (Ca^{2+}-DPA) is produced in the mother cell. DPA, formed by the conversion of aspartate, binds Ca^{2+} in the forespore, and the complex is then transported into the forespore in exchange for water (95). Ca^{2+}-DPA binds to the forespore's DNA to protect it against heat damages (57, 58, 148). In the meantime, coat proteins assemble around the membrane of the forespore (149). These events are coordinated by proteins regulated by σ^G in the prespore and σ^K in the mother cell. In contrast to pathogenic clostridia, these late stages of sporulation are barely studied in solventogenic clostridia. Nonetheless, σ^K and σ^G deficient mutants in *C. acetobutylicum* confirmed their crucial role in sporulation (17, 109). Disruption of *sigG* interrupted sporulation after engulfment as described in *Bacillus* and did not affect solventogenesis. The disruption of *sigK*, though, did not yield the same phenotype as the *Bacillus* mutant. In *C. acetobutylicum,* σ^K regulates sporulation initiation and sporulation maturation, as described in other clostridia (23). No Spo0A proteins were detected in *sigK* mutants of *C. acetobutylicum*, and the introduction of an extrachromosomal copy of *spo0A* under control of the *ptb* promoter leads to the formation of heat-sensitive spores.

During Stages VI and VII, the spore matures as the size of both cortex and coat increases. Sporulation finishes with the lysis of the mother cell and the release of the spore in the environment. A study on an autolysin deficient mutant, *C. acetobutylicum lyc::int* (72), generated in *C. acetobutylicum* ATCC 824, showed that autolysins are needed to complete sporulation (108). The number of viable spores produced by *C. acetobutylicum lyc::int* (72) decreased by 30% compared to the wild-type strain. The authors hypothesized that cell lysis provides additional nutrients to sporulating cells.

Only a few studies were done about the molecular regulation of sporulation in solventogenic clostridia (15, 17, 90, 109–111, 113, 118, 119, 130). These studies were mainly realized in *C. acetobutylicum* ATCC 824, and studies with other solventogenic strains show variations in the role of Spo0A and SpoIIE in the regulation of sporulation (60, 127). Moreover, the post-translational regulation for sigma factor activation, observed in *Bacillus* and *C. difficile*, still needs to be investigated in solventogenic clostridia. Thus, several grey areas remain concerning the regulatory mechanisms controlling the sporulation cascade in solventogenic clostridia.

The clostridial endospore

At the end of sporulation, an endospore is released into the environment. The endospore is highly dehydrated and organized in proteinous layers protecting the core, which hosts the DNA bound to small acid-soluble proteins (Sasps), ribosomes, enzymes and DPA. The DPA content can reach up to 25% of the spore's dry weight (142). The core is surrounded by several layers: the inner membrane, the germ cell wall, the cortex, the outer membrane and the coat (Figure 7-C). These layers ensure robust protection of the core against chemicals, oxygen, enzymes and heat.

Spores are characterized by their size, shape, and location in cells (3, 21). There is not a typical morphology for all clostridial spores, and few studies have been done to compare the spore morphology of clostridial strains. One study (5) compared the spore morphology of the four main solventogenic species (*C. acetobutylicum*, *C. beijerinckii*, *C. saccharobutylicum,* and *C. saccharoperbutylacetonicum*). Their endospores show all an oval shape. This characteristic cannot be generalized to all solventogenic clostridia since *C. tetanomorphum* yields round-shape spores (147). The spore's location may vary within the mother cell (Figure 7-A & 6-B) from eccentric to terminal (3).
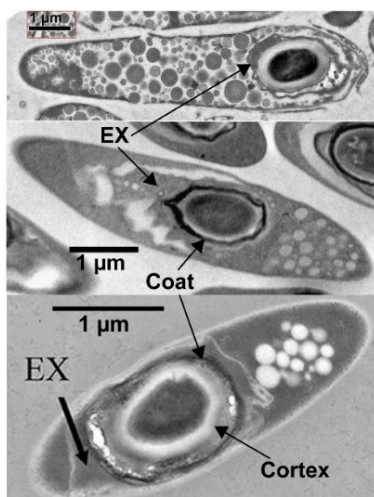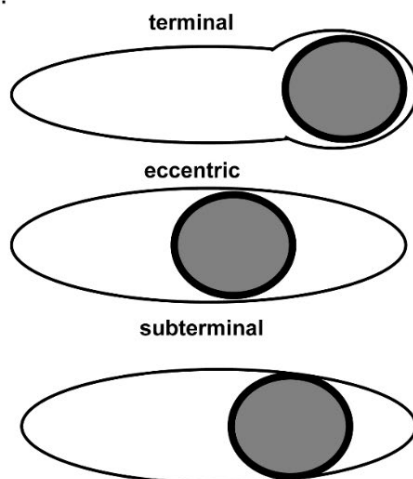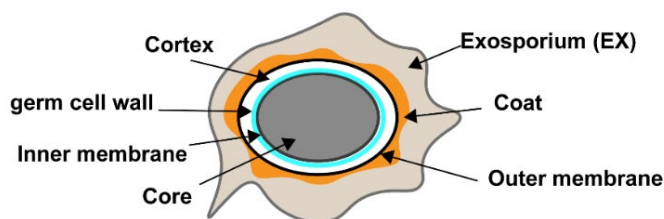
Figure 7: Morphology and composition of the clostridial endospore; (A) Transmission electron micrographs of mature endospores from three solventogenic clostridia. Depending on the species, the size of the cortex, as well as the location of the endospore in the mother cell, changes; (B) Possible location of the endospore in solventogenic clostridia; (C) Composition of the endospore. EX stands for exosporium.

The spore coat (35), the cortex and even the core composition can vary because these features are species-specific. For instance, differences in the spore's DPA and Ca^{2+} content were detected, resulting in a difference in heat resistance depending on the species. Jamroskovic et al. (58) showed that the higher the Ca^{2+}/DPA ratio is, the more resistant the spores are. *C. acetobutylicum* spores can sustain a long heat treatment (> 5 min), at 70°C, while *C. beijerinckii* spores germinate better with a short heat treatment (1 min) at higher temperatures (around 90°C). The heat-shock treatment has to be adapted to the species to give the highest germination efficiency. Knowledge of these differences is crucial as it affects germination efficiency after a heat-shock treatment before fermentation (90, 150, 151).

Spore formation in fermentation: enemy or ally?

Connections between sporulation and solventogenesis

In solventogenic clostridia, sporulation and solvent production are both stress responses to unfavorable environmental conditions. Solventogenesis is thought to be triggered before sporulation (138, 147), although several studies mention a link between sporulation and solvent production (32, 83, 92, 93, 111, 118, 152). Moreover, both cellular events can be lost simultaneously in a phenomenon named "strain degeneration", which is observed after repeated batch fermentation or continuous fermentation (81, 123, 153–157). Genomic and transcriptomic studies have been done on degenerated strains to unravel the mechanism of degeneration and find ways to prevent it. In *C. acetobutylicum*, it has been linked to the loss of the pSol megaplasmid, which contains genes encoding enzymes crucial for solvent production and SpoVD, a protein involved in the engulfment in *Bacillus* (155, 156). In *C. beijerinckii* and *C. saccharoperbutylacetonicum*, which do not harbor the same type of megaplasmid, strain degeneration seems to be caused by mutations. Even though no consensus on the location, number, or type of mutations causing degeneration exists, the cell physiology of these degenerated strains is similar. Changes in medium composition were reported to prevent (158) or reestablish the solvent and sporulation ability, but the molecular mechanism is still unknown (157).

Nevertheless, common regulators of sporulation and solventogenesis have been identified. The study of cellular signaling pathways proved a close link between sporulation and solventogenesis initiation. The RRNPP quorum-sensing systems seem to promote both cellular events. Several regulators enabling the transition from exponential phase to stationary phase are crucial to both cellular events. The disruption of general regulators such as CcpA and σ^L impaired both phenomena (Table 1). As for the role of Spo0A, it seems to differ depending on the species. In contrast to the phenotype described in *C. acetobutylicum* and *C. beijerinckii* (see section on the initiation of sporulation), the expression of *spo0A* is not required for solvent formation in *C. pasteurianum* (128). Proteins involved in Spo0A activation or inhibition, such as AbrB, orphan kinases and σ^K, contribute to the regulation of sporulation and solvent production. Studies showed, nonetheless, that decoupling sporulation and solvent formation is possible. Several solvent-producing but asporogenous mutants and recombinant strains have been isolated and engineered (Table 1) (137, 159).

Spores in industrial processes

Sporulation in solventogenic clostridia is considered a drawback for several reasons. Once sporulation is initiated, cell growth stops, and the cell's energy is used to generate metabolically inactive cells (18). These events are undesirable in industrial settings as they

negatively impact solvent productivity and cause cell wash-out in a continuous process (10, 13, 160). Therefore various attempts were made, like random mutagenesis, inactivation of early-stage sporulation proteins, and engineering of degenerated strain (160) to obtain asporogenous solventogenic strains. Alternatively, modifications of the process can reduce the proportion of sporulating cells. For instance, in continuous culture, the dilution rate is controlled to keep the cells in the growing phase to reduce sporulation (161).

In contrast, bacterial spores can have useful applications, e.g. the pharmaceutical and agroindustry (162–164). Their high resistance to heat and radiation makes them excellent bio-dosimeters. *B. subtilis* spores, for instance, are used to test the UV disinfection performance during drinking water purification (165). *B. subtilis* spores are also known to be effective biopesticides. The use of endospores for enzyme immobilization is actively explored (166). Compared to other immobilization supports, spores are very cheap, and several studies have highlighted the positive effects of spore immobilization on enzyme activity, stability and recovery (167–170). *Clostridium* spores are already used in the food and feed industry in Japan and China. Indeed, *C. butyricum* spores have been used as probiotics for several years. In fact, a large number of acid-producing *Clostridium* species are found in the gut of healthy individuals. They metabolize nutrients that cannot be degraded by the host. Studies have also proven that *C. butyricum* can prevent antibiotic-associated diarrhea and prevent *C. difficile* infections (171, 172). The use of clostridial spores in the pharmaceutical industry is in the pipeline. Spores of *C. novyi* and *C. sporogenes* are being investigated as potential carriers for enzymes involved in chemotherapy (21, 173, 174).

Some research on the integration of sporulation in the bioprocess for butanol production has been done. Spores can be used for cell immobilization. Low biomass is one of the issues of continuous culture with clostridia, and cell immobilization may prevent wash-out at high dilution rates (10). Spore can be immobilized on porous carriers (175, 176) and microencapsulation (177) to prevent cell wash-out. Microencapsulated spores can even be reused several times, enabling the production of butanol at high yields.

Moreover, sporulation enables consistent conservation of the strain characteristics. Due to the oxygen sensitivity of vegetative cells, the storage of spores is preferred. A study performed at Wageningen Food and Biobased Research (178) explored the possibility of integrating a heat treatment during the product recovery by gas stripping in repeated batch cultures. This treatment killed the remaining vegetative cells and triggered the germination of the spores present in the culture. This procedure improved solvent recovery and prolonged strain stability.

Conclusion and perspectives for future studies

Next to solventogenesis, sporulation is a major stationary phase event occurring during ABE fermentation. Changes in carbon sources, media components, and cell density were described as the primary triggers of sporulation. Since studies were mainly conducted with first-generation substrates but with advanced biofuel production in mind, more studies on the effect of second and third-generation feedstocks on sporulation need to be realized.

Despite being a major part of the cell growth cycle, knowledge on its regulation mechanism in solvent-producing clostridia is scarce. Sporulation in *Bacillus* and pathogenic clostridia is well described, but noteworthy differences in the regulatory network exist between the two genera (100) and even within the clostridial pathogens (149). The model established for those species might not be applicable to solventogenic clostridia. As highlighted in a previous review (18), most research is done on *C. acetobutylicum* ATCC 824. Recent studies in other solventogenic strains have revealed differences in the transcriptional regulation of sporulation. Hence, more research in different solventogenic clostridia is required to understand better the complex regulation of sporulation and its interaction with other cellular events.

Knowledge in sporulation might be applied in the design of fermentation processes at different levels; by tailoring the medium composition to reduce/increase sporulation as desired or by integrating clostridial spore in the pretreatment of the substrate by displaying hydrolases on the spore surface or for cell immobilization during a continuous process. A better knowledge of sporulation in solventogenic clostridia would contribute to an improvement of the ABE-fermentation processes for the production of fuels and chemicals from renewable resources as a step towards a more sustainable industry.

1

Thesis outline

Metabolic engineering of *Clostridium* used to be tedious and hampered the studies on strain physiology. An efficient and fast engineering tool was highly desired. **Chapter 2** describes the adaptation of a CRISPR-Cas9 method for use in *Clostridium beijerinckii*. The tool was initially developed for *Clostridium acetobutylicum*, but several modifications of the tool were necessary for use in *C. beijerinckii*. This method allows the deletion and insertion of genes in *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* NRRL B593, an isopropanol-producing strain.

The method developed in Chapter 2 was used to generate a SpoIIE deficient mutant in *C. beijerinckii* NCIMB 8052. The mutant phenotype is described in **Chapter 3**. Morphology, fermentation profile, and transcriptome of the mutant were investigated and compared to the wild-type. Thanks to this mutant strain, we could study the impact of the interruption of the sporulation cycle before forespore formation on other cellular events.

The mutant investigated in Chapter 3 was not the only SpoIIE deficient isolate we obtained. Five SpoIIE deficient strains were isolated, and each of them had a slightly different genotype, but all of them were asporogenic. Four of the five strains showed a solventogenic phenotype, and one was deficient in solvent production. The characterization of a non-solventogenic SpoIIE deficient strain is reported in **Chapter 4.** Morphology, fermentation and transcriptome were compared to the solvent producing Δ*spoIIE* strain described in Chapter 3.

Apart from spores, *C. beijerinckii* can form other intracellular compartments called bacterial microcompartments (BMCs). Like spores, BMCs are formed due to changes in the environment and, more precisely, when *C. beijerinckii* metabolizes L-rhamnose. In **Chapter 5**, the metabolism of *C. beijerinckii* NRRL B593 on rhamnose was investigated with respect to its fermentation profile, product formation, transcriptome**.**

Chapter 6 summarizes the results described in this thesis and discusses their contribution to *Clostridium* research in general. Moreover, an outlook on further studies on the physiology of industrially relevant clostridia is given. Based on the work described in this thesis, three interesting future research lines are presented.

Chapter 2-

Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*

Mamou Diallo [1*#], Rémi Hocq [2*], Florent Collas [1], Gwladys Chartier [2], François Wasels [2], Hani Surya Wijaya [1], Marc W. T. Werten [1], Emil J. H. Wolbert [1], Servé W. M. Kengen [3], John van der Oost [3], Nicolas Lopes Ferreira [2#], Ana M. López-Contreras [1]

[1]Bioconversion group, Wageningen Food and Biobased Research, Wageningen, The Netherlands.
[2] IFP Energies nouvelles, Biotechnology Department, Rueil-Malmaison, France.
[3] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands
[*] Mamou Diallo and Rémi Hocq contributed equally to this work.

2

Keywords: *Clostridium beijerinckii*; CRISPR-Cas9; nuclease; genome editing

Abbreviations: ABE, Acetone Butanol Ethanol; **CRISPR**, Clustered Regularly Interspaced Short Palindromic Repeats; **Cas**, CRISPR-associated protein; **Cas9n**, Cas9 nickase; **DAE**, Double Allelic Exchange; **DSB**, Double-Strand Break; **gDNA**, genomic DNA; **gRNA,** guide RNA; **HR**, Homologous Recombination; **NHEJ**, Non-Homologous End Joining; **PAM**, Protospacer Adjacent Motif; **PCR**, Polymerase Chain Reaction; **SAE,** Simple Allelic Exchange.

Abstract

Recent developments in CRISPR technologies have opened new possibilities for improving genome editing tools dedicated to the *Clostridium* genus. In this study, we adapted a two-plasmid tool based on this technology to enable scarless modification of the genome of two reference strains of *Clostridium beijerinckii* producing an Acetone/Butanol/Ethanol (ABE) or an Isopropanol/Butanol/Ethanol (IBE) mix of solvents. In the NCIMB 8052 ABE-producing strain, inactivation of the SpoIIE sporulation factor encoding gene resulted in sporulation-deficient mutants, and this phenotype was reverted by complementing the mutant strain with a functional *spoIIE* gene. Furthermore, the fungal cellulase-encoding *celA* gene was inserted into the *C. beijerinckii* NCIMB 8052 chromosome, resulting in mutants with endoglucanase activity. A similar two-plasmid approach was next used to edit the genome of the natural IBE-producing strain *C. beijerinckii* DSM 6423, which has never been genetically engineered before. Firstly, the *catB* gene conferring thiamphenicol resistance was deleted to make this strain compatible with our dual-plasmid editing system. As a proof of concept, our dual-plasmid system was then used in *C. beijerinckii* DSM 6423 Δ*catB* to remove the endogenous pNF2 plasmid, which led to a sharp increase in transformation efficiencies.

Introduction

Solventogenic *Clostridia* are mesophilic, rod-shaped gram-positive, anaerobic and sporogenic bacteria belonging to the *Firmicutes* phylum. Model species such as *Clostridium acetobutylicum* and *Clostridium beijerinckii* can produce a wide range of valuable chemicals, among which acetone/isopropanol, butanol and ethanol during the A/IBE fermentation (4, 153, 179–181). Moreover, *Clostridium* species have been reported to produce other chemicals such as acetoin, 2,3-butanediol (180, 182), or 1,3-propanediol (183–185). However, the A/IBE fermentations suffer from the low end-products final concentrations mainly related to their toxicity (186).

Over the last decades, many efforts were made to improve the solvent tolerance and productivity of solventogenic *Clostridia* by genetic engineering. Forward genetics (i.e., random mutagenesis) have been mostly used to study molecular mechanisms in *Clostridia* (187, 188), because targeted mutagenesis has long been impeded by the lack of efficient genetic tools for the *Clostridium* genus. Since the adaptation of the TargeTron technology (189) to the *Clostridia* genus in 2007 and the creation of a genetic tool called ClosTron (190, 191)), targeted approaches have been greatly facilitated. This technology allows gene disruption through the precise insertion of a mobile group II intron into the genome. However, polar effects on flanking genes have been reported, making the subsequent analysis of mutant phenotypes difficult (192). Moreover, considering that this method does not enable large genetic insertions, other tools were created to alleviate these bottlenecks. In particular, strategies using counter-selection markers have been developed to select rare allelic exchange events (104, 107, 193). In 2013, Al-Hinai et al. (193) described a markerless genome editing strategy using the MazF toxin from *Escherichia coli* as a counter-selection marker. The methodology enabled the deletion of several genes and the insertion of a 3.6 kb DNA fragment, leaving an FRT scar in the modified chromosome. Ehsaan et al. (104, 194) developed the ACE system (Allele Coupled Exchange) using the *pyrE* gene, encoding the pyrimidine biosynthesis enzyme "orotate phosphoribosyltransferase", as a positive and negative selection marker. However, these methods are time-consuming as they require several selection and subculturing steps. The cells in which the plasmid has been integrated into the genome via a single crossover event are selected and then subcultured in selective media to further allow the allelic exchange. This lengthens the engineering process and favors the occurrence of spontaneous mutations.

CRISPR-Cas9 technology for genetic modification of eukaryotic and prokaryotic cells provides an alternative to these genome editing methods (195). Recently, several publications reported the use of CRISPR-Cas9 tools for scarless genome editing in *Clostridium cellulolyticum* (196), *C. acetobutylicum* (105) and *C. beijerinckii* (105, 197–199). Most groups reported difficulties when the native Cas9 nuclease from *S. pyogenes* was constitutively expressed. Indeed, the double-strand break (DSB) generated by Cas9 is

reported to be lethal because of the absence of non-homologous end-joining (NHEJ) mechanisms in *Clostridia* (196, 200). Therefore, the Cas9 nuclease was often replaced by a Cas9 nickase (105, 198) in which the RuvC domain was mutated (D10A) (201). This methodology enabled the insertion of DNA fragments up to 1.7 kb after serial transfers (196). Alternatively, Wang et al. (197) overcame the DSB lethality by placing the *cas9* gene under the control of a lactose inducible promoter (202), enabling deletions, insertions or single-nucleotide modifications (SNM) within the genome of *C. beijerinckii*. All the systems described above are based on a "single vector strategy". However, such strategies are limited due to the length of the *cas9* gene (4.1 kb), which restricts the cargo capacity of the transformation vectors and therefore limits the size of potential genome insertions.

Approaches using two separate plasmids for the sequential introduction of the *cas9* gene and the gRNA sequence into a targeted strain have been recently described in the solventogenic *C. acetobutylicum* (203) and in the acetogen *C. authoethanogenum* (204). In *C. acetobutylicum*, the system enabled precise genome editing with an efficiency of 100%. Indeed, the system enabled several modifications ranging from a single base pair substitution to the genomic insertion of a 3.6 kb DNA fragment. In the case of *C. autoethanogenum*, the system enabled a partial deletion of an alcohol dehydrogenase gene and the deletion of a 2,3-butanediol dehydrogenase gene (1 kb) with an efficiency of 50%. Although being powerful, the two-plasmid editing strategy requires two functional selection markers, which might not always be available for some strains.

In this study, we adapted our recently-described dual-plasmid CRISPR-Cas9 genetic tool (203) to edit *C. beijerinckii* strains. In the ABE model strain *C. beijerinckii* NCIMB 8052, the *cas9* gene and the gRNA expression cassette were placed on two distinct plasmids. Both *cas9* and gRNA were placed under the control of a xylose inducible promoter (205) to enable tight control of nuclease activity. The homologous DNA repair template, required for allelic exchange, was placed on the same plasmid as the gRNA expression cassette. The lethal effect of DSB was taken as a strong asset to use the CRISPR system as a counter-selection against unedited cells. This system enabled the scarless deletion of 2.379 kb within the *spoIIE* gene (cbei_0097) as well as the insertion of the *celA* gene from *Neocallimastix patriciarum* in *C. beijerinckii* NCIMB 8052 with 100% and 75% efficiency, respectively.

The applicability of this two-plasmid approach was also investigated in the IBE producing *C. beijerinckii* DSM 6423 strain, for which successful genetic modifications have never been reported so far. Being only able to use the erythromycin resistance gene as a selection marker, a preliminary strategy involving a single-plasmid CRISPR-Cas9 tool was undertaken to inactivate the CIBE_3859 gene (*catB*), encoding a putative amphenicol acetyltransferase. The Δ*catB* mutant was sensitive to thiamphenicol and therefore compatible with a two-plasmid editing approach, which was subsequently used to delete

the endogenous pNF2 plasmid as a proof of concept. Interestingly, the removal of this mobile genetic element resulted in a drastically increased transformation efficiency, paving the way to further genome editing of this natural isopropanol producer.

Materials and Methods

Strains and culture condition

Relevant characteristics of the bacterial strains and plasmids used in this study are listed in Table 1. *C. beijerinckii* NCIMB 8052 and DSM 6423 are laboratory strains. The *spoIIE* mutant was stored as vegetative cells in mCGM with 20% glycerol at -80ºC. All the other strains were kept as spore suspension. Except for fermentation assays, liquid cultures of *C. beijerinckii* NCIMB 8052 and related mutants were grown in mCGM containing per liter: yeast extract, 5 g; KH_2PO_4, 0.75 g; K_2HPO_4, 0.75 g; asparagine•H_2O, 2 g; $(NH_4)_2SO_4$, 2 g; cysteine, 0.5 g; $MgSO_4$•7 H_2O, 0.4 g; $MnSO_4$•H_2O, 0.01 g; $FeSO_4$•7 H_2O, 0.01 g; glucose, 10 g. These microorganisms were grown and selected on mCGM agar containing per liter: yeast extract, 1 g; tryptone, 2 g; KH_2PO_4, 0.5 g; K_2HPO_4, 1 g; $(NH_4)_2SO_4$, 2 g, $MgSO_4$•7 H_2O, 0.1 g; $MnSO_4$•H_2O, 0.01 g; $FeSO_4$•7 H_2O, 0.015 g; $CaCl_2$, 0.01g; $CoCl_2$, 0.002g; $CaCl_2$, 0.002 g; glucose, 50 g and agar 12 g. *C. beijerinckii* DSM 6423 and related mutants were cultivated in liquid 2YTG (per liter: tryptone, 16 g; yeast extract, 10 g; NaCl, 5 g; glucose, 20 g, pH 5.2) and grown on solid 2YTG (same as above, with neutral pH, 15 g/L agar and 5 g/L glucose). *Clostridium* mutants were selected in a culture medium containing 50 µg/mL of erythromycin (Duchefa), 650 µg/mL spectinomycin (Duchefa) and/or 15 µg/mL thiamphenicol (Sigma-Aldrich), as needed. Liquid media were made anaerobic by flushing with nitrogen gas or by incubation for at least 4h in an anaerobic chamber. Cultivation was performed at 37 °C, without shaking, and anaerobically in i) an anaerobic chamber or ii) in glass serum vials as described above. *E. coli* and *S. cerevisiae* strains were stored as vegetative cells in 15% glycerol solution at -80°C. *E. coli* was grown on LB medium, supplemented with 100 µg/mL of ampicillin (Duchefa), 100 µg/mL spectinomycin (Duchefa), 25 µg/mL chloramphenicol (Sigma-Aldrich) and/or 500 µg/mL erythromycin (Duchefa).

2

Table 2 : Strains and plasmids used in this study.

Strains and plasmids	Relevant information	Reference or source
Clostridium		
***Clostridium beijerinckii* NCIMB 8052**	ABE producer, sensitive to erythromycin (25-50 µg/mL) and spectinomycin (650 µg/mL)	(206)
***Clostridium beijerinckii* DSM 6423**	IBE producer, *adh* gene, sensitive to erythromycin (25-50 µg/mL), resistant to thiamphenicol (15 µg/mL)	(102)
***Clostridioides difficile* str. 630**	Gene *xylR* and promoter of the gene *xylB*	(205)
Other microbial strains		
***E. coli* XL1-blue**	endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)	Agilent
***E. coli* NEB 10-beta**	Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (Str^R) rph spoT1 Δ(mrr-hsdRMS-mcrBC)	NEB
***E. coli* INV110**	F' {traΔ36 proAB lacI^q lacΔM15} rpsL (Str^R) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) Δ(mcrCmrr)102::Tn10(Tet^R)	Invitrogen
***S. cerevisiae* W303**	MATa/MATα {leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15} [phi+]	ATCC
***C. beijerinckii* NCIMB 8052 genome editing plasmids**		
pMTL500E	Ap^R ; Em^R; colE1 ; pAMβ1	(207)
pSEC500E	Ap^R ; Em^R; colE1 ; pAMβ1 ; 2µ ; URA3	This study
pYC100S	Sp^R; colE1; pCB102; 2µ ; URA3	This study
pWUR100S (pS)	Sp^R; colE1; pCB102	This study
pE_X_*cas9*	pSEC500E derivative with *cas9* (inducible *xylB* promoter) insertion	This study
pS_X^{R}KR	pWUR100S derivative with *Kan^R*; chimera reprogrammable gRNA (inducible *xylB* promoter) and MCS	This study
pS_X^R_*celA*_S1	pSX^{R}KR derivative with cassette "celA" and gRNA celA_S1 (PAM sequence located 36 bp downstream of the *hbd* gene (cbei_0325) stop codon) insertion	This study
pS_X^{R}ThlPcelA_S1	pS_X^R_*celA*_S1 derivative with the thl promoter from cbei_0411	This study
pS_X^R_Δ*spoIIE*_S1	pSX^{R}KR derivative with cassette "ΔSpoIIE" and gRNA spoIIE_S1 (PAM located 571 bp downstream the start codon of *spoIIE*) insertion	This study

Strains and plasmids	Relevant information	Reference or source
pWUR3	pMTL500E derivative with *celA* (constitutive promoter *eglA* and its signal peptide) insertion	(206)
SpoIIE complementation plasmid		
pSpoIIE	pWUR100S derivative with the *spoIIE* (cbei_0097)	This study
***C. beijerinckii* DSM 6423 genome editing plasmids**		
pFW01	Em^R; colE1; pCB102	(203)
pCas9$_{ind}$	pFW01 derivative with *cas9* (inducible Pcm-tetO2/1 promoter) and *tetR* (miniPthl promoter) insertions	(203)
pEC750C	Tm^R; colE1; pIP404	(203)
pCas9$_{ind}$-Δ*catB*_S1	pCas9$_{ind}$ derivative with cassette "ΔcatB" and gRNA catB_S1 (PAM sequence located 76 bp upstream of the *catB* gene (CIBE_3859) stop codon (inducible Pcm-2tetO1 promoter)) insertion	This study
pGRNA-pNF2_S1	pEC750C derivative with gRNA pNF2_S1 (PAM sequence located 285 bp downstream of the CIBE_p20002 gene start/stop codon (inducible Pcm-2tetO1 promoter)) insertion	This study

Ap, ampicillin; Em, erythromycin; Sp, spectinomycin; Tm, thiamphenicol; Kan, kanamycin; Tet, tetracycline.

Plasmid construction and transformation

Except where noted, plasmids and DNA templates used for *C. beijerinckii* NCIMB 8052 CRISPR system were generated as yeast/*E. coli*/*Clostridium* shuttle vector in the yeast *S. cerevisiae* by homologous recombination as described previously (208). Yeasts were transformed by the lithium acetate method (209). Competent cells were transformed with linear plasmids and insert(s) produced by PCR. Primers are listed in Table S1. Each primer contained a 30 bp 5'-tail specific to another DNA fragment to allow homologous recombination. Final yeast/*E. coli*/*Clostridium* shuttle vectors obtained from yeast were digested by AvrII and self-ligated using the T4-DNA ligase (NEB) to evict the yeast replication origin (2microns) and selection marker (URA3 gene).

The SpoIIE complementation plasmid was constructed using the Circular Polymerase Extension Cloning (210). The pS shuttle vector was digested by SphI and XhoI enzymes for linearization, and the insert was amplified by PCR on genomic DNA from *C. beijerinckii* NCIMB 8052 using the primers Cbei0097prom_F and Cbei0097_R.

The plasmid pS_XR_celA_S1 was constructed (with the promoter of the *thl* gene, cbei_3630, driving expression of *celA*) via homologous recombination in *S. cerevisiae* using the primers listed in Table S1. Unfortunately, non-silent mutations were found in the *celA* gene. In the meantime, transcriptome analysis (211) showed that another copy of the *thl* gene, *cbei_0411* was more expressed, and thus its promoter would be a stronger promoter than *cbei_3630* and a better choice. Therefore, to simultaneously address both matters, a BsaI/XcmI fragment encompassing the relevant regions was replaced with an accordingly designed synthetic BsaI/XcmI fragment (Genscript). The resulting vector, pS_XRThlPcelA_S1, was sequence-verified and used the promoter of the *thl* gene (cbei_0411) to drive the expression of the *celA* gene.

Plasmids were introduced in *C. beijerinckii* NCIMB 8052 and related recombinant strains were prepared according to the protocol described previously (207).

To construct pCas9$_{ind}$-ΔcatB_S1, a DNA cassette containing the ΔcatB template DNA as well as the gRNA targeting the *C. beijerinckii* DSM 6423 *catB* locus (anhydrotetracycline inducible promoter) was synthesized by Eurofin Genomics and cloned in the pEX-A258 commercial vector. This cassette was amplified by PCR (ΔcatB_fwd and ΔcatB_rev primers) and cloned at the XhoI restriction site in pCas9$_{ind}$ (203).

To construct pGRNA-pNF2, a gRNA expression cassette composed of the Pcm-2tetO1 promoter (212) and a 20-nt guiding sequence targeting pNF2 fused with the chimeric gRNA sequence (201), pNF2-gRNA, was synthesized (Eurofins Genomics, sequence available in Table S1), and cloned into the SacI-digested pEC750C.

Prior to electroporation, plasmids were transformed in a *dam⁻ dcm⁻ E. coli* strain (INV110, Invitrogen). Dam⁻ Dcm⁻ DNA was then used to transform *C. beijerinckii* DSM 6423 as described previously (213), with a few modifications outlined thereafter. 2YTG 20 g/L glucose was used for liquid cultures. Mid-exponential phase (OD_{600} = 0.6-0.8) cells were used and 5-20 µg DNA was electroporated with a single pulse (100 Ω, 25 µF, 1,4 kV) in 2 mm-gap cuvettes.

Reprogramming the gRNA protospacer in pS_$X^R K^R$

The CRISPR-Cas9 target sequences were selected after a manual search in the target region for a 25 bp spacer sequence flanked in its 3' end by a 5'-NGG-3' motif named protospacer adjacent motif (PAM). The spacer was designed by the annealing of two primers; i) the primer "Target_Forw" consisting in the spacer sequence to which the motif "5'-GAGG-3'"was added at its 5' end and ii) the primer "Target_Rev" consisting in the reverse complement of the spacer sequence to which the motif "5'-AAAC-3'" was added at its 5' end. Both primers were mixed (1 µM each), heated at 95°C for 10 minutes and cooled down at room temperature. The primer dimer was then inserted in pS_$X^R K^R$ by Golden Gate cloning using the BsaI enzyme (NEB). Ligation and digestion were performed concomitantly for 1 hour at 37°C.

Selection of mutant strains

C. beijerinckii NCIMB 8052 recombinant strains harboring the two plasmids of the CRISPR-Cas9 system were grown at 37°C in 25 mL liquid mCGM medium containing 0 (without induction) or 40 g/L xylose (pre-induction) for 24 hours to 3 days to induce the CRISPR-Cas9 system. Once the optical density of the culture containing xylose reached 1.5 mAU at 600 nm, these cultures were concentrated in 100 µL of fresh mCGM media. 50 µL of the cell suspension were spread on mCGM selective plates supplemented 40 g/L xylose, 50 µg/mL erythromycin and 650 µg/mL spectinomycin and incubated for 24 hours to 3 days to obtain single colonies. Edited cells were screened by colony PCR to detect single and double allelic exchange profiles.

C. beijerinckii DSM 6423 recombinant colonies bearing a single- or dual-plasmid genome editing system were treated as described previously (203), with a 200 ng/mL anhydrotetracycline (Sigma-Aldrich) induction.

Fermentation assays

Fermentations were performed in CM2 medium (161), which contains per liter: yeast extract, 5 g; KH2PO4, 1 g; K2HPO4, 0.76 g; NH4Ac, 3 g; *p*-aminobenzoic acid, 0.1 g; MgSO4·7 H2O, 1 g; and FeSO4·7 H2O, 0.5 g, glucose, 60 g. 50 mL of media were inoculated with 250 µL of a heat-shocked fresh spore suspension and incubated overnight

at 35°C. The overnight cultures were inoculated (5% (v/v)) in a fresh culture medium and incubated at 35°C for 72 hours. The concentrations of metabolites were determined in clear culture supernatants after removal of microbial biomass by centrifugation. Glucose, acetate, butyrate, lactate, acetoin, acetone, butanol and ethanol titers were determined by HPLC as described previously (68, 180) using valeric acid (30 mM) as an internal standard.

Microscopy

Phase-contrast microscopy (Olympus BX51; Olympus, Tokyo, Japan) was used to determine the presence of endospores or pre-spore stages. Native bacteria were observed by phase-contrast microscopy at ×1000 magnifications.

Results

Genome editing strategy

The *cas9* gene was placed on a different plasmid from the gRNA expression cassette and the editing template so that 1) cloning was facilitated (i.e., with the use of smaller plasmids) and 2) space was available for large editing templates. Since previous CRISPR-Cas9-related studies in *Clostridia* hypothesized that DSBs catalyzed by Cas9 are likely to impair transformation efficiency severely (196, 214), both the *cas9* gene located on pE_X_cas9 and the gRNA expression cassette located on pS_XKR were placed under the control of the inducible promoter PxylB from *C. difficile* str. 630 (205) (Figure S1). To complete our 2-plasmid inducible system, the *xylR* regulatory gene was further cloned into the plasmid pS_XKR yielding pS_X^{R}KR.

Editing templates and gRNA expression cassettes were thus inserted into the pS_X^{R}KR plasmid (Figure 1a). Each editing template contains two homology sequences (HS) of at least 500 bp, named HS1 and HS2. For genomic deletion, the two HS flank the region that has to be removed, which contains the CRISPR target site. Genomic insertions were designed by placing the sequence of interest between HS1 and HS2. This configuration allows 3 different allelic exchange events between the genome and the plasmid containing the DNA template (Figure S2): two single-crossover allelic exchanges (SAE) and one double-crossover allelic exchange (DAE). The latter exchange is the only one that results in genomic DNA without the CRISPR-Cas9 target, which enables the survival of the transformants upon Cas9 induction.

The pS_X^{R}KR-based targeting plasmids were then introduced into a *C. beijerinckii* NCIMB 8052 recombinant strain already transformed with the pE_X_cas9 plasmid (Figure 1b-c).

Following plasmid selection, xylose was used to induce the expression of both *cas9* and the gRNA. This critical step enables the selection of correctly edited cells, as, in the absence of non-homologous repair systems, Cas9 targets the bacterial chromosome, resulting in cell death. The final recombinant strains were next obtained by curing the pS_X^{R}KR-based plasmid and the pE_X_*cas9* plasmid by serial subculturing in mCGM media lacking both antibiotics (Figure 1d-e). Alternatively, selectively curing the targeting plasmid allows to directly reuse the modified cells for another round of genome editing (Figure 1g).

To evaluate the robustness of this new genome-editing approach for modification of *C. beijerinckii* NCIIMB 8052, the system was used for: i) a gene deletion and ii) a gene insertion. The modifications performed in this study and their characteristics are summarized in Table 2.

Table 3: Chromosomic modifications performed in this study.

Strain	DNA template	Size of the sequences (base pairs)				Screening results[a]			
		Deleted	Inserted	HS1	HS2	DAE	WT	DAE & WT	"-"
NCIMB 8052	ΔspoIIE	2424	0	741	1020	7	0	0	1
	celA	19	1667	1001	1017	4	0	0	1
DSM 6423	ΔcatB	653	8	400	400	3	0	0	1
a: Screening was performed on colonies growing with the highest xylose / aTc concentrations									
HS1 and HS2 Homologous sequence; WT: wild-type; DAE: Double allelic exchange; SAE: Single allelic exchange; "-": no amplification									

2
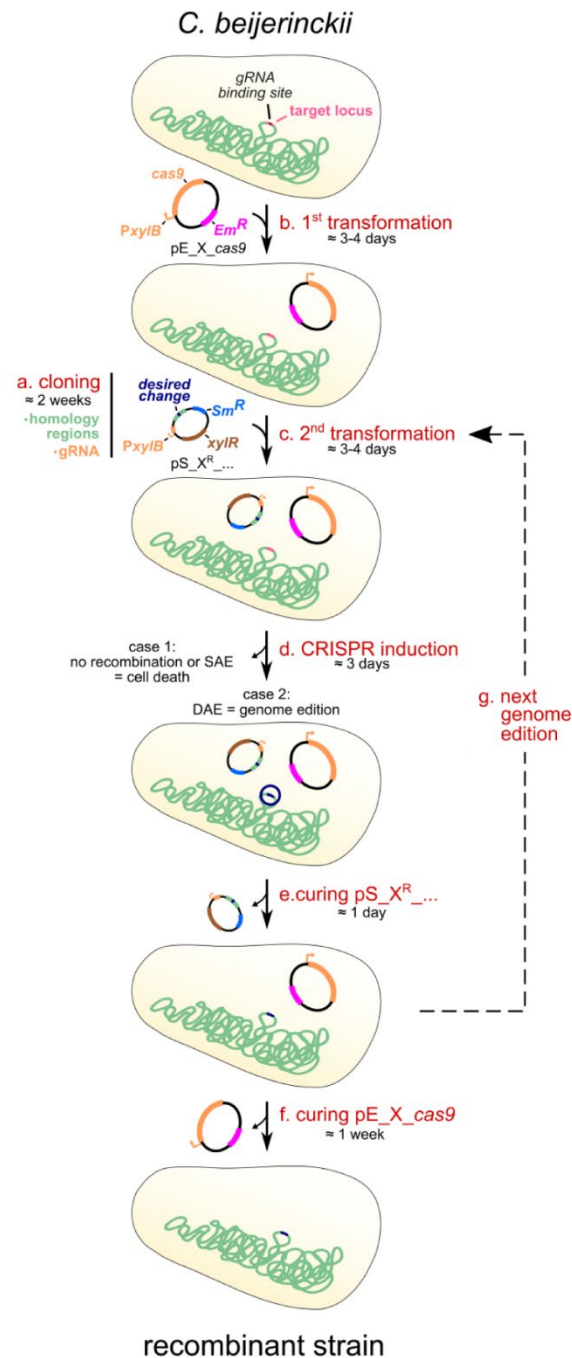
Figure 1: Dual-plasmid inducible CRISPR-Cas9 strategy for *C. beijerinckii* NCIMB 8052 genome editing.

a. Preliminary cloning of homology sequences and gRNA protospacer in pS_XRKR plasmid to obtain a targeting plasmid. b,c. Transformation of (b) pE_X_cas9 in the WT strain and (c) of the targeting plasmid in the latter strain. d. Induction of cas9 and gRNA expression to yield an edited strain. e,f,g. Curing of the targeting plasmid (e) and pE_X_cas9 plasmid (f) yields a markerless genome-edited mutant. Alternatively (g), the targeting plasmid can be specifically removed to promote another round of genome editing.

gRNA : guide RNA; EmR : erythromycin resistance marker; SmR: spectinomycin resistance marker.

Deletion of the *spoIIE* gene, characterization of the edited strains and complementation with a functional *spoIIE* gene

The *spoIIE* gene (cbei_0097) encoding a phosphatase involved in the sporulation regulation was targeted first. In model organisms *Bacillus subtilis* and *C. acetobutylicum*, SpoIIE is involved in the first steps (Stage II) of the regulation of the sporulation mechanisms cycle (23, 214). It enables the activation of the sporulation-specific sigma factor σ^F and is necessary to complete asymmetric division. In *C. acetobutylicum*, SpoIIE inactivation does not impair solvent production (15, 16). The pS_XΔ*spoIIE*_S1 plasmid contains the "Δ*spoIIE*" DNA editing template (Table 1 and Figure 2a) and the "*spoIIE*_S1" gRNA. HS1 and HS2 templates were designed to remove 2379 bp within the coding sequence of the *spoIIE* gene. Without induction of the CRISPR-Cas9 machinery, PCR screening showed that the full pS_XΔ*spoIIE*_S1 plasmid was frequently integrated into the host genome by SAE mechanisms either with HS1 or with HS2. However, no DAE event was observed (data not shown). Overnight xylose induction in the *C. beijerinckii* (pE_X_cas9, pS_XΔ*spoIIE*_S1) strain negatively affected its growth, suggesting a toxic effect of Cas9 activity (Figure 2C). This overnight culture was then plated on mCGM agar containing xylose (Figure 2D), which permitted the selection of correctly edited Δ*spoIIE* cells (7/7 DAE events, Figure 2E and Table 2). Microscopic analysis showed that the Δ*spoIIE* cells were more elongated than the wild-type cells and did not form viable endospores (Figure 3A-B). Similar to what was observed in *C. acetobutylicum*, mutant strains were still able to produce solvents (Table S2). To test whether the asporogeneous phenotype was effectively caused by the disruption of the *spoIIE* gene, complementation with a plasmidic copy of the *spoIIE* gene under the control of its native promoter was performed (Figure S3). As expected, sporulation was restored in the complemented mutant (Figure 3A-B).
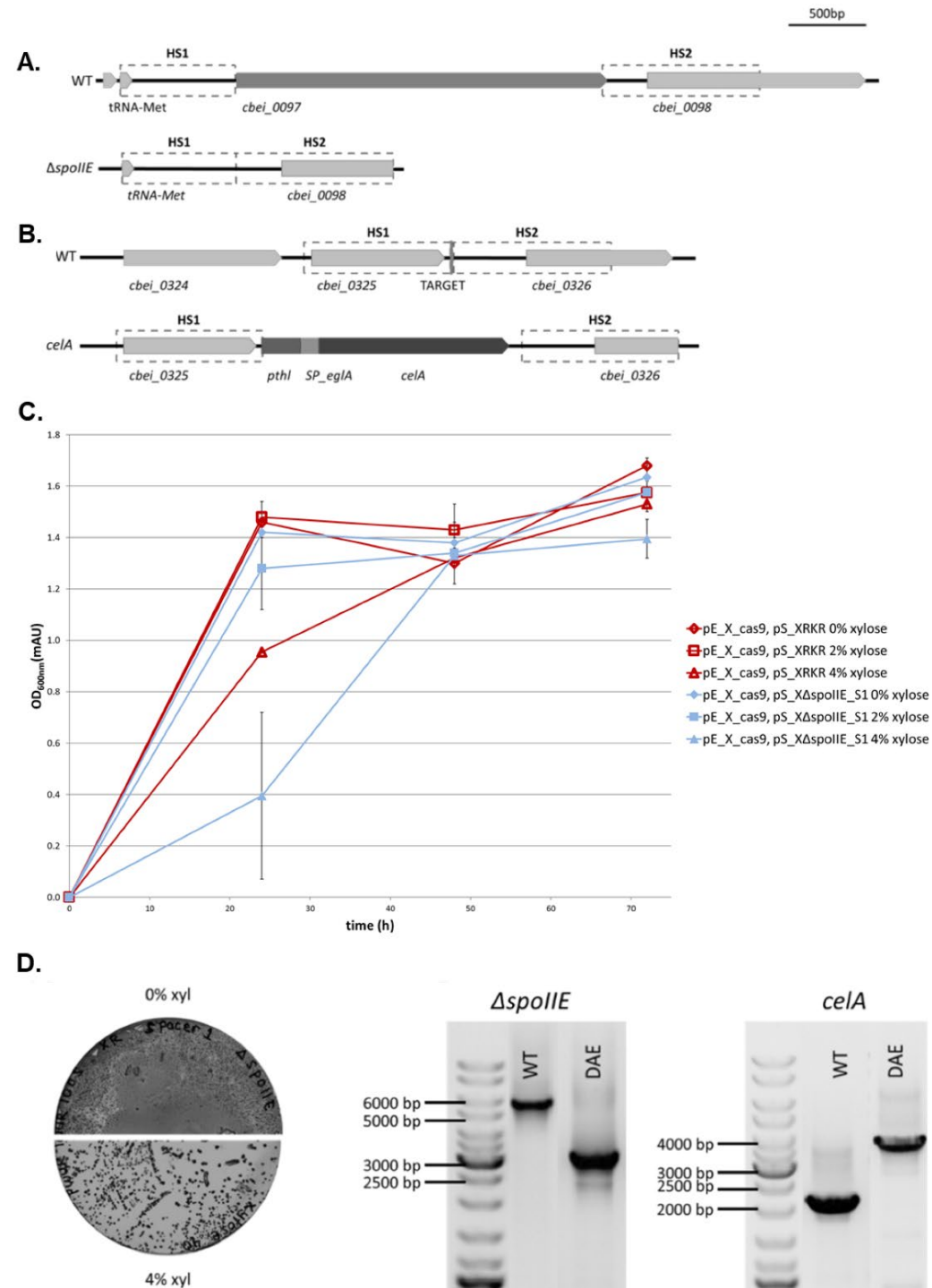
A.

B.

C.

D.

Figure 2 : Genome editing of the *C. beijerinckii* NCIMB 8052 strain: design and selection. A, B. Targeted loci and their relative DNA templates. (A) Disruption of *spoIIE*. (B) Insertion of *celA* from *N. patriciarum*. c, d, e. Selection and screening of edited strain. (C) Growth of *C. beijerinckii* NCIMB 8052 pE_X_cas9 transformed with the empty gRNA plasmid pS_XRKR or the spoIIE disruption plasmid pS_XΔspoIIE_S1. The transformants were grown in duplicates in mCGM supplemented with erythromycin and spectinomycin and with 0 to 4% of xylose. The error bars represent the error of the replicates (D) Selection of Δ*spoIIE* mutants on mCGM agar plate containing 0 or 4 % xylose (w/v) inoculated with overnight pre-induced cultures. (E) Growing xylose-resistant colonies were subsequently screened by colony PCR. The screening for *spoIIE* disruption was performed using primers cbei_0096_F/cbei_0098_R. Expected sizes: 5462 bp (WT) and 3083 bp (DAE). The screening of the insertion of the *celA* gene was performed using primers cbei_0325_F/cbei_0326_R. Expected sizes: 2083 bp (WT) and 3607 bp (DAE). HS1 and HS2: homologous sequence 1 and 2; SP_eglA: signal peptide eglA; pthl: promoter of *C. beijerinckii* thiolase gene (cbei_3630); WT: wild-type; DAE: double allelic exchange (edited cells).

Insertion of the *celA* gene from *Neocallimastix patriciarum*

The *celA* gene from the fungus *Neocallimatix patriciarum* (206, 215), encoding a cellobiohydrolase, was inserted downstream of the *hbd* gene (cbei_0325). The pS_XR_celA_S1 plasmid, harboring the "celA" DNA template (Table 1 and Figure 2b) and the "celA_S1" gRNA, was introduced in strain NCIMB 8052 (pE_X_cas9). The celA template was designed to enable the insertion of the *celA* gene under the control of the thiolase promoter from *C. beijerinckii* (1667 bp) and the deletion of a 19 bp sequence located 36 base pairs downstream of the stop codon of the *hbd* gene (cbei_0325).

Similar to what was observed for *spoIIE* inactivation, only SAE events were detected by PCR in the transformants containing pE_X_*cas9* and pS_XR_*celA*_S1 prior to xylose induction. Following xylose induction, 3 out of the 4 xylose-resistant colonies tested were correctly edited (Figure 2E and Table 2).

The resulting *celA* mutants displayed cellulolytic activity on mCGM plates containing 1 g.L^{-1} CMC (carboxymethylcellulose), similarly to a control *C. beijerinckii* strain carrying the *celA* transcription unit on a plasmid (pWUR3, (206)) (Figure 3B).
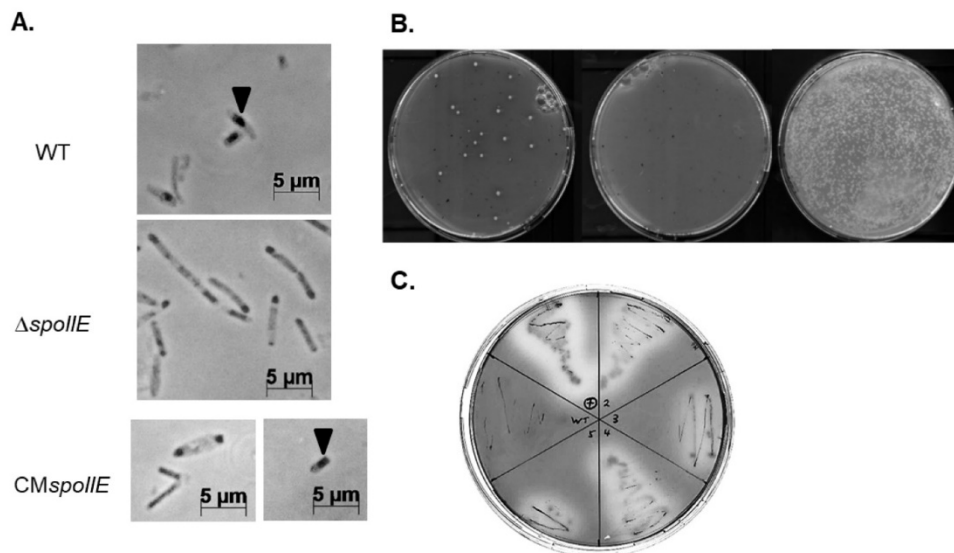
2

Figure 3: Novel phenotypes in genetically modified *C. beijerinckii NCIMB 8052* mutants. (A) Light microscopy pictures at 73 hours of fermentation of *C. beijerinckii* NCIMB 8052 WT, Δ*spoIIE*, CMspoIIE (complemented mutant) strains. Spores are marked by a triangle. (B) Growth of *C. beijerinckii* NCIMB 8052 WT, Δ*spoIIE*, CMspoIIE on CM2 agar plates after heat shocking 200 µL of a 48-hour-old culture. c. Detection of the cellulase activity of *C. beijerinckii celA* mutants. WT : *C. beijerinckii* wild-type, 2 - 5 : clone number (CelA mutants after plasmid curing), + : positive control (non-integrated self-replicating CelA plasmid) the halo surrounding the cells are caused by the degradation of carboxymethylcellulose.

Adaptation of the dual-plasmid genome editing strategy to *C. beijerinckii* DSM 6423

The advantages provided by a two-plasmid CRISPR-Cas9 genome editing system (i.e., scarless modifications, high efficiency, relatively high DNA cargo capacity) prompted us to develop a similar system in the poorly studied *C. beijerinckii* DSM 6423 strain. The first obstacle was the lack of knowledge on the transformability of this natural IBE producer. This hurdle was overcome by applying a DNA demethylation strategy prior to electroporation, as described by Kolek et al. (216) in the closely related *C. beijerinckii* NRRL B-598 strain (Figure S4). This treatment enabled the generation of transformants containing pFW01-derived vectors (126, 203). However, vectors bearing the spectinomycin resistance marker could not be introduced successfully in this strain (data not shown). This was a major bottleneck for the use of a two-plasmid editing strategy since this microorganism was also resistant to thiamphenicol, an antibiotic often used for selection in clostridial recombinant systems (217).

To alleviate this second obstacle, a single-plasmid CRISPR-based approach was developed to disrupt the *catB* gene (CIBE_3859), predicted to encode an amphenicol acetyltransferase probably involved in thiamphenicol resistance. pE_X_*cas9* plasmid was not chosen as a

backbone since this already large vector (≈ 13 kb, Figure S1) does not carry the *xylR* gene necessary for *cas9* inducibility. Instead, the 8874 bp pCas9$_{ind}$ plasmid previously used for genome editing in *C. acetobutylicum* ATCC 824 (203) was engineered to include an anhydrotetracycline-inducible gRNA expression cassette targeting *catB* as well as homology sequences designed to delete 99% of the *catB* gene (Figure 4A, Table 2). The resulting pCas9$_{ind}$-Δ*catB*_S1 vector was introduced in *C. beijerinckii* DSM 6423, and mutants with the expected *catB* deletion were obtained after anhydrotetracycline (aTc) induction (Figure 4B).

As anticipated, these Δ*catB* mutants were sensitive to thiamphenicol (Figure 4C) and allowed the maintenance of the pEC750C vector in this modified strain, which carries the *catP* marker from *C. perfringens* (218). Besides, this vector could be introduced in a strain already bearing a pFW01-based plasmid. In order to test whether a dual-plasmid CRISPR-Cas9 editing approach could be applied to the Δ*catB* mutant, we constructed a pEC750C-based plasmid carrying an aTc-inducible gRNA cassette targeting the endogenous pNF2 plasmid (pGRNA-pNF2_S1). The pNF2 plasmid (102) was discovered during the reconstruction of the *C. beijerinckii* DSM 6423 genome along with other extra-chromosomic genetic elements (i.e., another plasmid and a linear phage). The pNF2 plasmid is the smallest of these elements (4282 bp) and bears 4 putative genes, 3 of which have no predicted function, and the last one is predicted to encode a potential plasmid replication protein. The pGRNA-pNF2_S1 vector was subsequently introduced in *C. beijerinckii* DSM 6423 Δ*catB* (pCas9$_{ind}$), which allowed the highly efficient selection of ΔpNF2 mutants (Figure 4D-E). Fermentation assays further revealed that the deletion of ΔpNF2 did not affect alcohol production (Table S3). However, pNF2 removal resulted in increased transformation efficiencies (from 20 to 2000 fold depending on the plasmid, Figure 4F).

In summary, we successfully adapted the two-plasmid CRISPR-Cas9 editing tool to *C. beijerinckii* DSM 6423 and created successively two platform strains. The first modification made the strain sensitive to thiamphenicol, which allowed us to use two different resistance markers. This mutant strain was further engineered by removing pNF2, which allowed us to obtain a second strain with higher transformation efficiencies.
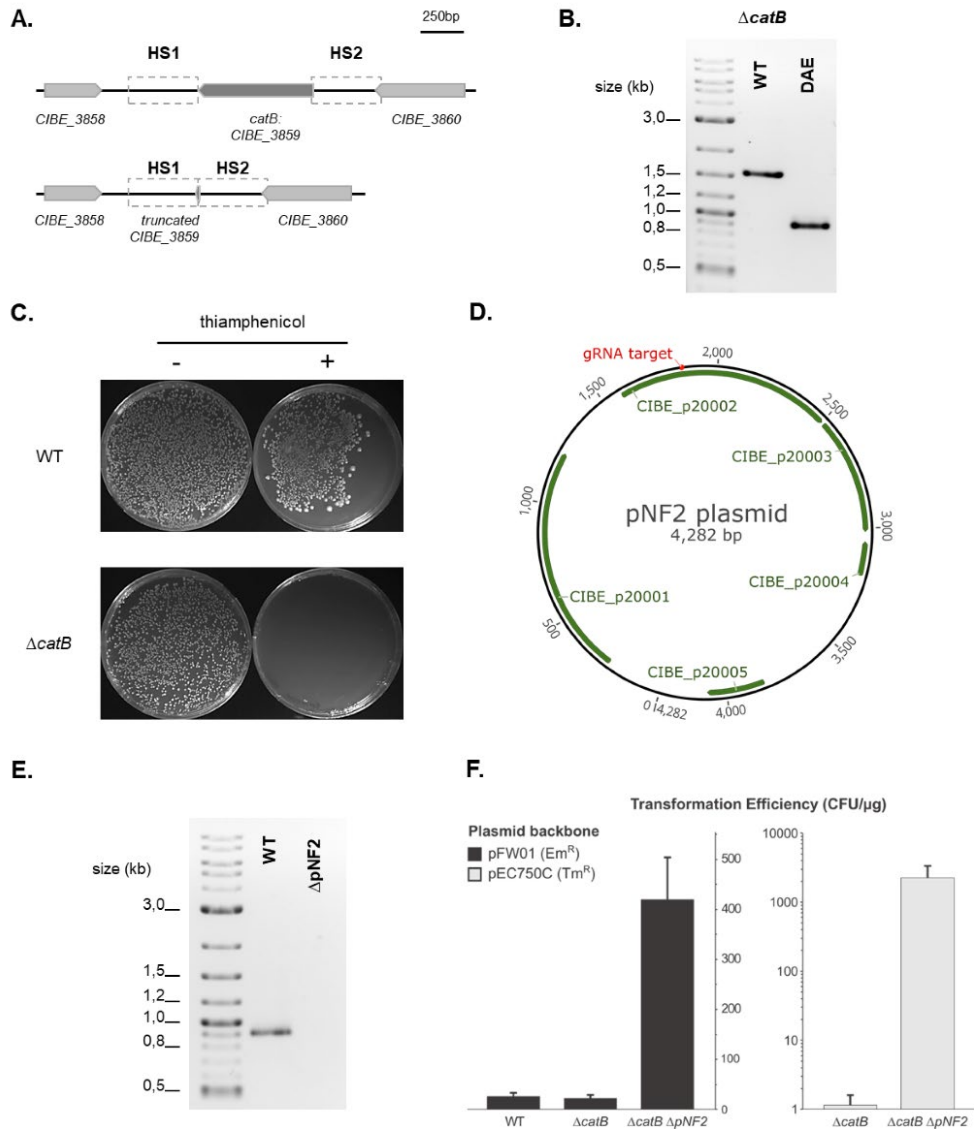
Figure 4: Adaptation of the two-plasmid CRISPR-Cas9 system to *C. beijerinckii* DSM 6423. (A) *catB* (CIBE_3859) genome deletion design by a single-plasmid strategy. (B) Verification of catB deletion by PCR with primers RH76 and RH77, which amplify a genome region encompassing the homology sequences (expected sizes: 1553 (WT) and 900 (Δ*catB*) bp). (C) Thiamphenicol resistance assay on the resulting ΔcatB strain. (D) Map of the natural pNF2 plasmid from *C. beijerinckii* DSM 6423. The locus targeted by the pGRNA-pNF2_S1 vector (containing the thiamphenicol resistance gene catP) is shown in red. (E) Verification of the deletion of the pNF2 plasmid with a dual-plasmid CRISPR-Cas9 strategy. Primers pNF2_fwd and pNF2_rev were used for PCR (expected size: 907 bp (WT) ; no amplification is expected from a ΔcatBΔpNF2 strain). (F) Transformation efficiencies in the wild-type, the Δ*catB* and the ΔcatB ΔpNF2 strain. Errors bars represent the s.e.m. of duplicate experiments. EmR : erythromycin resistance gene TmR : thiamphenicol resistance gene.

Discussion

In this study, we describe the successful adaptation of a CRISPR/Cas9 genetic tool in the solvent-producing *C. beijerinckii*. The two-plasmid system, in which the *cas9* gene was cloned on a first plasmid - the nuclease plasmid - while the gRNA and the DNA template were cloned on a second plasmid – the editing plasmid, was successfully used to edit the genome of two reference strains of *C. beijerinckii*. Such a method benefits its user on several points. Firstly, the cloning procedure is facilitated because the toxic genes of the CRISPR-Cas9 machinery are localized on separate genetic elements. Therefore, it is easier to construct plasmids for a given genetic modification, which will also be significantly smaller, likely resulting in higher transformation efficiencies in *Clostridium*.

Secondly, our methodology helps to reduce the time spent on the traditionally laborious genetic engineering steps. The strain bearing the nuclease plasmid can indeed be stored at -80°C and used multiple times for several genetic constructions. Similarly, the editing plasmid can be specifically cured after a genome editing experiment, permitting multiple successive rounds of genome editing. This editing plasmid was besides designed to fasten the cloning process by incorporating genetic elements allowing a quick and easy reprogramming of the gRNA. In our hands, CRISPR-Cas9-based counter-selection also proved to be extremely efficient with most clones – if not all – edited after the induction step, which also contributes to shortening the mutant screening process.

Lastly, the system displays a good genetic cargo capacity (up to 3.5 kb insertions with *C. acetobutylicum* (203)), which makes it suited for large genome modifications required for metabolic engineering. The additional space provided by the second plasmid also gives more freedom for the design of DNA templates and notably enables the use of longer homologous sequences to favor allelic exchanges (194).

In the CRISPR system described in this study, an inducible Cas9 nuclease (197, 204) was chosen over the Cas9 nickase previously used by other groups (105, 196). In those studies, edited cells were obtained at lower frequencies than in studies when the Cas9 nuclease was used. This can be explained by lower toxicity of a single nick in the genome rather than a DSB. Since the Cas9 nickase is less toxic and therefore less selective, its use increases the probability of having a mixed population, especially when the desired mutation negatively impacts the fitness of the bacterium or when the editing events are rare (e.g., insertion of large DNA fragments). As an example, the disruption of *adc* in *C. acetobutylicum* or *C. beijerinckii* using the nickase led to the isolation of 1 and 3 correctly edited colonies out of 16 colonies tested, respectively, whereas other mutations (*xylR*, cbei_3923) were obtained in almost all of the tested colonies (105).

In order to use Cas9, the inducible promoter PxylB from *C. difficile* and the *xylR* gene, whose product codes for the associated regulatory protein (205), were used to control the

expression of both *cas9* and the gRNA. The lethal effect of the DSB resulting from their expression following xylose induction was then used as a strong counter-selection marker. In this study, this approach allowed the isolation of correctly edited cells at an efficiency close to 100%. Moreover, unlike what is reported in studies using the Cas9 nickase, serial transfers of recombinant strains were not required to isolate clones that undergone large fragment insertions in their genome. All mutants generated in this study were therefore isolated within one week starting from the strain *C. beijerinckii* NCIMB 8052 (pE_X_cas9). Although the system is mediated by two plasmids, their removal required only a few serial transfers making the system as efficient as those previously described. The easy clearance of the plasmids carrying the gRNA and DNA template and the absence of selection marker in the final strain thus enable multiple rounds of genome editing.

In this study, we deleted a 2.379 kb DNA fragment within the *spoIIE* gene (cbei_0097), and inserted the 1.68 kb *N. patriciarum celA* gene. The disruption of the *spoIIE* gene, described here for the first time in *C. beijerinckii*, resulted in asporogenous mutant strains. These results are in accordance with studies in earlier *Bacillus subtilis* (219) and other solventogenic *Clostridia* (15, 36), which showed that a reduced *spoIIE* expression coincided with hampered sporulation and that *spoIIE* disruption prevents sporulation.

The *spoIIE* mutants generated in this study kept the ability to produce solvents at similar levels to the wild-type strain (Table S2). These observations indicate that this gene is necessary for sporulation but does not play a role in solvent production, contrary to *spo0A*, for example. This characteristic has also been observed in *spoIIE* negative mutants generated in *C. acetobutylicum* (15, 16). In contrast to the *spo0A* homologs in *Clostridia*, the *spoIIE* gene is not well conserved between clostridial species. Indeed, while *C. acetobutylicum spoIIE* gene has only 36% identity with *C. beijerinckii spoIIE*, *spo0A* homologs have as much as 74% sequence identity. We observed differences at the phenotypic level between *C. acetobutylicum* and *C. beijerinckii spoIIE* mutants. Indeed *spoIIE* mutants in *C. beijerinckii* harbor two dark spots at their poles (Figure 3), corresponding to the forespore condensed DNA, that are absent in *C. acetobutylicum spoIIE* mutants but were observed in *C. acetobutylicum sigE* mutants (109). Moreover, *C. beijerinckii spoIIE* mutant cells display an asymmetric septum and sometimes two septa at each pole of the cell, which is similar to the morphology of *B. subtilis spoIIE* mutants (214, 220). We can thus hypothesize that even though SpoIIE is necessary for the correct completion of stage II of the sporulation cycle in *C. beijerinckii* and *C.acetobutylicum*, it is not needed for asymmetric septum formation in *C. beijerinckii* in contrary to what is observed in *C. acetobutylicum*.

The insertion of the *"thl_p_celA"* transcriptional unit yielded mutant strains that showed cellulase activity on plate assays, not observed with the WT strain. The CelA enzyme from the anaerobic fungus *N. patriciarum* is a cellulase that contains an active domain belonging

to the Family 6 of glycosyl hydrolases. This enzyme has been expressed before in clostridial strains using a plasmid vector, either alone or as part of a mini-cellulosome (206, 221). This study shows for the first time the insertion of a cellulase gene into the chromosome of a solventogenic strain and its functional expression. This opens novel possibilities for the direct fermentation of sugar polymers like cellulose into ABE or IBE mixtures.

The success of our genome editing strategy prompted us to apply it also to another strain. The atypical *C. beijerinckii* DSM 6423 is genetically very close to the NCIMB 8052 model strain (222), but its modification by targeted genome editing techniques has, however, never been reported, in spite of the microorganism value as being one of the few natural IBE-producing strains. Taking into account this strain characteristics, we successfully adapted our dual-plasmid inducible CRISPR-Cas9 approach. Indeed, some adjustments were found necessary to 1) circumvent the restriction-modification system of the strain (using Dam⁻ Dcm⁻ DNA, similarly to the NRRL B-598 strain (216)) and 2) find an alternative to the spectinomycin resistance gene, which proved to be malfunctioning in our hands. This bottleneck was specifically tackled by using a single-plasmid inducible CRISPR-Cas9 strategy to disrupt a putative amphenicol resistance gene (CIBE_3859, encoding the *catB* gene). The resulting strain was demonstrated to be sensitive to thiamphenicol, and the *catP* resistance marker (from *C. perfringens* (218), used extensively in *C. acetobutylicum* (217)) could subsequently be used in this mutant. This mutant was further exploited to delete the pNF2 natural plasmid (102) with a dual-plasmid inducible CRISPR-Cas9 system, underlining the compatibility of the Δ*catB* strain with such a system. The curation of the pNF2 plasmid was preferred to the modification of chromosomal genes because it required the introduction of a small editing plasmid without an editing template. Indeed the low transformation efficiency of the DSM 6423 strain severely impedes our efforts. For the disruption of the *catB* gene, for example, several transformation experiments were required to introduce the single-plasmid inducible CRISPR-Cas9, and only one was successful, yielding one transformant when using 20 µg DNA. This problem underlines the need for better transformation efficiencies in *C. beijerinckii* DSM 6423 for future genome editing experiments.

CRISPR-Cas9-based removal of non-essential genetic elements has already been pioneered in solventogenic *Clostridia*, with notable effects on the primary/secondary metabolism and/or on exogenous plasmid transformation efficiencies (203, 223). If the product pattern and quantities of the resulting Δ*catB* ΔpNF2 strain were not strongly affected, the mutant interestingly showed much higher transformation efficiencies, most likely because the endogenous pNF2 plasmid interferes with the replication of non-natural plasmids introduced into the cell. As mentioned above, this improvement is critical as transformation efficiencies usually are extremely low in the wild-type genetic background (c.a. 4 CFU/µg

with an empty plasmid – Figure S4), which will likely considerably help to construct future strains. The rise of the CRISPR/Cas9 technology in the field of *Clostridium* genetics has been a real breakthrough. The relatively simple generation of scarless mutants in a short period of time will hopefully facilitate the research into the physiology of solvent production and lead to a better understanding of the regulatory mechanisms involved in gene expression. The dual-plasmid system described in this study should overcome issues reported previously (higher transformation efficiency, flexibility for the design of DNA templates, convenience of the cloning procedure, rapid mutant selection, cargo capacity). We showed here that this advantageous system, initially developed for *C. acetobutylicum*, could be simply adapted to work as efficiently in a model microorganism from another species. We also demonstrated that this approach could be expanded to atypical microorganisms, recalcitrant to genetic manipulations, by following simple steps (i.e., finding the correct DNA methylation pattern, removing antibiotic resistance genes, removing mobile genetic elements), which suggests that similar strategies are likely to function in other prokaryotes.

Funding information

Supplementary files
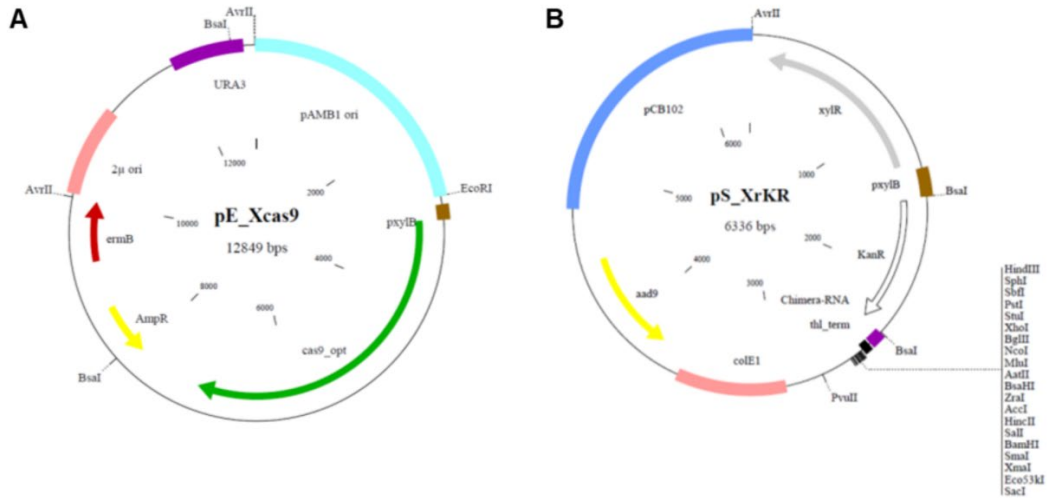
Figure S1 : Maps of the plasmids pE_X_cas9 (A) and pS_XRKR (B). pAMβ1 replication origin; xylBp: promoter of the operon *xylAB* from *C. difficile*; cas9opt: codon-optimized *cas9* gene; colE1, gram-negative replicon; ampR: gene resistance ampicillin; ermB: gene resistance erythromycin; 2µ: yeast replicon 2µ; xylR: gene of the regulation protein of the pxylB; kanR: kanamycin resistance gene; Chimera: minimum sequence for gRNA (26); thlT: terminator of the gene thl (cbei_3630). MCS: multi-cloning site; aad9: spectinomycin resistance gene, pCB102: replication origin.
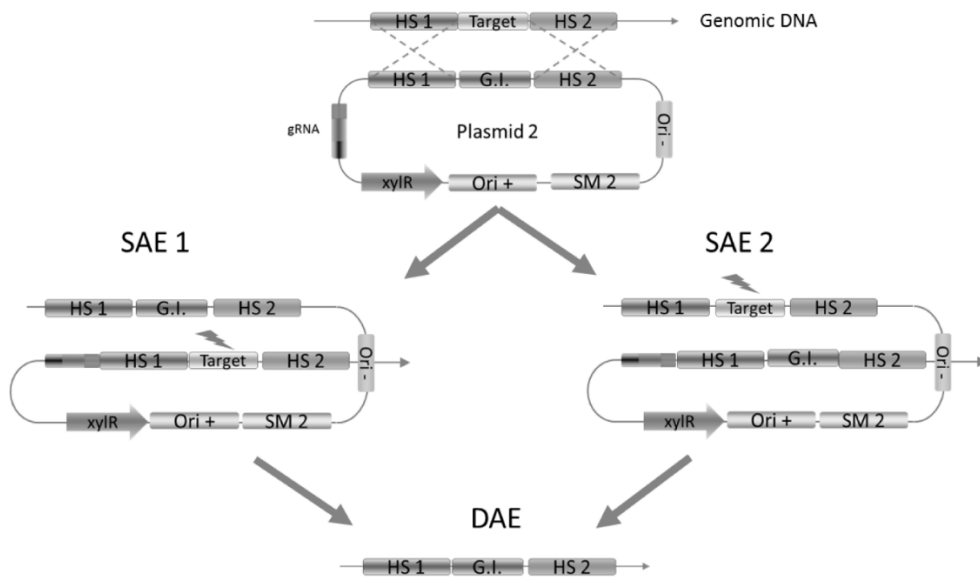
Figure S2 : Allelic exchange events between the genomic DNA and editing DNA template. In our two plasmid system, the plasmid 2 harbors the gRNA directed against the target on the genomic DNA, the regulator of the promoter *xylB* from *C. difficile* sp. 630, and the editing DNA template consisting of the 2 sequences flanking the target in the genome (HS1 and HS2, Homologous Sequences) and a gene of interest (G.I.) inserted between the two homologous sequences. SAE: Single Allelic Exchange and DAE Double Allelic Exchange. In cells with an SAE, the target site is still present in the genome, whereas it disappeared in the DAE cells. Strains harboring SAE will then be sensitive to the activated CRISPR-Cas9 system.
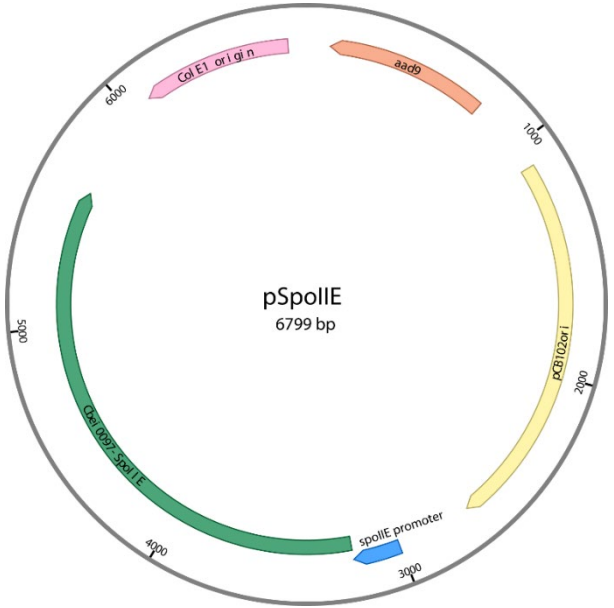
Figure S3: Maps of the pSpoIIE plasmids: *colE1***, Gram-negative replicon,** *aad9***: spectinomycin resistance gene, pCB102: Gram-positive replication origin.**
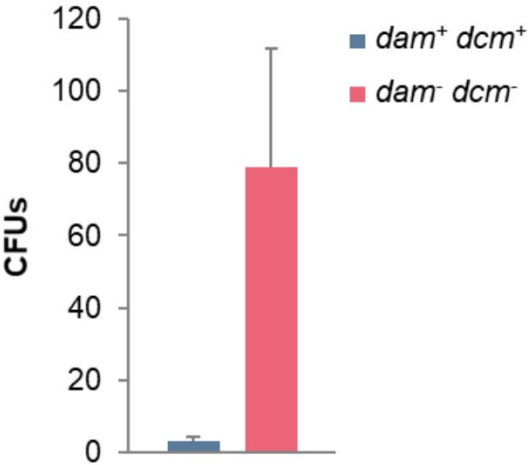
Figure S4: Comparison of transformation efficiencies in *C. beijerinckii* DSM 6423. 20 µg of the pCas9$_{ind}$ was used with and without *E. coli* Dam and Dcm methylations in triplicate experiments. Error bars represent the standard error of the mean (s.e.m.) of these replicates.

Table S1: list of primers used in this study

Name	SEQUENCES 5' --> 3'
Construction of pE_Xcas9	
P1_xylB_F	<u>TATTACGCCAGCTGGCGAAAGGGGGATG</u> AAATAAAAGGTTATTTTGCATTGACAAAG
Cas9_xylB_F	<u>GTACTTTTTATCCAT</u> CCCTCCTTGAATGCCACTTCATAACTATCG
xylB_cas9_F	<u>GGCATTCAAGGAGGG</u> ATGGATAAAAGTACAGTATTGGTCTAG
16-P2_cas9R	<u>GCTGGCCTTTTGCTCACATGTTC</u> TTTCTAAAAAAATTCTTAATCTCCGCCTAGTTGAC
Construction of pS_X^RKR	
P1_xylR_R	<u>TATTACGCCAGCTGGCGAAAGGGGGATG</u> AAATAAAAGGTTATTTTGCATTGACAAAG
bsaI_K_XylB_R	<u>CATAACACCCCTTGTAGGTCTC</u> CCTCCTTGAATGCCACTTCATAAC
bsaI_X_kan_F	<u>CATAACACCCCTTGTAGGTCTC</u> CCTCCTTGAATGCCACTTCATAAC
bsaI_R_Kan_R	<u>CTAGCTCTAAAACCGAGACC</u> CCAATTCTGATTAGAAAAACTC
bsaI_traCRNA_F	<u>AGAATTGGGGTCTCGGTTTTAGA</u> GCTAGAAATAGCAAGTTAAAATAAG
MCS_Term_R	AGATCTCGAGGCCTGCAGGCATGCAAGCTTGAATTTTTTTAAAAAATAACTCTGTAG
Construction of pS_X^R_KR_*celA*	
P1_hbd_F	AAAAAATTCAAGCTTGCATGCCTGCAGGCCATTTTAGCTAAGGATATCTATAATG
0326_celA_R	<u>AGTTATTTATATATTTCCAT</u> GATTACGAATTCGAGCTCGGTACCC
celA_0326_F	<u>CGAATTCGTAATCATG</u> GAAATATATAAATAACTACTAGATTACATAAATAAG
P2_0326_R	CGCGTCCATGGAGATCTCGAGGCCTGCAGG AGATGCTCCTGTCTCCTGTATTTC
CelA_S1_Forw	<u>GGAG</u> TATGGAAAAACTCAAGTTTATGCTT
CelA_S1_Rev	<u>AAAC</u> AAGCATAAACTTGAGTTTTTCCATA
Construction of pS_X^R_Δ*spoIIE*	
P1_Trna_F	TATTTTTTAAAAAAATTCAAGCTTGCATGC TCGCGGGGTGGAGCAGTTGGTAGC
16_0098_TRNA_R	<u>TTTTGAAACTACTAC</u> TTCCAAATAATGTAACTTGATTATAGCATC
16_TRNA_cbei_0098_F	<u>TTACATTATTTGGAA</u> GTAGTAGTTTCAAAAGTTTATTTAGATAGC
P2_0098_R	TACGAATTCGAGCTCGGTACCCGGGGATCC ATAAACCAGCGGATAATTTTTCAAGG

Name	SEQUENCES 5' --> 3'
spoIIE_S1_Forw	<u>GGAG</u> ATTAATACTAATTACTTCTTTTCTT
spoIIE_S1_Rev	<u>AAAC</u> AAGAAAAGAAGTAATTAGTATTAAT
Construction of pSpoIIE	
Cbei0097prom_F	GACGTTGTAAAACGACGGCCAGTGCCAAGC ACGAATTTTGGTCACTTAAGCCAAG
Cbei0097_R	AGTCGACGTCACGCGTCCATGGAGATCTCG CTTATTGCATGATAATCTTAGCTATC
Construction of pCas9$_{ind}$-ΔcatB	
ΔcatB_fwd	TGTTATGGATTATAAGCGGCTCGAG GACGTCAAACCATGTTAATCATTGC
ΔcatB_rev	AATCTATCACTGATAGGGACTCGAG CAATTTCACCAAAGAATTCGCTAGC
Construction of pCas9$_{ind}$-ΔpNF2	
pNF2-gRNA	GAGCTCACTCTATCATTGATAGAGTTTGAAACTCTATCATTGATAGAGTATAATATCTTTGTTCATTTAAGCC ATCTACTAAACAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG TGGCACCGAGTCGGTGCTTTTTTTGAAGCTTGAGCTC
Primer used for screening	
Cbei_325_F (*celA*)	ATAAGTAACAATTAGATAATTATGAAGTTAATCCTTAG
Cbei_326_R (*celA*)	CATTTGCTTTCAGGTCTTCTTTTGCTG
Cbei_0096_F (*spoIIE*)	CGAAGATATTATGTCTAAGTTTCTA
Cbei_0098_R (spoIIE)	CATTACATTCCATACAATATTTATTGTATAAACCAGC
RH076 (*catB*)	CATATAATAAAAGGAAACCTCTTGATCG
RH077 (*catB*)	ATTGCCAGCCTAACACTTGG
pNF2_fwd (pNF2)	GGGCGCACTTATACACCACC
pNF2_rev (pNF2)	TGCTACGCACCCCCTAAAGG
Bold: restriction sites (BsaI); <u>Underlined</u>: non-specific tail;	

Table S2: Final products concentration in g.L-1 after a 72 h-fermentation experiments in 100 mL bottles of *C. beijerinckii* NCIMB 8052 wild-type and *ΔspoIIE* strain. Values are given ± SEM of biological duplicates. Fermentations were performed in CM2 medium with 50 g glucose.

	Concentration (g.L^{-1})					
	Acetate	Butyrate	Acetone	Butanol	Ethanol	Total ABE
Wild-type	0.4 ± 0.0	0.1 ± 0.0	2.0 ± 0.5	7.7 ± 0.6	n.d	9.7 ± 0.6
ΔspoIIE	0.4 ± 0.0	0.8 ± 0.0	2.7 ± 0.3	7.0 ± 0.1	n.d	9.7 ± 0.3

Table S3: Final solvent concentration of 72h-fermentation experiments of C. beijerinckii DSM 6423 and derivative strains. Values are given ± SEM of biological triplicates. Fermentations were performed in Gapes medium (per liter: 2.5 g yeast extract, 1 g KH2PO4, 0.6 g K2HPO4, 1 g MgSO4 7H2O, 6.6 mg FeSO4 7H2O, 100 mg p-aminobenzoic acid, 2.9 g ammonium acetate, 60 g glucose).

	Concentration (g.L^{-1})				
	Ethanol	Acetone	Isopropanol	Butanol	Total
Wild-type	0.1 ± 0.0	0.2 ± 0.0	4.1 ± 0.1	7.0 ± 0.1	11.5 ± 0.1
ΔcatB	0.1 ± 0.0	0.1 ± 0.0	3.9 ± 0.1	6.8 ± 0.2	10.9 ± 0.3
ΔcatB ΔpNF2	0.1 ± 0.1	0.1 ± 0.0	4.5 ± 0.1	7.8 ± 0.5	12.5 ± 0.5

Chapter 3-

Transcriptomic and phenotypic analysis of a *spoIIE* mutant in *Clostridium beijerinckii*

Mamou Diallo[1,5]*, Nicolas Kint[2], Marc Monot[3], Florent Collas[1], Isabelle Martin-Verstraete[2,4], John van der Oost[5], Servé W. M. Kengen[5], Ana M. López-Contreras[1]

[1]Wageningen Food and Biobased Research, Wageningen, The Netherlands
[2] Laboratoire Pathogenèse des Bactéries Anaérobies, Institut Pasteur, UMR CNRS 2001, Université de Paris, Paris, France
[3] Biomics platform, C2RT, Institut Pasteur, Paris, France
[4] Institut Universitaire de France, Paris, France
[5] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

3

Abbreviations: ABE, Acetone Butanol Ethanol; **CRISPR**, Clustered Regularly Interspaced Short Palindromic Repeats; **Cas**, CRISPR-associated protein; **COG**; Clusters of Orthologous Groups

Abstract

SpoIIE is a phosphatase involved in the activation of the first sigma factor of the forespore, σ^F, during sporulation. A $\Delta spoIIE$ mutant of *Clostridium beijerinckii* NCIMB 8052, previously generated by CRISPR-Cas9, did not sporulate but still produced granulose and solvents. Microscopy analysis also showed that the cells of the $\Delta spoIIE$ mutant are elongated with the presence of multiple septa. This observation suggests that in *C. beijerinckii,* SpoIIE is necessary for the completion of the sporulation process, as seen in *Bacillus* and *Clostridium acetobutylicum*. Moreover, when grown in reactors, the *spoIIE* mutant produced higher levels of solvents than the wild-type strain. The impact of the *spoIIE* inactivation on gene transcription was assessed by comparative transcriptome analysis at three time points (4 h, 11 h and 23 h). Approximately 5% of the genes were differentially expressed in the mutant compared to the wild-type strain at all time points. Out of those only 12% were known sporulation genes. As expected, the genes belonging to the regulon of the sporulation specific transcription factors (σ^F, σ^E, σ^G, σ^K) were strongly down-regulated in the mutant. Inactivation of *spoIIE* also caused differential expression of genes involved in various cell processes at each time point. Moreover, at 23 h, genes involved in butanol formation and tolerance, as well as in cell motility, were up-regulated in the mutant. In contrast, several genes involved in cell wall composition, oxidative stress and amino acid transport were down-regulated. These results indicate an intricate interdependence of sporulation and stationary phase cellular events in *C. beijerinckii*.

Introduction

Even though butanol is nowadays mainly produced through the petrochemical route, it used to be made industrially by a bioprocess called ABE fermentation in the first half of the 20th century. This process returned to the forefront at the end of the 1990s with the emerging interest in biobased chemicals. ABE fermentation relies on the ability of several bacteria from the *Clostridium* genus to convert carbohydrates to acetone, ethanol, butanol (ABE) and isopropanol. Clostridia are anaerobic bacteria that can form spores to protect themselves from unfavorable environmental conditions, including oxygen exposure. The main representatives of the solventogenic clostridia group are *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccharobutylicum* and *Clostridium saccharoperbutylacetonicum*. These clostridia produce solvents while they form spores. Once the spores are mature, the solvent-producing cells lyse, and the metabolically inactive spores are left behind (23). That is why in industry, spores are seen as undesirable (2, 160), and many efforts were made to engineer asporogenous solvent-producing strains (15–17). However, the sporulation process and the associated regulatory network in these microorganisms are still poorly characterized (18).

In solventogenic clostridia, several studies revealed a link between sporulation and solvent production (17, 23) but, none was able to explain the involved mechanism. The regulatory pathway controlling sporulation was first described and intensively studied in *Bacillus subtilis,* which is considered as a model organism for the sporulation process. Comparative studies between bacilli and clostridia show similarities in the sporulation process and its regulation, including the presence of the main actors such as Spo0A and σ^H as well as the four sporulation specific sigma factors, σ^F, σ^E, σ^G and σ^K (23). However, differences in the sporulation regulatory networks are also observed between bacilli and clostridia and even among clostridia (23). Substantial deviations from the *B. subtilis* paradigm exist in clostridial spore formers, especially concerning the communication between the forespore and the mother cell, a weaker connection between gene expression and morphogenesis, and modifications in the interplay between sigma factors (101, 224).

In *B. subtilis*, the SpoIIE protein is a phosphatase that plays a crucial role in the sporulation regulation mechanism (146). SpoIIE acts in stage II of the sporulation process. SpoIIE plays a central role in the asymmetric septum formation separating the mother cell and the forespore. Studies in *B. subtilis* showed that SpoIIE interacts with cell division proteins and peptidoglycan synthesis proteins to enable a correct localization and thickness of the asymmetric septum (225–227). Following the asymmetric division, SpoIIE enables the activation of σ^F (Figure 1), the first sigma factor of the forespore (228, 229). σ^F is held inactive by the anti-sigma factor and kinase SpoIIAB. At the beginning of stage II of the sporulation cascade, SpoIIE dephosphorylates the anti-anti-sigma factor SpoIIAA to enable its interaction with SpoIIAB, which releases σ^F (Figure 1) following asymmetric division

3

(23). In *C. acetobutylicum,* considered as the model solventogenic *Clostridium*, SpoIIE has also been reported to function as a phosphatase involved in the early stages of the sporulation regulation cascade. Previous studies have also shown that *spoIIE* mutants of *C. acetobutylicum* were asporogenous but still produced solvents (15, 16). However, no studies to date confirmed if this model can be applied to other ABE-producing strains.

C. beijerinckii is the second most studied solventogenic species and was used industrially for acetone production already at the beginning of the 20th century (4). *C. beijerinckii* is known for its ability to catabolize a wide range of carbohydrates (230, 231). However, few studies on the sporulation mechanism have been conducted in this species. We have recently inactivated the *spoIIE* gene (Cbei_0097) in *C. beijerinckii* NCIMB 8052 using a CRISPR-Cas9 system for Clostridia developed in our laboratory (119). In this work, we analyzed the impact of *spoIIE* inactivation in *C. beijerinckii* NCIMB 8052 on the sporulation process and the solvent production through fermentation assays, microscopy observations and transcriptome analysis**.**
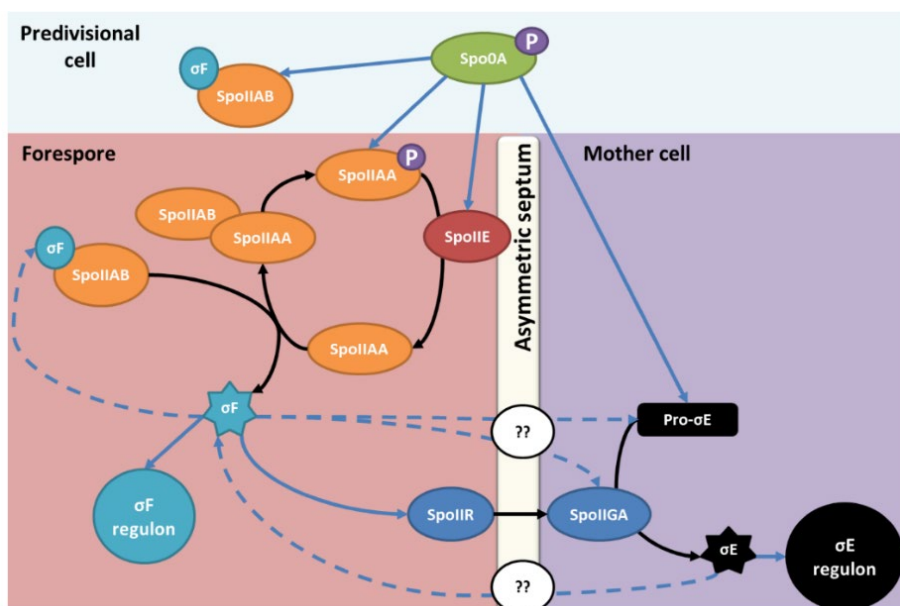
Figure 1: SpoIIE, an essential protein for sporulation in *B. subtilis* and *C. acetobutylicum* in stage II of sporulation. After sporulation initiation, the expression of *spoIIE* and the *spoIIAA-spoIIAB-sigF* operon is promoted by phosphorylated Spo0A. In *B. subtilis* and *C. acetobutylicum*, SpoIIE dephosphorylates SpoIIAA, the anti-anti sigma factor that sequestrates σF. The dephosphorylated SpoIIAA binds to SpoIIAB and enables the release of σF in the prespore. The blue arrows indicate an impact on the gene transcription; dotted arrows are only valid for *C. acetobutylicum*. In contrast to what is observed in Bacillus, σF and σE impact events in both the prespore and the mother cell compartment in *Clostridium*; however, the mechanism enabling the interaction is unknown. Adapted from Al-Hinai et al. (23) and amended with permission from the American Society for Microbiology.

Methods

Bacterial strains and culture conditions

Bacterial strains and plasmids are listed in Table 1. The *C. beijerinckii* wild-type (WT) strain was stored as spore suspension and the mutants as vegetative cells in 15% glycerol solution at – 20°C. Spore suspensions were heat-shocked 1 min at 98°C before inoculation in a liquid medium to kill any vegetative cell present and enable the germination of the spores. Except for fermentation assays, liquid cultures of the WT, mutant and complemented strains were grown in liquid modified CGM (mCGM) as described previously (68) containing per liter: yeast extract, 5 g; KH_2PO_4, 0.75 g; K_2HPO_4, 0.75 g; asparagine•H_2O, 2 g; $(NH_4)_2SO_4$, 2 g; cysteine, 0.50 g; $MgSO_4$•7 H_2O, 0.40 g; $MnSO_4$•H_2O, 0.01 g; $FeSO_4$•7 H_2O, 0.01 g; glucose, 10 g. Liquid media were made anaerobic by flushing with nitrogen gas. Cultivation was performed at 35°C anaerobically in an anaerobic chamber without shaking (Sheldon Manufacturing, USA). The gas mixture used consists of 15% CO_2, 4% H_2 and 81% N_2. *Clostridium* strains were grown and selected on modified CGM agar described previously (188) containing per liter: yeast extract, 1 g; tryptone, 2 g; KH_2PO_4, 0.50 g; K_2HPO_4, 1 g; $(NH_4)_2SO_4$, 2 g, $MgSO_4$•7 H_2O, 0.10 g; $MnSO_4$•H_2O, 0.01 g; $FeSO_4$•7 H_2O, 0.015 g; $CaCl_2$, 0.01 g; $CoCl_2$, 0.002 g; $CaCl_2$, 0.002 g; glucose, 50 g and agar 12 g. *Escherichia coli* XL1 blue cells (Agilent, USA) were used for cloning. *E. coli* strains were stored at -80°C in the presence of 15% glycerol. *E. coli* was grown on LB medium, supplemented with 100 µg/mL spectinomycin (Duchefa, The Netherlands).

Plasmid construction

The primers used for plasmid construction are listed in Table S1 and were synthesized by Integrated DNA Technologies. To replace the *catp* gene by *aad9* in pRAN473, pRAN473 was linearized by PCR using the primers M392 and M393. The *aad9* gene was amplified from pSpoIIE with the primers M394 and M395. The fragments were fused using the Circular Polymerase Extension Cloning (232) to generate pRAN473S. pRAN473S::cbei_0097 was constructed according to the protocol from (233). The *spoIIE* gene (Cbei_0097) was amplified from pSpoIIE by PCR using M396 and M397. The obtained PCR product and pRAN473S were digested by Sal1 and BamH1. The fragments were ligated using the T4 DNA ligase (New England Biolabs, USA) according to the manufacturer protocol.

Table 1: Strains and plasmids used in this study. catP: chloramphenicol resistance gene, aad9: spectinomycin resistance gene

Strains or plasmids	Relevant characteristics	Reference
Strains		
C. beijerinckii NCIMB 8052 (WT)	Wild-type, sensitive to spectinomycin (650 µg/mL) and erythromycin (25–50 µg/mL)	NCIMB
C. beijerinckii Δ*spoIIE*	NCIMB 8052, Δ*spoIIE*	(119)
E. coli XL1-blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIq ZΔM15 Tn10 (Tetr)]	Agilent
Plasmids		
pSpoIIE	pCB102, colE1, aad9, Cbei_0097	(119)
pRAN473	repA, colE1, Ptet::mCherryOpt–MCS, catP	(234)
pRAN473S	repA, colE1, Ptet::mCherryOpt–MCS, aad9	This study
pRAN473S::cbei_0097	repA, colE1, Ptet::mCherryOpt–cbei_0097, aad9	This study

Fermentation

Fermentations were performed at 35°C in CM2 medium (230), which contains per liter: yeast extract, 5 g; KH_2PO_4, 1 g; K_2HPO_4, 0.76 g; ammonium acetate, 3 g; *p*-aminobenzoic acid, 0.10 g; $MgSO_4 \cdot 7\,H_2O$, 1 g; and $FeSO_4 \cdot 7\,H_2O$, 0.50 g, glucose, 60 g. Metabolites were determined in culture supernatants after removal of cells by centrifugation. Glucose, acetate, butyrate, lactate, acetone, butanol and ethanol titers were determined by high-performance liquid chromatography (HPLC) as described previously using 4 methyl valeric acid (30 mM) as an internal standard (68, 180).

DNA extraction and sequencing

Genomic DNA of *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* Δ*spoIIE* mutants was purified using the GenElute bacterial genomic DNA kit (Sigma-Aldrich, USA). The concentration of genomic DNA was determined using a nanodrop spectrophotometer (Thermo Fisher Scientific, USA) and quality checked on 1 % agarose gel. PCR reactions were carried out using the Q5 Master mix (New England Biolabs, USA). DNA sequencing of clones and genome assembly were performed by BaseClear (Leiden, The Netherlands). The sequences of the WT and the Δ*spoIIE* clones were compared to the publicly available sequence of *C. beijerinckii* NCIMB 8052 on NCBI. SNPs between the genome of our WT and the Δ*spoIIE* mutant's genome with a frequency above 98% were considered in our study.

Granulose staining

Granulose accumulation was monitored by iodine staining. Each *C. beijerinckii* strain was grown on CM2 agar plates and incubated anaerobically at 37°C. After 24 h of incubation, the petri dish was opened and inverted over I_2 crystals for approximately 1 min. The

colonies of granulose-negative mutants were unstained by the sublimed I_2 vapors, while the granulose-positive strains were labeled (90, 235).

Spore viability assay

To verify the presence of viable spores, overnight cultures of each strain were prepared in 5 mL CM2 medium. The next day, two tubes containing 15 mL of fresh CM2 liquid medium were inoculated with 50 µL of overnight culture. Aliquots of 100 µL were collected at 24 h and 48 h, treated at 98°C for 1 min, and used for 4 serial dilutions. 50 µL of each dilution was spread on CM2 agar plates in the anaerobic chamber and incubated for 24 h at 37°C. The colonies on each plate were then counted to determine the average number of viable spores per mL (spores/mL). This method was adapted from Steiner et al. (108).

Microscopy analysis

Phase-contrast microscopy (Olympus BX51) was used to observe the morphology of WT and Δ*spoIIE* strains at ×400 and x1000 magnifications. Cells were cultivated for 72 h in liquid CM2 medium, samples were collected at 48 and 72 h, and centrifuged to stain them according to the Schaeffer-Fulton technique. A cell film was made on a glass slide and stained with malachite green and safranin to visualize spores (17, 236)**.** The stained cells were observed by phase-contrast microscopy at x1000 magnification to detect the presence of spores. For the fluorescent microscopy analysis, cells were washed three times by centrifugation (5000 g, 3 min) and resuspended in 1 mL of PBS. Following washing, the cells were resuspended in 1 mL of PBS supplemented with the membrane dye Mitotracker Green (MTG, 0.5 mg.mL^{-1}) (Molecular Probes, Invitrogen Thermo Fisher Scientific, USA). Cells were mounted on 1.7% agarose coated glass slides and observed in a home-built and designed microscope. Fluorescent signals were visualized with a Nikon CFI SR HP Apochromat TIRF 100XC Oil objective, and images were captured using an Andor Zyla 4.2 PLUS sCMOS camera. To observe the localization of the SpoIIE protein, the *spoIIE* gene was fused to a mCherry coding sequence, and the translational fusion was expressed under the control of the Ptet promoter inducible in the presence of anhydrotetracycline (atc). *C. beijerinckii* Δ*spoIIE* cells harboring the pRAN473S and pRAN473S::cbei_0097 were grown in CM2 containing 650 µg/mL of spectinomycin. After 10 h of growth, atc (Sigma-Aldrich, USA) was added at 200 ng/mL. Samples for microscopy analysis were collected 8 h after induction. Before MTG staining, the cells were fixed as previously described (233).

Transcriptome analysis

RNA isolation and sequencing protocol

The same procedure was repeated three times in three different weeks to obtain three independent biological replicates. Each week, a fresh preculture was used to inoculate two identical bioreactors. The cultures were grown, as described in the method section. Samples were taken over the early exponential, late exponential and stationary phases (samples at 4, 11 and 23 h). Following centrifugation of the samples, cell pellets were washed with chilled RNase free water and resuspended in RNase free water to obtain a suspension having an OD_{600nm} of approximately 1. A 3-mL diluted sample was centrifuged, the supernatant was discarded, and the cell pellet was stored at -80°C for subsequent isolation. Frozen samples were thawed on ice, and RNA was isolated using High pure RNA isolation kit (Roche Diagnostics, Switzerland). Quality and concentration of RNA samples were checked using a nanodrop spectrophotometer (Thermo Fisher Scientific, USA). The absence of DNA in the RNA samples was evaluated by qPCR analysis performed with BioRad CFX 96 Touch™ (BioRad, Hercules, USA) and the PowerUP SYBr green reaction mix (Applied Biosystems, Thermo Fisher Scientific, USA). Reactions were performed in an overall volume of 10 μL with concentrations of components and reaction conditions, as described in the master mix protocol. RNA quality and integrity were determined using the Qsep 100 bioanalyzer (Bioptic Inc., Taiwan).

All the RNA samples collected were used for library construction and sequencing. The Ribo-zero kit (Illumina, USA) was used to enrich the samples in mRNA. The stranded library was prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina, USA) according to the manufacturer's recommendation. AMPure XP beads (Beckman Coulter, USA) were used to clean up the cDNA fragments after each process. Library quality was checked using the bioanalyzer, and the library was then loaded onto HiSeq 2500 for high-throughput sequencing.

Bioinformatics analysis

After Illumina sequencing, all the reads were mapped to the *C. beijerinckii* genome using Bowtie (237) then converted into BAM files with the Samtools (238). Differential analysis of RNA-seq data and statistical analysis were performed with Sartools pipeline (239) using the DESeq2 package. The data were visualized in a strand-specific manner using COV2HTML (http://mmonot.eu/COV2HTML/) (240). Genes were considered differentially expressed when padj < 0.05 and gene expression with a | $\log_2$ fold change |>1.5. Gene expression was considered down-regulated if $\log_2$ fold change < 1.5 or up-regulated if $\log_2$ fold change >1.5.

RT-qPCR analyses were performed with BioRad CFX 96 Touch™ and the PowerUP SYBr green reaction mix to confirm the RNA-seq results. Primer3 website was used for

oligonucleotide design (Table S2). Relative expression at 23 h of 12 genes was monitored (Table S2); gene *gyrA* was chosen as the reference gene from a selection of candidate genes (including *gyrA, 16S rRNA, polIII* and *alaS*) based on analysis by RefFinder algorithms to verify the stability of their expression. Reaction efficiency was determined for each assay using a ×5 serial dilution of cDNA samples. A sample collected after 4 h of cultivation in CM2 medium was chosen as a calibrator and a reference value for calculation of expression fold-change of each gene in other samples. All RT-qPCR analyses were performed in triplicate. The relative quantification was evaluated using the mathematical model described by Dr. Pfaffl (241).

Results

SpoIIE disruption and mutant characterization during the sporulation cycle

The *spoIIE* gene (Cbei_0097) from *C. beijerinckii* NCIMB 8052 is 2412 bp long and encodes an 803 amino acid long transmembrane protein. The *spoIIE* gene is present in all the strains of *C. beijerinckii* with a genome available. In all *C. beijerinckii* strains, *spoIIE* shows a very high level of identity ranging from 96% to 100%. SpoIIE orthologs were found in all spore-forming Bacilli and Clostridia (98). In *Clostridium*, its role was studied only in *C. acetobutylicum* ATCC 824, where SpoIIE (Cac_3205, 795 amino acids) was reported to be involved in sporulation but not to affect directly solvent production (15, 16). SpoIIE of *C. beijerinckii* NCIMB 8052 (cbei_spoIIE) shows 36% identity [62% similarity] with SpoIIE of *C. acetobutylicum* ATCC 824, 26% identity with SpoIIE of *B. subtilis* and 42% identity with SpoIIE of *Clostridium perfringens*. However, compared with SpoIIE of *C. acetobutylicum*, the protein of *C. beijerinckii* is longer. An extra transmembrane helix at the N terminus of the *C. beijerinckii* protein (Figure 2A) was predicted by the TMHMM software (242). This region is conserved in *B. subtilis* and *C. perfringens*. Synteny of the genomic region, in which the *spoIIE* gene is located, was observed in *Clostridium* and *Bacillus* (Figure 2B). Indeed in *C. beijerinckii* NCIMB 8052 and *C. acetobutylicum* ATCC 824, *spoIIE* is in the middle of a conserved genomic region larger than 30 kb harboring more than 20 genes (Figure 2B). Moreover, SpoIIE inactivation in *B. subtilis* and *C. acetobutylicum* resulted in an asporogenous phenotype (15, 219, 243), and the bacteria could not complete stage II of the sporulation cycle (23). All these observations suggest that the protein has a similar function in *Bacillus* and *Clostridium*.
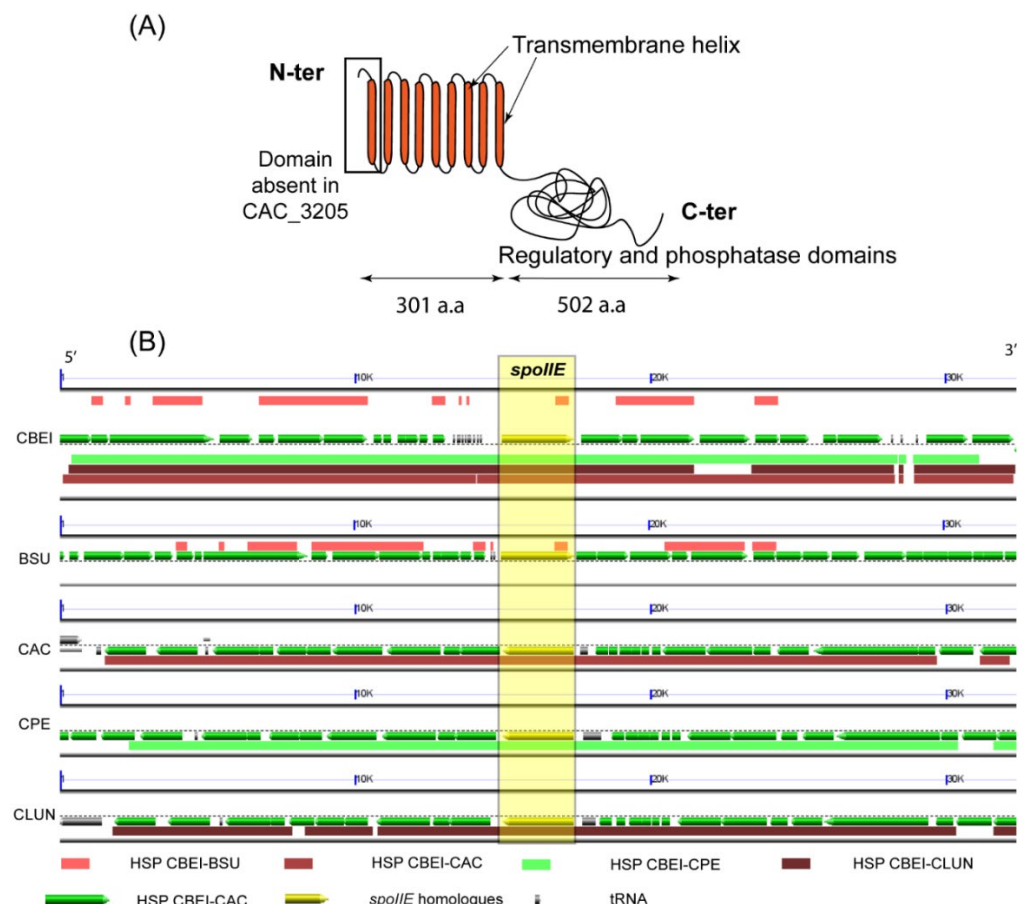
Figure 2: SpoIIE a sporulation protein conserved in spore-forming Firmicutes. (A) Illustration of the predicted SpoIIE protein with its transmembrane regions (according to the TMHMM software) and its phosphatase domain terminus, a.a: amino acids (B) Synteny map of the spoIIE region from Cbei_0084 to Cbei_0107, with spoIIE (Cbei_0097), in *C. beijerinckii* NCIMB 8052 (CBEI) and its homologs in *B. subtilis* str.168 (BSU), *C. acetobutylicum* ATCC 824 (CAC), *C. perfringens* str.13 (CPE) and *C. ljundahlii* ATCC 49587 (CLUN), the homologous regions (HSP) are highlighted by color blocks for each organism. Image generated using the COGE platform https://genomevolution.org/r/1ceci, and the GEvo tool (244)

Using a xylose inducible CRISPR-Cas9 system, we recently constructed a *spoIIE* mutant by deleting a 2.379 kb fragment of the coding sequence of *spoIIE* (119). The genomes of four independent mutants were sequenced to confirm the *spoIIE* deletion and the absence of any other mutations. Compared to the genome of the WT strain used in our laboratory, the mutants' genomes harbored between 7 and 23 SNPs (Table S3). The C5 mutant harbored the fewest modifications. In this mutant, only 7 SNPs were detected. Out of these SNPs, 4 were located in CDS, of which 3 were found in the genome of the other mutants. The mutant C5 was selected for further studies of the *spoIIE* mutant. After 24 h and 48 h of growth in liquid medium, samples of C5 and WT cultures were collected and incubated at 98°C for 1 min and then plated on solid medium to determine their sporulation efficiency.

This heat treatment kills the vegetative cells and induces spore germination. No viable spores were detected in the C5 culture, as no colonies grew on the respective plates, in contrast to the growth observed in the plates corresponding to the WT cultures (Figure 3A). Sporulation capacity was partially restored once the C5 mutant was complemented (Figure S1) with a plasmid carrying a functional *spoIIE* gene with its native promoter (119), confirming the direct link between *spoIIE*'s disruption and the mutant's asporogenous phenotype. After incubation of the strains for 24 h on plates and iodine staining, we observed that both the mutant and WT strains produced granulose (Figure 3B). Moreover, when grown in serum bottles in CM2 medium, the Δ*spoIIE* mutant metabolized glucose and produced solvents even if the global amount of solvent was reduced and the production of acid increased as compared to the WT strain (Figure 3C). SpoIIE inactivation abolished sporulation but did not prevent granulose nor solvent production.
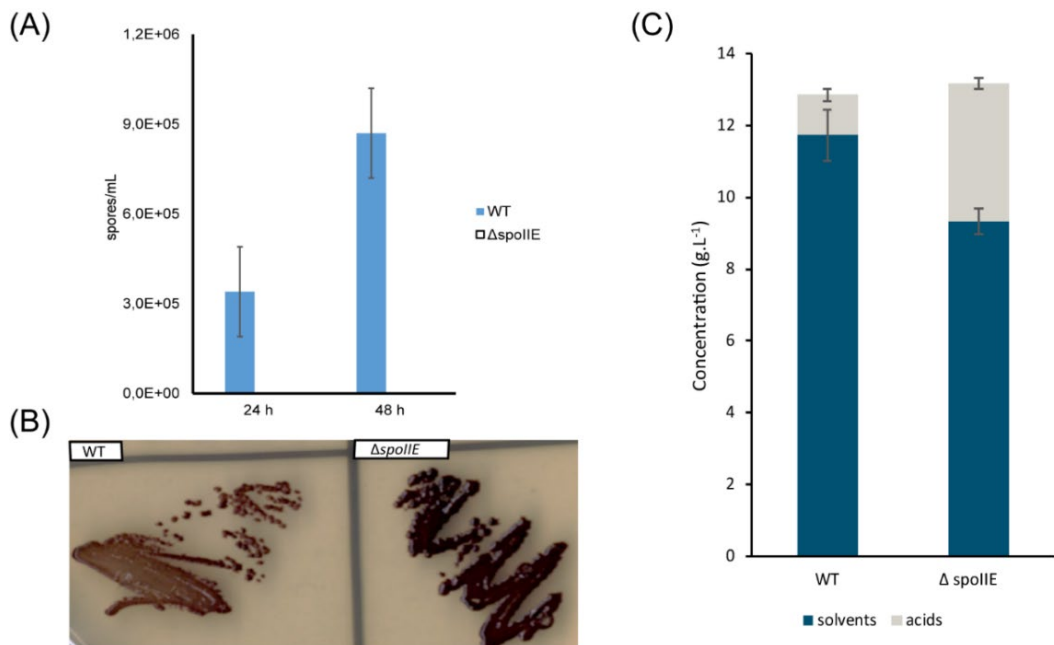
Figure 3: Phenotypic comparison of Δ*spoIIE* mutant and WT. The Δ*spoIIE* mutant could not produce viable spores but still produced granulose and solvents. (A) Spore viability assay; (B) Granulose detection after 24 h of incubation; (C) Fermentation end products after 72 h of culture at 35oC in 50-mL serum bottles, the error bars indicate one standard deviation of the mean, which was determined based on the data from biological duplicates (n=2).

We observed the morphology of the cells after 6 h, 24 h, 35 h, 48 h and 72 h of cultivation with phase-contrast microscopy (Figure 4) and after 20 h with fluorescent microscopy (Figure 5A, Figure S2A-B). While mature spores were detected after 24 h of growth in WT cultures, no regular pre-spores nor spores were seen in Δ*spoIIE* mutant cultures, even after 72 h of incubation (Figure 4A). At 48 h and 72 h, several mutant cells harbored phase-dark masses at the poles of the cells (Figure 4B black arrows) that were absent in WT cells. Even in the complemented Δ*spoIIE* mutant, in which sporulation was restored, some cells still harbored these phase-dark bodies at the poles (119). These masses were stained by malachite green (Figure 4B) and could be a result of DNA condensing in forespore-like compartments, as observed in the *sigE* mutant of *C. acetobutylicum* (109) and *spoIIE* mutants of *B.subtilis* (219). Moreover, mutant cells had a different morphology compared to WT cells, from 20 h of cultivation onwards, as observed by fluorescent microscopy after MTG labeling. The cells of the Δ*spoIIE* mutant became filamentous and seemed to contain several septa (Figure 5A, Figure S2A-B). In *C. beijerinckii, spoIIE* disruption seems to impact the division site for asymmetric septation, as observed in *B. subtilis* (219).

To localize the SpoIIE protein in the cells, SpoIIE was fused to the mCherry fluorescent protein (234) and introduced into the Δ*spoIIE* strain. After 10 h of cultivation, atc was added in the cultures inducing the expression of the mCherry protein in both Δ*spoIIE* strains harboring, either pRAN73S or pRAN73S::cbei0097 plasmid. After induction, the cells were incubated for 8 h and then observed by fluorescent microscopy (Figure 5B, Figure S2C). As described in *B. subtilis*, SpoIIE proteins seemed to gather mainly near the poles of the cell.

3
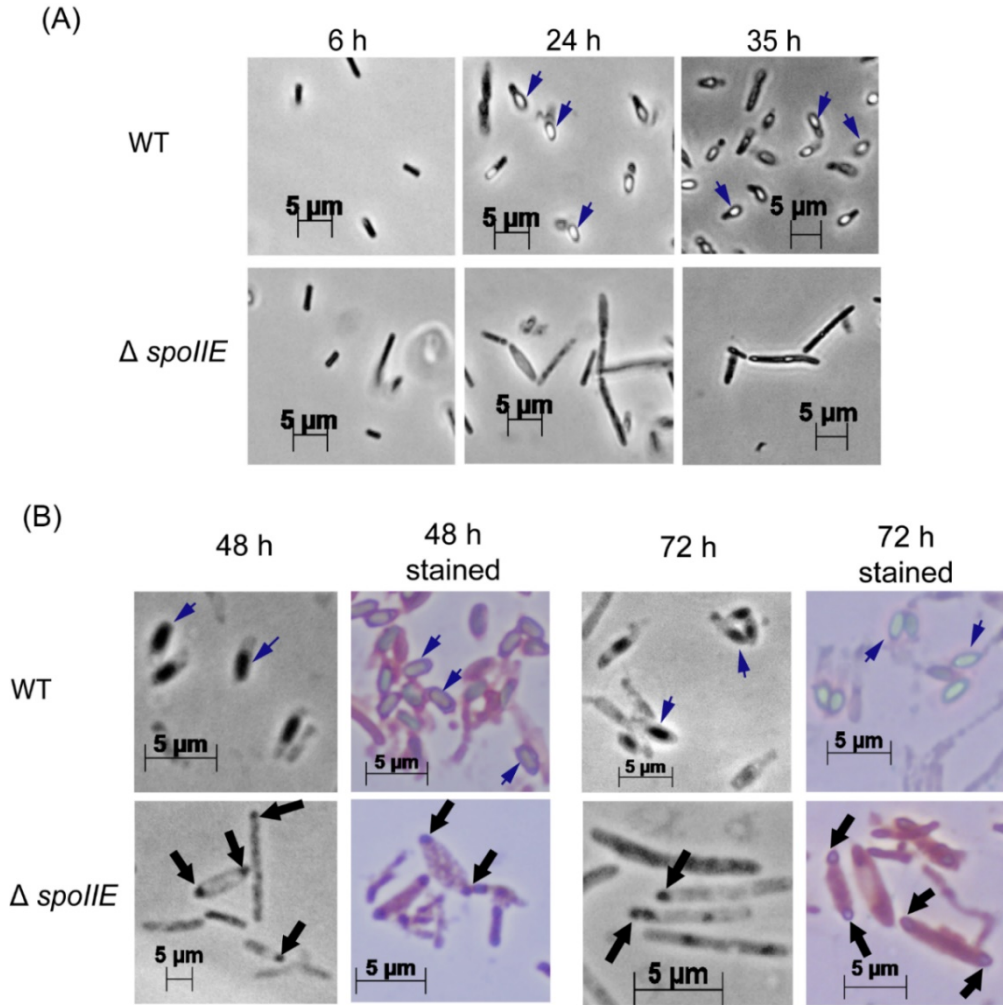
Figure 4: Phase-contrast microscopy images of *C. beijerinckii* WT and Δ*spoIIE* cells during the fermentation. While mature spores were seen in the WT images after 24 h of cultivation, no spores were seen in ΔspoIIE cultures even after 72 h of cultivation. However, phase-dark bodies, stained in blue by Shaeffer-Fulton staining, were observed in the ΔspoIIE cells after 48 h of cultivation. (A) Pictures with x400 magnification at 6 h, 24 h and 35 h; (B) Pictures at 48 h and 72 h of culture with x1000 magnification with and without Shaeffer-Fulton stain. The short dark blue arrows indicate mature spores and the black arrows indicate the phase-dark masses at the poles of the mutant cells.
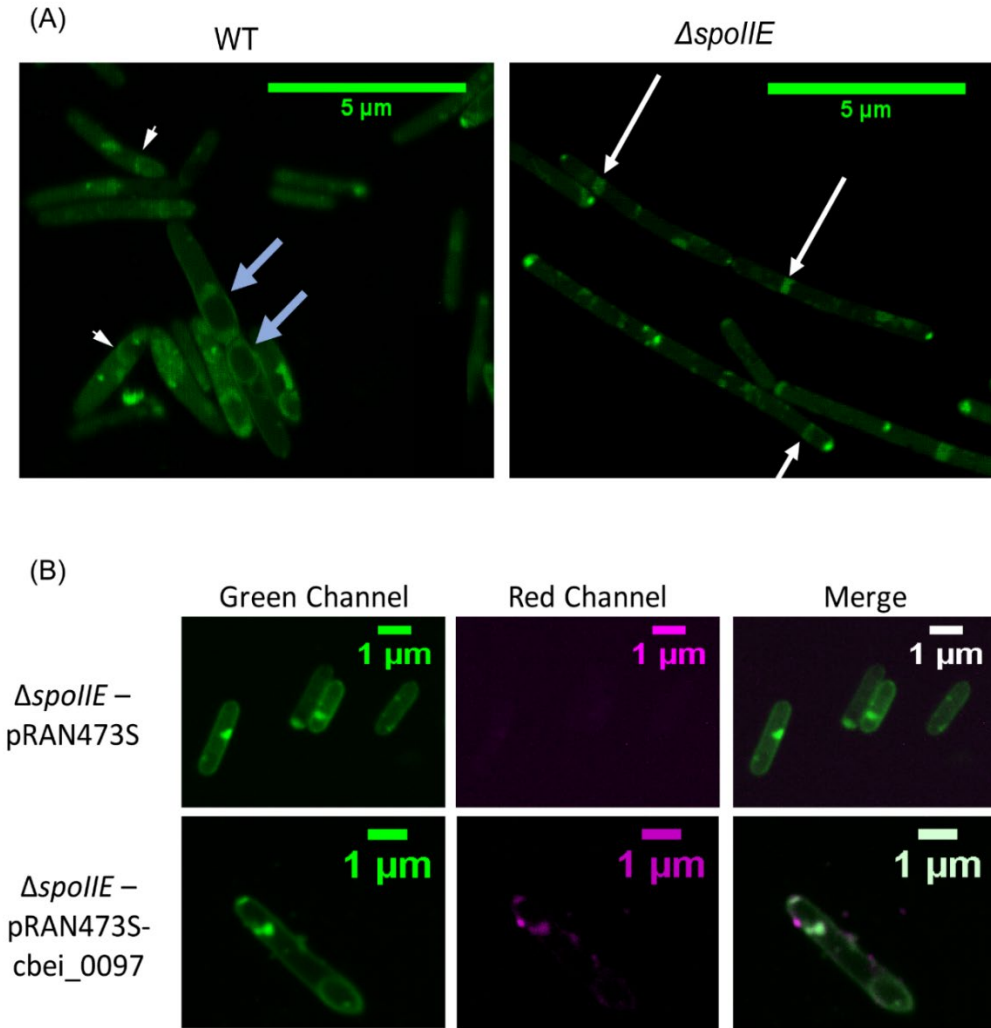
Figure 5: Fluorescence microscopy images of *C. beijerinckii* WT and Δ*spoIIE* cells. (A) Fluorescence images of the wild-type and the mutant cells after 20 h of cultivation and stained by the membrane staining MTG; (B) Fluorescence images of ΔspoIIE mutant cells stained by MTG and harboring either the mCherry empty plasmid (pRAN73S) or the plasmid expressing mCherry fused to Cbei_0097 (pRAN73S::cbei_0097) at 18 h of cultivation (after 8 hours of atc induction). The white arrows indicate septa observed in the mutant strain. The short white arrows indicate wild-type cells with an asymmetric septum. The light blue arrows indicate the prespores observed in the wild-type.

Comparison of the fermentation profile of the wild-type strain and Δ*spoIIE* mutant

The WT and Δ*spoIIE* strains were grown in a chemostat. Growth, substrate consumption, product formation and pH were monitored for 73 h (Figure 6 and Table 2). The disruption of *spoIIE* did not affect biomass creation as the mutant strain reached the stationary phase after 23 h, like the WT strain (Figure 6A). As observed in small-scale fermentations (Figure 3C), the mutant consumed glucose to produce acids and solvents. However, the pH in the Δ*spoIIE* mutant culture dropped slightly earlier than in the WT culture (Figure 6A). The reassimilation of the acids started earlier in the mutant culture as the solvent titer at 11 h was twice higher than in the mutant culture compared to the WT. After the reassimilation of butyrate and acetate, the pH at 11 h rose only to 5.7 +/- 0.1 in the mutant culture while reaching 6.1 +/- 0.0 in the WT culture. From 11 h to 23 h, the pH decreased in both cultures abruptly to 5.0 +/- 0.1 in the Δ*spoIIE* mutant culture and 5.3 +/- 0.0 in the WT culture. After 23 h, the pH increased slightly to 5.6 +/- 0.0 in the WT culture but still decreased in the mutant culture to 4.9 +/- 0.2. At the end of the fermentation, higher amounts of butyrate and acetate accumulated in the mutant culture. Indeed at 49 h, the acetate concentration reached 0.8 g.L^{-1} +/- 0.2 in the mutant culture and 0.3 g.L^{-1} +/- 0.0 in the WT culture. The concentration of acids after 73 h of cultivation in the chemostat was higher in the Δ*spoIIE* mutant culture, as observed in the serum bottles (Figure 3C). This difference is probably due to the absence of reassimilation of these acids in the Δ*spoIIE* mutant. The mutant also produced more acetone and butyric acid than the WT (Table 2); but, we did not observe an elongation of the solvent production phase but rather a switch to acidogenesis after 35 h.

Table 2: **Fermentation data on substrate consumption and product formation of cultures of WT and Δ*spoIIE* strains** after 30 h of cultivation in bioreactors, one standard deviation of the mean was determined based on the data from independent biological and technical duplicates (n=4) * acetate was also produced by the mutant strain, n.d = not detected

	Wild-type	Δ**spoIIE**
Substrates consumed [g.L^{-1}]		
Glucose	32.5 +/- 0.5	37.0 +/- 1.4
Acetate	1.1 +/- 0.1	0.8 +/- 0.1
Products at the end [g.L^{-1}]		
Acetate*	0.5 +/- 0.0	0.7 +/- 0.1
Butyrate	0.6 +/- 0.0	0.7 +/- 0.1
Lactate	0.5 +/- 0.1	0.1 +/- 0.0
Acetone	1.9 +/-0.1	2.7 +/- 0.2
Butanol	5.8 +/- 0.1	6.3 +/- 0.3
Ethanol	n.d	n.d
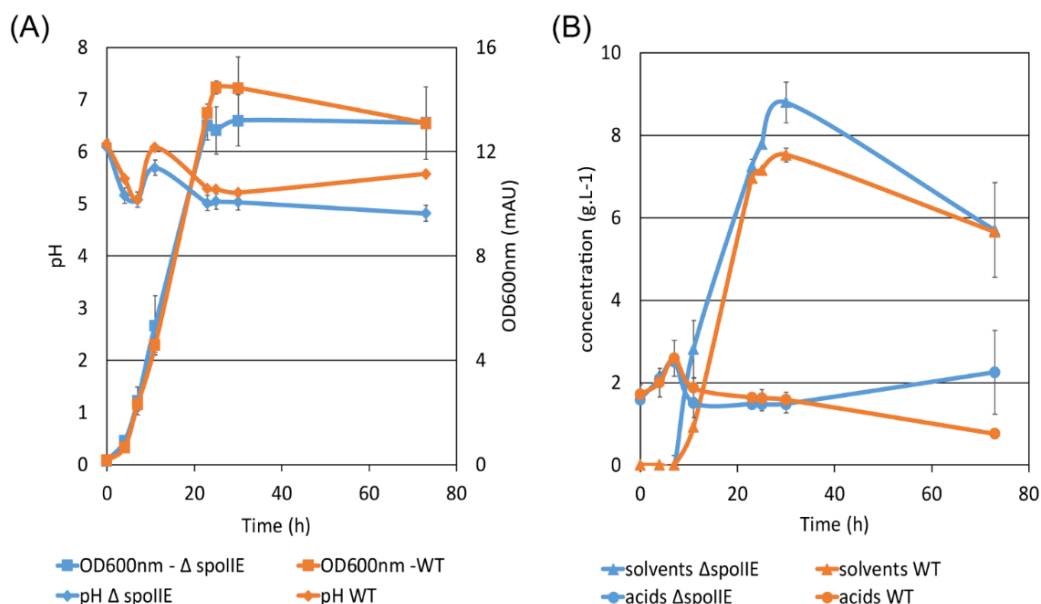
Figure 6: **Fermentation profiles of C. beijerinckii strains. Fermentations were performed in biological and technical duplicates in CM2 medium in chemostats at 35oC during 73 h of cultivation.** (A) Growth curve and pH variation in the cultures; (B) Acids and solvent titers during the fermentation. The error bars indicate one standard deviation of the mean, which was determined based on the data from biological and technical replicates (n=4).

Impact of the *spoIIE* inactivation on the transcriptome

Overview of the transcription data

To study the repercussion of *spoIIE*'s disruption on the transcriptome, samples for mRNA isolation were collected from three independent chemostat fermentations of the WT and the Δ*spoIIE* mutant at three time points (4, 11, and 23 h), corresponding to early exponential, mid-exponential and entry into the stationary phase. After RNA isolation, library construction and sequencing, the data were mapped against the published genome (NCBI). The differential expression was calculated using the SARtools pipeline on the Institut Pasteur Galaxy platform (239). Out of the 5026 coding genes annotated by NCBI Prokaryotic Genome Annotation Pipeline (PGAP), 5021 were detected in our transcriptomic data. We then compared the expression profile of the WT and Δ*spoIIE* mutant strains. The inactivation of *spoIIE* had a significant impact on the transcriptome, as about 40% of the total CDS was differentially expressed in the mutant at least at one time point (2005 out of the 5021 genes transcribed). The differentially expressed genes were clustered using the Clusters of Orthologous Groups (COGs) database (NCBI). Thirty-seven percent of these 2005 genes encode for proteins of unknown function. The rest encodes mainly proteins involved in sporulation, metabolism, signal transduction, and membrane/cell wall biogenesis. At each time point, more than 70% of the genes (Figure 7A) were significantly down-regulated ($\log_2$ fold change <1.5 and padj <0.05). The

difference between mutant and WT transcriptome was the largest at 23 h. Indeed, while the number of differentially expressed genes with a $|\log_2$ fold $|>1.5$ and padj <0.05 was equal to 589 and 578 at 4 h and 11 h, respectively, it reached 1312 genes at 23 h. This increase at 23 h was mainly due to a rise in the number of differentially expressed genes involved in metabolism (from 147 genes at 11 h to 432 at 23 h) and encoding proteins of unknown function (from 218 to 469).

Two hundred and forty-three genes were significantly up or down-regulated ($|\log_2$ fold $|>1.5$ and padj <0.05) in the mutant compared to the WT at the three time points (Figure 7B, Table S4), 97% of them being down-regulated at each time point in the Δ*spoIIE* mutant. Furthermore, 45% of these 243 genes encode proteins of unknown function while the remaining genes encode proteins involved in various cellular processes such as sporulation, solventogenesis, or chemotaxis (Figure 7C). Only four genes were up-regulated at the three time points. They form an operon of 4 genes annotated as a histidine kinase, two diguanylate cyclases, and a transcription regulator.

To analyze in more detail the data, the genes were clustered according to the putative function of the encoded proteins. In addition to the differential expression of the genes involved in sporulation, the differential expression of the genes linked to five other cellular processes was analyzed: stress response, cell wall and membrane formation, signal transduction and motility, carbohydrate metabolism, and amino acid, ion and vitamin transport and metabolism. The COG clustering of the differential expressed genes revealed that these five functional clusters were the most affected by SpoIIE's inactivity after the sporulation cluster (Figure 7C).
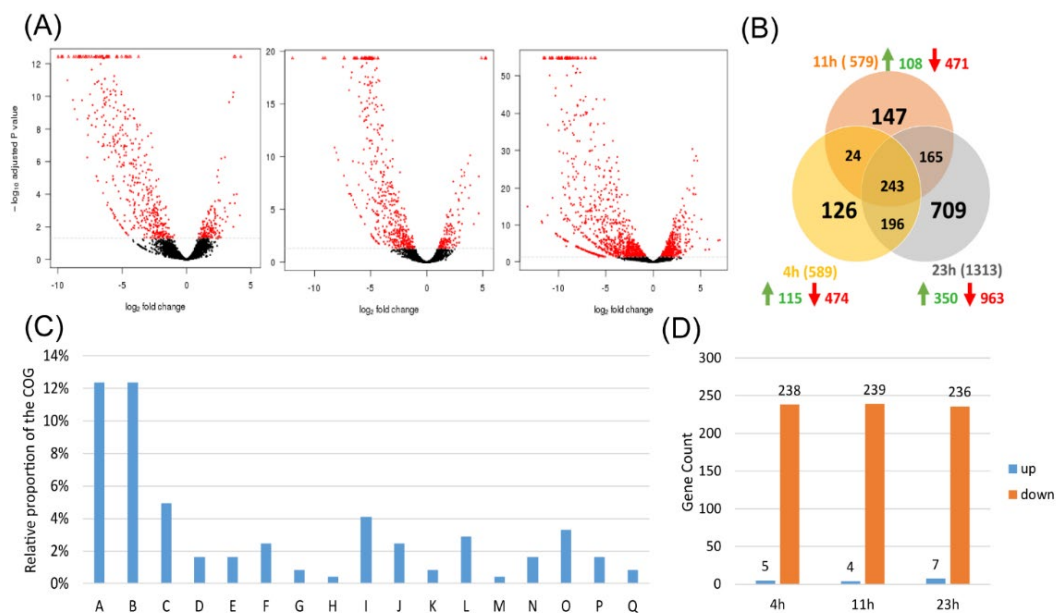
3

Figure 7: Overall differential expression dynamics. (A) Volcano plots of each comparison. Red dots represent significantly differentially expressed features; (B) Venn diagram showing the number of differentially expressed genes in the mutant at each time point; (C) COG class repartition of the 243 genes differentially expressed at all time points in the Δ*spoIIE* culture: A: Sporulation, B: Cell wall/membrane/envelope biogenesis, C: Amino acid transport and metabolism, D: Acidogenesis and Solventogenesis, E: Stress response, F: Other posttranslational modification, protein turnover, chaperones, G: Secondary metabolites biosynthesis, transport and catabolism, H: Coenzyme transport and metabolism, I: Carbohydrate transport and metabolism, J: Signal transduction mechanisms, K: Replication, recombination and repair, L: Inorganic ion transport and metabolism, M: Intracellular trafficking, secretion, and vesicular transport, N: Lipid transport and metabolism, O: Energy production and conversion, P: Transcription, Q: Defense mechanisms; (D) Proportion up and down-regulated genes at each time point in the cluster of the genes differentially expressed at all time points in the ΔspoIIE culture. The RNAseq data analyzed contains three independent biological replicates per time point and per strain (n=3).

Impact of the *spoIIE* disruption on the expression of sporulation genes

The genes involved in the sporulation process were the most down-regulated in the Δ*spoIIE* mutant. Indeed 84% of them are differentially expressed at least at one time point, and they have the lowest $\log_2$ fold change at 11 h and 23 h (Table S5). To better assess the impact of *spoIIE*'s disruption, the genes involved in sporulation regulation were clustered based on the genes present in *C. beijerinckii*'s genome and on previous studies on sporulation in *C. acetobutylicum* (23), *C. difficile* (245, 246) and *B. subtilis* (247, 248). According to the data concerning the corresponding regulon in *C. acetobutylicum*, in *C. difficile* and *B. subtilis,* the genes involved in sporulation were divided into six groups: initiation of sporulation, Spo0A regulon, σ^F regulon, σ^E regulon, σ^G regulon and σ^K regulon (Table S6). However, 23 genes, homologs to known sporulation genes, could not be linked to any known regulon.

The expression levels of most genes involved in the initiation of the sporulation and the activation of Spo0A by phosphorylation did not change in the Δ*spoIIE* mutant (Table S6). The expression of *spo0A* remains relatively stable. Only the expression of two genes (Cbei_4885 and Cbei_3375) encoding AbrB-type regulators significantly differed in the mutant compared to the WT. In *C. acetobutylicum*, *abrB* genes were shown to have a crucial role in the transition from acidogenesis to solventogenesis (131). Out of the three *abrB* genes identified in *C. acetobutylicum*, two (Cac_3647 and Cac_0310) encode transcriptional regulators down-regulating genes necessary for solventogenesis: the *sol* operon as well as *ald, bdh, adhE2* and *adc* genes. In *C. beijerinckii*, five *abrB* genes, including Cbei_3375 and Cbei_4885, are annotated. Cbei_4885 is highly similar to Cac_3647 and Cac_0310 with 88 and 89% identity, respectively. The expression of Cbei_4885 increased at 11 h and decreased at 23 h (log$_2$ fold change of 2.09 and -2.75) while the expression of Cbei_3375, 54% identity with Cac_3647 and Cac_0310, increases slightly at 23 h (log$_2$ fold change of 1.86).

In the Δ*spoIIE* strain, expression levels of the sporulation genes belonging to the Spo0A regulon were similar to those of the WT strain. As expected, no mRNA of the *spoIIE* gene was detected in the mutant cells. The expression level of *spoIIGA* and *sigE* was not differentially expressed in the mutant; at 4 h log$_2$ fold change is low (-2.69 and -2.60) but was not statistically relevant, padj > 0.05. At 11 h and 23 h, their expression returned to WT levels. The *spoIIAA-spoIIAB-sigF* operon was also slightly less expressed at 4 h in the mutant compared to the WT, as observed for the *spoIIGA-sigE* operon. As seen in other spore formers (139, 249), Spo0A boxes were detected upstream from these operons, suggesting that Spo0A controls their expression.

By contrast, the expression of genes involved in the later stages of the sporulation was more severely impacted by *spoIIE*'s disruption. The genes belonging to the σ^F regulon were all down-regulated at the three time points except *spoIIR,* suggesting that *spoIIR* is not part of the σ^F regulon in *C. beijerinckii*, in contrary to its homolog in *B. subtilis*. Also, the fold-change of expression varied throughout the σ^F regulon. Indeed, while some genes like *spoIIQ*, *spoIIP* and *spoIVB* were strongly down-regulated at the three time points (Table S6), other genes like *gpr* and *dacF* were mainly down-regulated at 4 h and 23 h. In the case of *dacF,* this could be due to the low coverage of the gene already in the WT at 11 h, but that was not the case for *gprR*. This result suggests that another transcription factor might be involved, enabling the expression of these genes during exponential growth in the *spoIIE* mutant. Most genes involved later in the sporulation process (Stage III to VII) belonging to the σ^E, σ^G, or σ^K regulon were down-regulated. Only four genes (*spoVD1*, *spoVD2*, *spoVE* and *yzbD*), associated with the σ^E regulon in *B. subtilis,* were not differentially expressed in our strain.

However, out of the 50 most differentially expressed genes at 11 h and 23 h (Table S5), only 14% and 24% of genes correspond to known proteins directly involved in the sporulation process, while 54% and 58% of the genes, respectively were of unknown function. This large number of genes with unknown function shows that other genes might be involved in the sporulation process and that *spoIIE* inactivation and the cells block at stage II of the sporulation cycle has an impact on other gene clusters.

The disruption of *spoIIE* affects other cellular processes

As observed in the COG clustering of the genes three times differentially expressed, the inactivation of *spoIIE* has an impact on several cellular processes besides sporulation, including the central metabolism (Figure 7C). Indeed the results that were obtained in this study showed the differential expression of genes associated with five functional groups: stress response, cell wall and membrane formation, signal transduction mechanisms and motility, carbohydrate metabolism, and amino acid, ions, and vitamin transport and metabolism.

Stress response

Genes involved in stress response were also strongly differentially expressed in the Δ*spoIIE* mutant (Table S7). The genes coding for rubrerythrins (Cbei_0569, Cbei_2325, and Cbei_3257) and superoxide dismutase (Cbei_1507, Cbei_1856) belonging to the σ^G regulon in *C. difficile* (245) were down-regulated in the mutant at all time points. In contrast, the genes coding for chaperonin proteins Cpn10, GroEL, and DnaJ, known to be involved in butanol tolerance in *C. acetobutylicum* (76), were up-regulated at 23 h.

Cell wall and membrane formation

As shown in the microscopy pictures (Figure 4, Figure 5 and Figure S2A-B), the morphology of Δ*spoIIE* mutant cells is very different from WT cells'. Thus, a difference in the expression of genes encoding membrane- and cell wall-associated proteins (Table S7) was expected. Among the 243 genes differentially expressed in the mutant at the three time points, 42 belonged to the cell wall/membrane cluster. These 42 genes were all down-regulated in the mutant ($\log_2$ fold change ranging from -12 to -1.9). Several of them encode glycosyltransferases, which might suggest a possible modification of the cell wall. An up-regulation of the genes involved in septum formation, *fstZ, minD* and *refZ* was also observed. These genes were proven to be crucial for the positioning of the septum formation and sporulation in *B. subtilis* (250, 251). This result can be linked to the presence of several septa observed in the mutant (Figure 4, Figure S2A-B).

Signal transduction mechanisms and motility

The initiation of the sporulation cycle is triggered by several environmental signals (23). Environmental changes are detected and processed by the cells through several transduction mechanisms. In *C. acetobutylicum,* two quorum-sensing mechanisms were studied, an Agr system (90) and an RNPP type system (252). Homologs of the Agr system were identified in *C. beijerinckii,* but no homologs of the RNPP proteins were found. In the Δ*spoIIE* mutant*,* the *agr* operon was not differentially expressed. In *C. acetobutylicum*, another type of bioactive metabolite, the clostrienose, synthesized by a PKS system, is also involved in the control of solvent production and sporulation, mainly through σ^K activation (83). No homolog to the clostrienose producing PKS enzyme was found in *C. beijerinckii*. Still, a PKS-NRPS gene cluster of 64.5 kb (80) is present, enabling the secretion of a secondary metabolite, circularin A (253). This cluster of 50 genes, from Cbei_0233 to Cbei_0283, encodes biosynthetic proteins (Cbei_0249-0254) and four transporters (Cbei_0260-0264). These genes were differentially expressed in the mutant (Table S8). The genes encoding the biosynthetic proteins are up-regulated at 4 h ($\log_2$ fold change from 1.9 to 2.9) and down-regulated at 23 h ($\log_2$ fold change from -4.6 to -1.1), while the genes encoding transporters were up-regulated at 4 h and down-regulated at 11 h.

About 20% of the genes involved in signal transduction are differentially expressed at least one time point in the Δ*spoIIE* mutant. Several genes encoding serine/threonine kinases and PAS/PAC sensor hybrid histidine kinases were differentially expressed. In particular, six genes were differentially expressed at all three time points. Genes encoding two serine/threonine kinases and an unknown gene were consistently down-regulated ($\log_2$ fold change between -10 and -3.8). Three genes belonging to the same operon were up-regulated ($\log_2$ fold change between 2.7 and 4.2). This operon is unique to *C. beijerinckii.* It harbors a PAS/PAC sensor hybrid histidine kinase, a diguanylate cyclase/phosphodiesterase with PAS/PAC sensor(s) and a diguanylate cyclase with a transcription regulator.

In *Clostridium*, the loss of motility is usually coupled with the transition from exponential phase to stationary phase (211, 254). At 4 h, no significant differential expression was observed for chemotaxis and motility genes (Table S7) in the mutant. However, at 11h and 23 h, respectively, 33% and 16% of genes from this cluster were up-regulated in the *spoIIE* mutant. The chemotaxis gene operon Cbei_4307 to Cbei_4312, as well as the motility genes *motA* and *motB,* were more transcribed in the mutant compared to the wild-type with a $\log_2$ fold change between 2 and 5 (Table S7).

3

Carbohydrate metabolism

Granulose biosynthesis

Granulose, also called bacterial glycogen, is a starch-like polymer produced at the onset of the sporulation cycle (235). Granulose biosynthesis is one of the indicators of sporulation initiation. No differential expression of granulose biosynthesis genes was observed at 4 h or 11 h (Table S9). However, at 23 h (Figure 8A), the genes *glgA, glgC* and *glgD* involved in granulose formation were all up-regulated in the Δ*spoIIE* mutant strain. In contrast, the expression of genes involved in granulose degradation (Cbei_0983 and *glgP*) remained unchanged (Figure 8B). The regulation of granulose synthesis by the Agr quorum-sensing system has been described in *C. acetobutylicum* (90). However, no differential expression of the *agr* operon was observed at any time point in our dataset. As described in *E. coli* (255–257), for the molecular regulation of glycogen formation, granulose regulation might be more complex, involving several transcriptional regulators such as CsrA and PhoP/PhoQ. While no difference in *csrA* expression was detected, both *phoP* and *phoQ* genes were up-regulated at 23 h (log$_2$fold change of 2.812 and 2.21). In *E. coli*, PhoP and PhoQ promote the expression of the glycogen formation genes *glgA, glgC* and *glgD,* which could be the case for granulose formation in the *spoIIE* mutant.
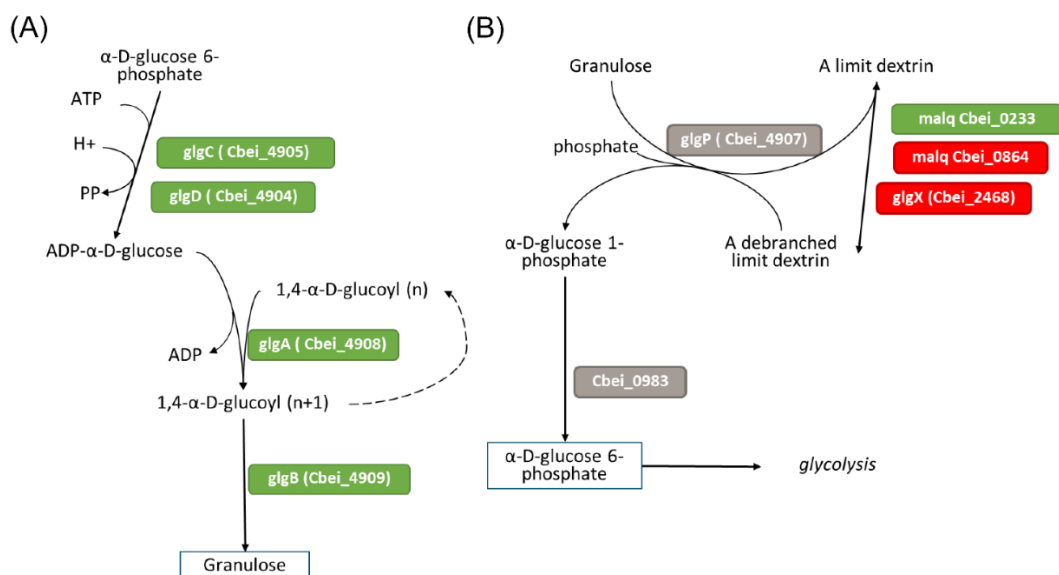
Figure 8: Differential expression at 23 h in the mutant of genes involved in (A) granulose formation and (B) granulose degradation, Figures adapted from Microcyc. Gene names are in colored boxes according to the differential expression in the Δ*spoIIE* mutant compared to the WT; genes in grey boxes have no change in expression, genes in green boxes are up-regulated, and genes in red boxes are down-regulated.

Central metabolism

Most of the genes involved in glycolysis were not differentially expressed in the ΔspoIIE mutant (Table S9). Only one of the two copies of the pyruvate kinase, Cbei_1412, was down-regulated at the three time points, but no change in expression was observed for the second copy, Cbei_4851. However, half of the genes involved in acidogenesis and solventogenesis were differentially expressed in the mutant compared to the WT strain at 23 h (Figure 9). Indeed, while the genes putatively involved in lactate and ethanol production were down-regulated in the mutant strain, the genes involved in butyrate and butanol production were up-regulated. However, no change in the expression of the genes coding for the proteins involved in acetate or acetone production was detected. The up-regulation of the *sol* operon, as well as genes encoding several alcohol dehydrogenases, might be linked to the down-regulation of the *abrB* gene, Cbei_4885, as a similar control was observed in *C. acetobutylicum* when its homolog was disrupted. We saw no differential expression of the *ctfA/B* operon, which could be linked to the accumulation of butyrate observed in the fermentation broth after 23 h, despite the up-regulation of the other genes involved in the butanol production.

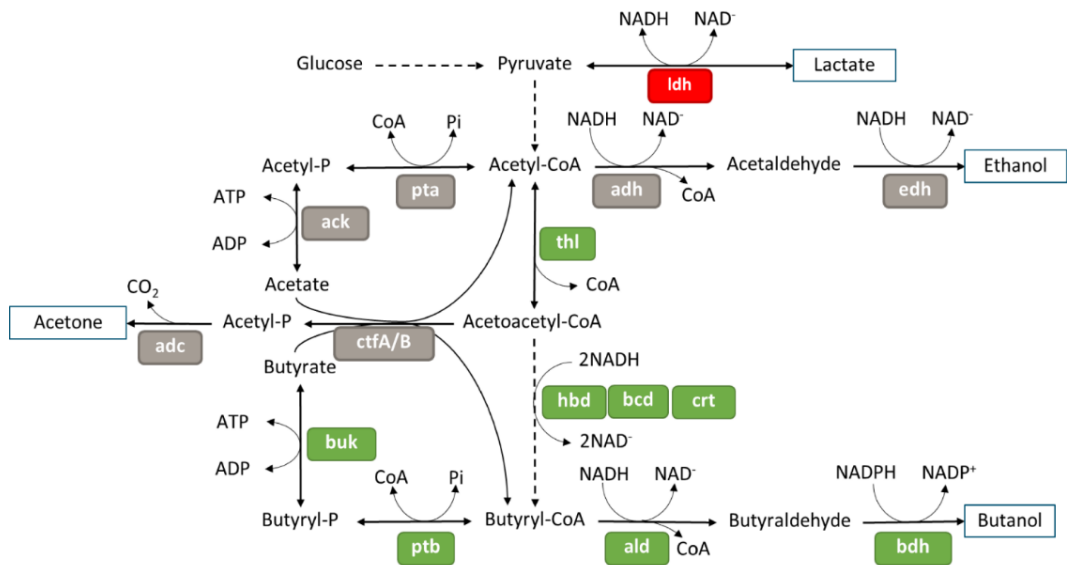**Figure 9: Differential expression in the mutant at 23 h of genes involved in acids and solvent formation, figure adapted from Microcyc.** Gene names are in colored boxes according to the differential expression in the ΔspoIIE mutant compared to the WT; genes in grey boxes have no change in expression, genes in green boxes are up-regulated and genes in red are down-regulated.

Amino acid, ions and vitamin transport and metabolism

Several vitamins and ions are essential for solventogenesis and sporulation in *Clostridium* (47, 49, 58, 258). These findings contributed to the modification of growth media to improve either the sporulation or the solvent formation. As shown in a previous transcriptomic study on *C. beijerinckii* (59), the expression of genes involved in amino acids, vitamins, and ions transport varied depending on the culture's growth stage. Based on this work, close to 600 genes, putatively involved in amino acid, ions and vitamin transport and metabolism, were identified. The transcription of 32% of them differed in the Δ*spoIIE* mutant at one time point, at least (Table S10). Twenty-one genes, including the *app* operon, were down-regulated in the mutant at the three time points. In *B. subtilis* and *C. difficile*, the *app* operon codes for permeases allowing the uptake of small quorum signaling peptides essential for the initiation of sporulation (259).

At 4 h, genes involved in ascorbate, cobalt and iron transport were down-regulated while genes encoding riboflavin and vitamin D transporters were upregulated. Moreover, at 11 h and 23 h, genes coding for cysteine-, glutamine- and manganese-transporters and the metabolism of these compounds were strongly down-regulated in the mutant ($\log_2$ fold change between -11 and -4). By contrast, we observed an up-regulation of genes encoding a proline transporter (Cbei_2870 and Cbei_2871) that has been linked to a rise in proline in the cell, which may act as a stress protectant (43, 260). At the same time points, the genes coding for a ferritin and a niacin transporter were up-regulated. Niacin is a precursor of NAD(H) and NADP(H), and its presence in the medium can lead to an increase in butanol production (261). The up-regulation of the niacin transporter could be explained by a lack of NAD(H) in the cell, which can be linked to the observed accumulation of butyrate.

At 23 h, the genes linked to the transport of phosphate were up-regulated. By contrast, the genes coding for methionine and glycine/betaine transporters ($\log_2$ fold change from -5 to -4) were strongly down-regulated, as well as the genes encoding for ascorbate- and iron transporters. The up-regulation of ferritin, coupled with the down-regulation of iron transporters, indicates an accumulation of iron in the cell, requiring the action of ferritin and an interruption of the iron uptake. The role of iron in sporulation in *Clostridium* is not clear. In *C. botulinum*, iron is needed for the formation of heat-resistant spores, while in *C. sporogenes,* its addition impairs sporulation (53). In the case of *C. beijerinckii*, the data suggests that iron is mostly needed for sporulation. The Δ*spoIIE* mutant limits its uptake and has to store the excess of iron present in the cell.

Genes of unknown function

Our transcriptomic analysis highlighted the presence of a large cluster of genes of unknown function impacted by *spoIIE*'s inactivation. Indeed, 45% of the genes differentially

expressed at the three times points did not have a known function. Except for one gene coding for a histidine triad (HIT) protein up-regulated at 23 h, these genes were down-regulated at all time points. Moreover, when focusing only on the genes differentially expressed at the 23 h, 56% of the most differentially expressed genes ($\log_2$ fold change above 5) belonged to the unknown function cluster. Some of these strongly down-regulated genes might code for spore coat proteins. Indeed, some coat proteins are species-specific and of small size, and they remain to be identified in *C. beijerinckii* (262). Spore coat proteins mainly belong to σ^E and σ^K regulon in *B. subtilis* and *C. difficile*. In *C. difficile*, the expression of a few genes encoding spore coat proteins is regulated by σ^G (142). Spore coat genes are expressed at stages IV-VII of sporulation (23). The Pan/Core Genome analysis tool (263) was used to investigate whether some differentially expressed genes of unknown function were species-specific. Thirteen genomes of 5 clostridial species (*C. acetobutylicum, C. beijerinckii, C. butyricum, C. saccharolyticum, C. saccharoperbutylicum*) were compared. One hundred and seventy-two genes of unknown function were down-regulated at 23 h and species-specific. Amongst them, thirty-one genes encoded small proteins, shorter than 100 amino acids that were strongly down-regulated at 23 h (Table S11).

3

Discussion

C. beijerinckii is a solventogenic bacterium that is well-studied because of its potential to produce biofuels and biochemicals (264, 265). However, while its central metabolism has been studied in detail, only a few studies on the regulation of its sporulation were conducted up to date. As *C. beijerinckii* belongs to the solventogenic clostridia, it has been assumed that its sporulation mechanism is identical to the one described in *C. acetobutylicum*. In this study, we described the phenotype of a Δ*spoIIE* mutant in *C. beijerinckii*, generated by CRISPR-Cas9, to verify this assumption. CRISPR-Cas9 genome engineering enables the generation of a more stable phenotype than the methods used for the *C. acetobutylicum spoIIE* mutants, generated by RNA silencing (15) and single-crossover (16). As described in *C. acetobutylicum*, the Δ*spoIIE* mutant of *C. beijerinckii* was not able to produce viable spores. The mutant was also able to produce solvents and acids at a higher level than the WT (16). Moreover, we confirmed the presence of granulose in the mutant cells, while granulose production was not tested in the *C. acetobutylicum* Δ*spoIIE* mutant. Granulose formation is the first visible indication of the cell's entry in the sporulation cycle. However, the regulatory network linking granulose formation and sporulation has not been elucidated. While several studies report that granulose and solvent formation can be totally decoupled (109, 266), this is not the case for sporulation and granulose formation (90). This study describes the first solventogenic asporogenous *Clostridium* mutant able to produce granulose.

Furthermore, the morphology of Δ*spoIIE* mutants in both clostridia is also different. In contrast to *C. acetobutylicum* Δ*spoIIE* cells, *C. beijerinckii* Δ*spoIIE* cells were elongated with several septa and showed phase-dark bodies at their polar ends. *C. acetobutylicum* mutant cells do not display any differentiation phenotype. Indeed *C. acetobutylicum* Δ*spoIIE* cells looked like vegetative cells; no internal structures apart from symmetrical septal membranes or other changes in morphology were described, even after 72 h of cultivation. However, in *B. subtilis,* the described Δ*spoIIE* mutant has a morphology close to the one observed in our study (219), as illustrated in Figure 10. We can thus hypothesize that the role of SpoIIE in *C. beijerinckii* more closely resembles the function described in *B. subtilis* (23)*.* In *B. subtilis*, SpoIIE enables the activation of σ^F and the positioning of the asymmetric septum, a role that has not been described for *C. acetobutylicum*.
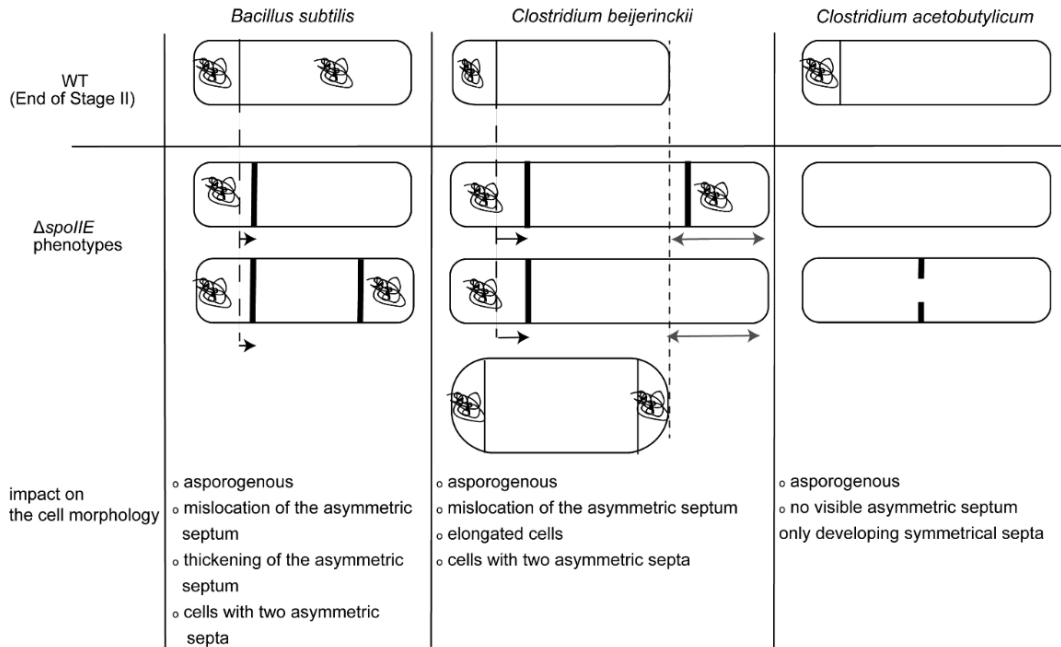
Figure 10: Schematic representation of the differences in cell morphology of Δ*spoIIE* mutants in *B. subtilis*, *C. beijerinckii* and *C. acetobutylicum*. The figure was constructed from data reported in this study and in previous studies (23, 219, 250)**.**

To better understand the role of SpoIIE, we compared the expression profile between the *spoIIE* mutant and WT strains by RNA sequencing at early-, mid-exponential and stationary phases. Forty percent of the total CDS were differentially expressed between both strains, with the 23 h time point showing the most changes. The genes involved in the regulation of sporulation after stage II were down-regulated after 23 h of culture. In contrast to what was observed in *C. acetobutylicum*, *sigF*, *sigE* and *sigG* were still expressed at 11 h. In *C. acetobutylicum*, a significant decrease of *sigF* transcript and σ^F production in the Δ*spoIIE* mutant was observed (16), suggesting that the impact of *spoIIE* inactivation on *sigF* expression is less pronounced in *C. beijerinckii*. The steady expression of *sigF* at 11 h in the *C. beijerinckii* Δ*spoIIE* mutant also suggests that σ^F does not influence its own expression, which contrasts with the regulation model described in *C. acetobutylicum* (23). In addition, the inactivation of *spoIIE* or *sigF* abolishes *sigE* expression and σ^E production in *C. acetobutylicum* (16, 110), while this is not the case for *sigE* expression in *C. beijerinckii*.

The *sigE* and *sigG* transcripts were slightly less abundant in the mutant than in the WT. These results suggest that in *C. beijerinckii*, the expression of *sigE* and *sigG* is only partially regulated by σ^F. Despite being expressed, σ^E and σ^G might not be functional since the genes

belonging to their respective regulons were weakly expressed in the mutant. This observation could be explained by (i) the need for σ^E and σ^G to reach a certain threshold level in the cell for them to allow the transcription of their regulons, (ii) a lack of active σ^E in the mother cell. Pro-σ^E might not be processed in the mutant, and σ^E might then stay inactive. In *B. subtilis*, the SpoIIGA protease processes Pro-σ^E, after being activated by SpoIIR, which is expressed under the control of σ^F. It is interesting to note that in *C. beijerinckii,* we failed to detect a clear impact of *spoIIE* inactivation on s*poIIGA* and *spoIIR* expression, even though *spoIIR* belongs to the σ^F regulon in *B. subtilis*. Thus, another protein belonging to the σ^F regulon might be required to process Pro-σ^E. It is worth noting that also in other clostridia, the expression of *spoIIR* is not strictly dependent on σ^F (110, 245).

The transcriptomic data confirmed that next to sporulation, other stationary-phase phenomena such as stress response, signal transduction, motility, carbohydrate metabolism, and the transport of amino acids, vitamins and ions were modified in the *ΔspoIIE* mutant. The interruption of the sporulation cycle led to an extension of the metabolic activity of the cells. Thus, the asporogenous *ΔspoIIE* mutant displayed three exciting features that, as observed in other solventogenic clostridia, appear to be associated with an increase in solvent production. Firstly, several genes involved in butyrate and butanol production were up-regulated. While a definite rise in acid titer was observed in the medium at the end of the fermentation, no substantial increase in butanol was measured. This rise in acid titer might have been caused by a lack of CtfA/B proteins. The CtfA/B complex enables the reassimilation of acids; however, their genes were not up-regulated in the Δ*spoIIE* mutant. The introduction of these genes in the *ΔspoIIE* mutant, expressed under the control of a constitutive promotor, might enhance the solvent production, as previously observed in *C. acetobutylicum* (180). Another explanation for the increase of the acid titer could be the lack of reducing power to convert the CoA esters into alcohols. Secondly, the genes encoding the chaperone proteins GroESL and DnaJ were up-regulated. Overexpression of these heat shock proteins in *C. acetobutylicum* increases tolerance to chemical stress (267, 268). Lastly, a change in the expression of genes encoding transporters of amino acids, ions and vitamins was observed in the *ΔspoIIE* mutant. These results suggest that a modification of the medium could also increase the solvent titer. Indeed, the addition of niacin has been demonstrated to improve butanol production (261).

In conclusion, this study has enabled us to identify genes potentially involved in the regulation of stationary phase phenomena. The expression of genes involved in secondary metabolism and signaling pathways, as well as a large number of genes of unknown functions, were strongly impacted in the *spoIIE* mutant. These genes should be investigated in more detail to obtain a better insight into the regulation of sporulation and other stationary-phase events. Indeed, the complete sporulation regulation pathway has not yet

been elucidated; critical information such as activation of σ^G (23) or the spore coat composition of *C. beijerinckii* are unknown.

The impact of the interruption of the sporulation cycle at Stage II in *C. beijerinckii* was evaluated through fermentation, microscopy, genome and transcriptome analysis. The fermentation potential of the mutant strain could be evaluated. The transcriptomic analysis provided directions to better exploit this potential by gene engineering or specific changes in media composition. Furthermore, this study reveals the complexity of the regulation of sporulation in *Clostridium* and its interconnection with other cellular regulatory networks. While some features are conserved, others seem to differ even between solventogenic clostridia. Nevertheless, this work constitutes a solid basis for further investigation of the molecular regulation of sporulation in *C. beijerinckii* and its potential for industrial application.

3

Data availability statement

The genome sequence of *Clostridium beijerinckii* NCIMB 8052 Δ*spoIIE* strain described in this study is available on the European Nucleotide Archive (ENA) under the accession number PRJEB39199. The RNA-seq data generated and analyzed for this study have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7481.

Funding

Acknowledgments

Author Contributions

MD co-designed the study, performed the experiments, collected and analyzed the data, and wrote the manuscript; NK collected the transcriptomic data; MM contributed to the transcriptomic data analysis; FC supervised the experimental work; IMV, JVDO and SWMK contributed to data interpretation, discussions and revised the manuscript; AMLC co-designed the study, supervised the work, contributed to data interpretation, discussions and revised the manuscript.

Supplementary files

(A)

(B)
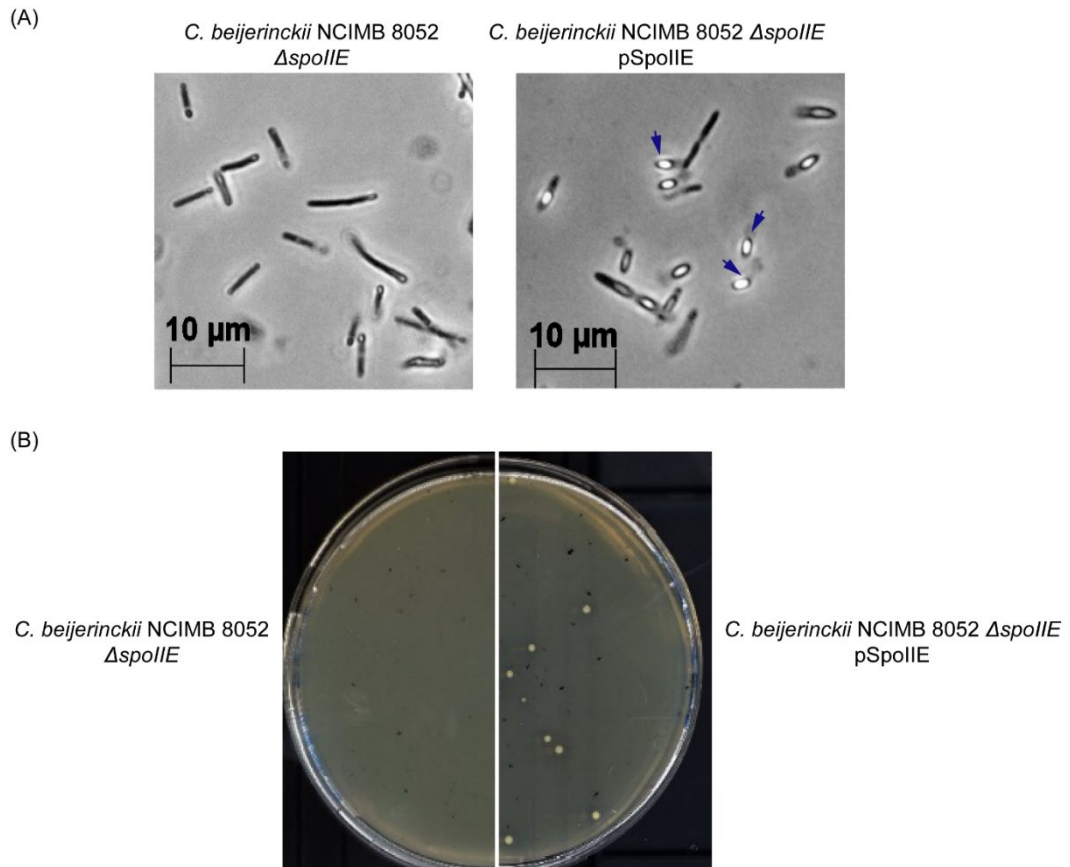
Figure S1: Complementation of the Δ*spoIIE* mutant restores sporulation. (A) Phase-contrast microscopy pictures (x400) after 35 h of cultivation of the Δ*spoIIE* mutant and the complemented strain; (B) Growth on plates after heat-shock treatment of 48 h old Δ*spoIIE* and complemented ΔspoIIE cultures. The short dark blue arrows indicate mature spores.
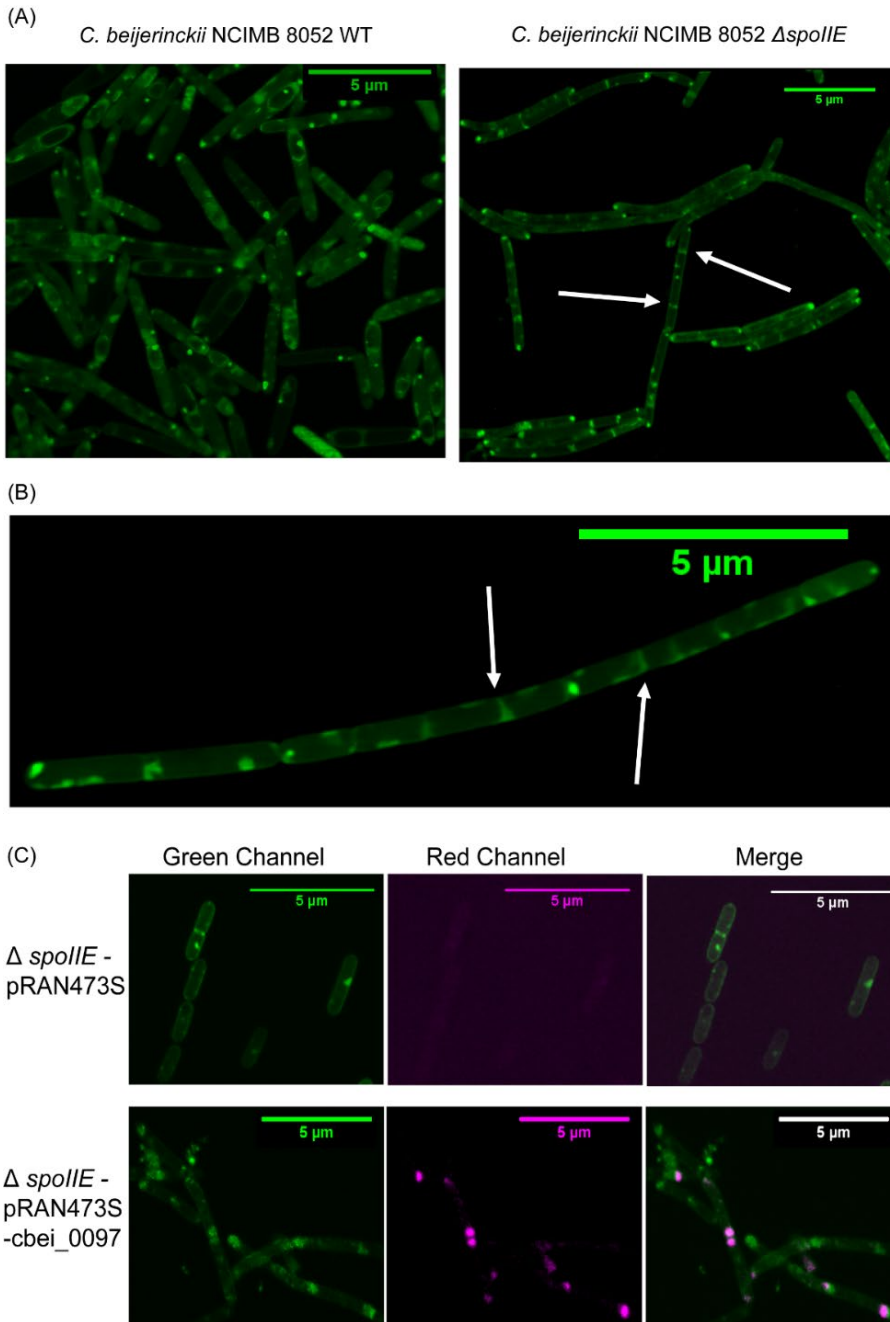
Figure S2: Additional fluorescence microscopy images of wild-type and Δ*spoIIE* mutant cells. (A) Images of several WT and Δ*spoIIE* mutant cells after 20 h of cultivation and stained by the membrane staining MTG; (B) Image of two Δ*spoIIE* mutant cells after 20 h of cultivation and stained by the membrane staining MTG ; (C) Fluorescence images of the Δ*spoIIE* mutant cells stained by MTG and harboring either the mCherry empty plasmid (pRAN73S) or the plasmid expressing mCherry fused to Cbei_0097 (pRAN73S::cbei_0097) at 18 h of cultivation (after 8 hours of atc induction). The white arrows indicate septa observed in the mutant strain.

Table S1: Oligonucleotides used in this study; the restriction sites are underlined.

Primer	Purpose	Sequence (5'-3')
M392	Removing *catP* et cloning *aad9* into pRAN473	ACACTGTCAGACACTTATCACTAAGTTCCCTCTCAA ATTCAAGTTTATCGC
M393	Removing *catP* et cloning *aad9* into pRAN473	CGTTCCACTGAGCGTCAGACCTACTACTGACAGCTT CCAAGGAGC
M394	Cloning *aad9* into pRAN473	TGATAAGTGTCTGACAGTGTCACAG
M395	Cloning *aad9* into pRAN473	GTCTGACGCTCAGTGGAACGAAAACTC
M396	Cloning *cbei_0097* into pRAN473S	AAATGAACGCATGC<u>GTCGAC</u>AACAACAACATGAGG CGAGCTAATGTGCAG
M397	Cloning *cbei_0097* into pRAN473S	TTATTAAAACTTATA<u>GGATCC</u>TTAGCTATCTAAATA AACTTTTGAAACTACTACTGTC

Table S2: Sequences of primers used in Q-RT-PCR experiments for the validation of the RNAseq analysis and Q-RT-PCR values.

Validation of RNA-seq data at 23 h
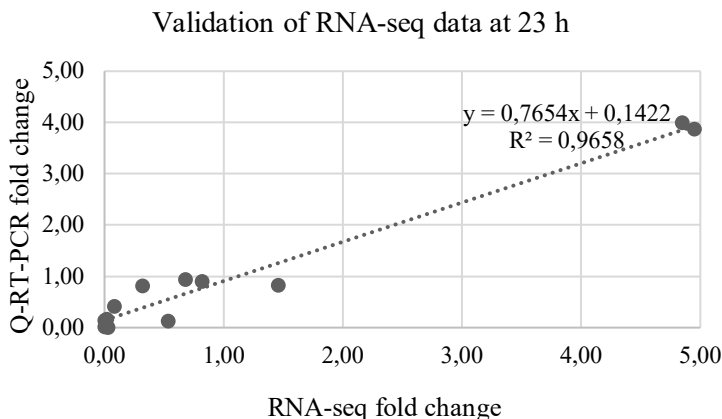
Target gene	Left Primer (5'-3')	Right Primer (5'-3')	PCR efficiency	Fold change at 23 h	
				RNAseq	q-RT-PCR
gyra (Cbei_0007)	aggaatcgcagttggtatgg	cccatcaattatttcagccaaa	2,18	1,46	0,83
spo0A (Cbei_1712)	tgaaagagcaataagacatgcaa	ccaaataacctgttaatagcctcaa	1,99	0,82	0,91
sigF (Cbei_0814)	cacctactatagaagagcttgca	ggagctccgtcatcctgat	2,10	0,68	0,93
sigG (Cbei_1121)	tcactgtttgagccaatttatca	tttcaagccaatgttcatctgt	2,13	0,32	0,82
spoIIQ (Cbei_0094)	ttccatctgcctgcttatcc	gcgctatctcctactttttgca	1,85	0,00	0,02
spoIIR (Cbei_0395)	tgtagacgaatcaaaggcaca	ttccatctgcctgcttatcc	1,93	0,53	0,13
yabG (Cbei_0388)	gaaggccaggaaagattctcc	atagctctccccactgcatc	1,99	0,01	0,15
smpa (Cbei_0051)	aaggcgatggaagaaatgga	accgagttacatgctgctct	2,11	0,03	0,00
sipL (Cbei_0391)	gggactcgcagtaaattcgg	tcaatattacactctgcctcgc	2,02	0,02	0,17
buk1 (Cbei_0204)	ttgaatggatagccccagtt	gcaccttgagctagagcaagtaa	2,06	4,96	3,87
gpr (Cbei_0822)	agtgacaccagatgctttagg	cctgtaactccaagcactcc	2,08	0,08	0,41
dnaJ (Cbei_0831)	tgcttttgatggtagtggcg	atttgatcctccgcctccaa	1,99	4,85	4,00

Table S3: SNPs detected in each Δ*spoIIE* mutant.

Δ*spoIIE* isolate	Gene ID	Predicted function	Base Change	CDS Position	Codon Change	Amino Acid Change	Polymorphism Type	Protein Effect
E, G, M			A -> G				SNP (transition)	
E, G, M			GAGGCGAG -> AGTAGTTT				Substitution	
E, G, M			T -> A				SNP (transversion)	
G, M			T -> A				SNP (transversion)	
E, G	Cbei_0198	peptidase	C -> A	1083	GAC -> GAA	D -> E	SNP (transversion)	Substitution
G	Cbei_0388	peptidase	C -> T	307	CCA -> TCA	P -> S	SNP (transition)	Substitution
E, G			W -> A	160	WTA -> ATA	X -> I	SNP	Substitution
C5, E, G, M	Cbei_0826	hypothetical protein	C -> A	170	TCC -> TAC	S -> Y	SNP (transversion)	Substitution
E, G, M			G -> A				SNP (transition)	
G			TT -> GC				Substitution	
G			TTC -> GCA				Substitution	
E, M			TC -> CA				Substitution	
E, G	Cbei_0933	ABC transporter substrate-binding protein	TGTACCGA -> CTGCAGGT	2	ATG,TAC,CGA -> ACT,GCA,GGT	MYR -> TAG	Substitution	Substitution
C5	Cbei_1353	hypothetical protein	C -> T	2858	GCA -> GTA	A -> V	SNP (transition)	Substitution
C5, E, G			G -> A				SNP (transition)	
E, G	Cbei_1939	ABC transporter permease	T -> C	362	GTC -> GCC	V -> A	SNP (transition)	Substitution
C5, E, G	Cbei_2581	hypothetical protein	G -> T	113	CCT -> CAT	P -> H	SNP (transversion)	Substitution

ΔspoIIE isolate	Gene ID	Predicted function	Base Change	CDS Position	Codon Change	Amino Acid Change	Polymorphism Type	Protein Effect
C5, G, M	Cbei_3078	PAS domain-containing two-component system sensor histidine kinase/response regulator	C -> A	1066			SNP (transversion)	Truncation
E, G	Cbei_3390	baseplate J protein	C -> A	1087			SNP (transversion)	Truncation
C5, E, G, M			G -> A				SNP (transition)	
C5, E, G, M			C -> A				SNP (transversion)	
E, G, M	Cbei_4594	2-hydroxyacid dehydrogenase	C -> A	244	GAC -> TAC	D -> Y	SNP (transversion)	Substitution
G	Cbei_4853	DNA polymerase III subunit alpha	C -> A	1072	GAC -> TAC	D -> Y	SNP (transversion)	Substitution
E, G			W -> A				SNP	

Table S4: Genes significantly differentially expressed in the Δ*spoIIE* mutant at the three time points and their expression levels in the mutant relative to the wild-type in log₂ fold change
.
Table S5: Most differentially expressed genes at 11 h and 23 h and their expression levels in the mutant relative to the wild-type in log₂ fold change.

Table S6: Differential expression of the genes belonging to the sporulation cluster in log₂ fold change. The genes in italic were not significantly differentially expressed (|log₂ fold| <1.5 or padj > 0.05) during the fermentation.

Table S7: Differential expression of the genes involved in stress response, cell wall/membrane composition, chemotaxis and motility in the Δ*spoIIE* mutant relative to the wild-type, in log₂ fold change. Only the genes for which |log₂ fold| >1.5 and padj ≤ 0.05 at least once during the fermentation.

Table S8: Differential expression of the PKS-NRPS gene cluster in the Δ*spoIIE* mutant relative to the wild-type, in log₂ fold change. The genes in italic were not significantly differentially expressed (|log₂ fold| <1.5 or padj > 0.05) during the fermentation.

3

Table S9: Differential expression of the genes involved in the granulose formation, glycolysis, acidogenesis and solventogenesis in the Δ*spoIIE* mutant relative to the wild-type, in log₂ fold change. The genes in italic were not significantly differentially expressed (|log₂ fold| <1.5 or padj > 0.05) during the fermentation.

Table S10: Differential expression of the genes involved in the amino acid, ion and vitamin transport and metabolism in the Δ*spoIIE* mutant relative to the wild-type, in log₂ fold change. Only the genes for which |log₂ fold| >1.5 and padj ≤ 0.05 at least once during the fermentation.

Table S11: Differential expression at 23 h of small species-specific genes of unknown function down-regulated in the ΔspoIIE mutant relative to the wild-type, in log₂ fold change; genes in bold are down-regulated at all time points. Only the genes for which |log₂ fold| >1.5 and padj ≤ 0.05 at least once during the fermentation.

Table S12: Differential expression at 23 h of small species-specific genes of unknown function down-regulated in the Δ*spoIIE* mutant relative to the wild-type, in log$_2$ fold change; genes in bold are down-regulated at all time points.

	Wild-type	Δ*spoIIE*	Δ*spoIIE* pSpoIIE
Glucose consumption [g.L^{-1}]	35.7 +/- 1.4	33.4 +/- 2.2	32.8 +/- 1.24
Acid production [g.L-1]	1.1 +/- 0.1	2.7 +/- 0.1	3.2 +/- 0.1
Solvent production [g.L^{-1}]	11.7 +/- 0.6	9.1 +/- 0.3	9.1 +/- 0.1
Final pH	5.8 +/- 0.0	5.3 +/- 0.0	5.0 +/- 0.0
Final OD (mAU)	6.6 +/- 0.6	7.9 +/- 0.1	7.9 +/- 0.4

3

Chapter 4-

Characterization of a degenerated *Clostridium beijerinckii* mutant strain

Mamou Diallo[1,5], Nicolas Kint[2], Marc Monot[3], Isabelle Martin-Verstraete[2,4], Servé W. M. Kengen[5], Ana M. López-Contreras[1]

[1]Wageningen Food and Biobased Research, Wageningen, The Netherlands
[2] Laboratoire Pathogenèses des Bactéries Anaérobies, Institut Pasteur, Université de Paris, Paris, France
[3] Biomics platform, C2RT, Institut Pasteur, Paris, France
[4] Institut Universitaire de France, Paris, France
[5] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

Abbreviations: ABE, Acetone Butanol Ethanol; **c-di-AMP**; cyclic di-adenosine monophosphate; **CRISPR**, Clustered Regularly Interspaced Short Palindromic Repeats; **Cas**, CRISPR-associated protein; **COG**; Clusters of Orthologous Groups; **ECF**, extracytoplasmic function; **SNP,** single nucleotide polymorphism

4

Abstract

S train degeneration, an uncontrolled change in strain physiology, leads in solventogenic clostridia to a decrease of solvent production and the loss of sporulation capacity. In a previous study, we described *Clostridium beijerinckii* Δ*spoIIE*C5, which was asporogenous and capable of solventogenesis. Here, we describe *C. beijerinckii* Δ*spoIIE*C2, a degenerated strain derived from *C. beijerinckii* Δ*spoIIE* that is no longer solventogenic. To study the effect of strain degeneration, we analyzed the phenotype, genotype and transcriptome of Δ*spoIIE*C2 and compared them to those of the Δ*spoIIE* mutant and the wild-type strain. Compared to cultures of wild-type and other Δ*spoIIE* strains, biomass production in Δ*spoIIE*C2 cultures was reduced by 3,5-fold and entered earlier in stationary phase. Moreover, six point mutations in coding sequences were detected in the Δ*spoIIE*C2 genome. The transcriptome of Δ*spoIIE*C2 cultures was further compared to Δ*spoIIE*C5 cultures at early exponential phase and the onset of the stationary phase. More than 30% of the genome was differentially expressed in Δ*spoIIE*C2 cultures. Furthermore, out of the six mutated genes, two were differentially expressed; *cbei_0201*, encoding the putative transcriptional regulator CdaR, was up-regulated at both time points, whereas *cbei_1712* encoding Spo0A was down-regulated at the entry into stationary phase, indicating that one of those mutations might have caused the degeneration phenotype. Strain degeneration was reflected on the transcriptome level by a down-regulation of the genes involved in translation, carbohydrate and amino acid metabolism. Concomitantly, an up-regulation of genes involved in signal transduction, regulatory networks and stress response was observed, indicating that the scarce energy resources were reallocated to the adaptation to environmental stressors.

4

Introduction

From the production of antibiotics to flavors and biofuels, fermentation is a widely used process in industry. Moreover, with the transition from a petrochemical-based industry to a more sustainable biobased industry, fermentation processes are increasingly popular since it relies on the production of biochemicals by specific microbes under controlled conditions. However, the use of bacteria at an industrial scale comes with several risks, such as phage infection and strain instability. Strain stability enables a constant productivity and fermentation profile in fixed conditions. However, when bacterial strains are submitted to strict fermentation conditions for a long time, the strain mutation rate increases, and low-productive mutants tend to take over the population, decreasing the process efficiency. This phenomenon is called strain degeneration and is defined as the loss of the ability to produce a desired chemical compound and the ineptitude for a bacterial population to keep desirable traits for several generations (269). This phenomenon affects a wide range of bacteria and predicting it remains impossible (270). Strain degeneration has been mostly studied in three genera *Streptomyces*, *Bacillus* and *Clostridium*. Three mechanisms have been linked to strain degeneration; transposable elements, SOS response and adaptive mutations. In *Clostridium*, strain degeneration has been studied only in solventogenic clostridia. In *C. acetobutylicum,* it has been linked to the loss of the pSol megaplasmid, harboring the *sol* operon that encodes crucial enzymes for solvent production (155, 156). In *C. beijerinckii* and *C. saccharoperbutylacetonicum,* strain degeneration seems to be linked to mutations. But no consensus on the location, number, or type of mutations causing degeneration exists. Quorum-sensing mechanisms seem to be involved in the degeneration of *C. saccharoperbutylacetonicum* N1-4, as suggested by the work of Kosaka et al. (81), in which solvent production could be restored in a degenerated mutant after the addition to the cultivation medium of concentrated broth extract from a wild-type fermentation. A better understanding of the causes and consequences of strain degeneration is crucial to improve the industrial use of solventogenic clostridia in biorefineries. In a previous study, we generated *C. beijerinckii spoIIE* deficient mutants by CRISPR-Cas9 (119), which were all asporogenous. In contrast to all the other *C. beijerinckii* Δ*spoIIE* strains, the Δ*spoIIE*C2 mutant could not produce solvents. In this study, genotype, phenotype and transcriptome of the degenerated strain Δ*spoIIE*C2 were analyzed and compared to solventogenic Δ*spoIIE* strains. This study enabled us to investigate strain degeneration through the loss of solvent production capacity and its relation to cell morphology, metabolism and transcriptome.

4

Methods

Bacterial strains and culture conditions

Relevant characteristics of the bacterial strains used in this study are listed in Table 1.

Table 1: Bacterial strains used in this study

Strains	Relevant characteristics	Source or reference
C. beijerinckii **NCIMB 8052 (WT)**	Wild-type	NCIMB
Δ*spoIIE*C5 (C5)	NCIMB 8052; *spoIIE* deletion; solvent producing	(119)
Δ*spoIIE*C2 (C2)	NCIMB 8052; *spoIIE* deletion; degenerated	This study

The WT was stored as spores suspension and the mutants as vegetative cells in 15 % glycerol solution at – 20°C. Spores suspension were heat-shocked 1 min at 95 °C before inoculation in liquid medium. Except for fermentation assays, liquid cultures of the wild-type and mutant were grown in liquid modified CGM (mCGM) containing per liter: yeast extract, 5.00 g; KH_2PO_4, 0.75 g; K_2HPO_4, 0.75 g; asparagine•H_2O, 2.00 g; $(NH_4)_2SO_4$, 2.00 g; cysteine, 0.50 g; $MgSO_4$•7 H_2O, 0.40 g; $MnSO_4$•H_2O, 0.01 g; $FeSO_4$•7 H_2O, 0.01 g; glucose, 10.00 g. Liquid media were made anaerobic by flushing with nitrogen gas. Cultivation was performed at 37 °C, without shaking, and anaerobically in i) an anaerobic chamber (Sheldon Manufacturing, Oregon USA; gas mixture consisting of 15 % CO_2, 4 % H_2 and 81 % N_2) or ii) in glass serum vials as described above. The *Clostridium* strains were grown on mCGM agar containing per litre: yeast extract, 1.00 g; tryptone, 2.00 g; KH_2PO_4, 0.50 g; K_2HPO_4, 1.00 g; $(NH_4)_2SO_4$, 2.00 g, $MgSO_4$•7 H_2O, 0.10 g; $MnSO_4$•H_2O, 0.01 g; $FeSO_4$•7 H_2O, 0.015 g; $CaCl_2$, 0.01 g; $CoCl_2$, 0.002 g; $CaCl_2$, 0.002 g; glucose, 50.00 g and microbial agar 12.00 g.

DNA extraction and genomic analysis

Genomic DNA of *C. beijerinckii* NCIMB 8052 was purified using the GenElute bacterial genomic DNA kit (Sigma-Aldrich, Saint-Louis, USA). Concentration of gDNA was determined using a nanodrop spectrophotometer (Thermofisher), and quality was verified on 1 % agarose gel. PCR reactions were carried out using the Q5 Master mix (New England Biolabs). Genomic DNA of *C. beijerinckii* NCIMB 8052, Δ*spoIIE*C5 and Δ*spoIIE*C2 was purified using the GenElute bacterial genomic DNA kit (Sigma-Aldrich, Saint-Louis, USA). DNA sequencing of clones was done by BaseClear (Leiden, The Netherlands).

Phenotypical characterization

Granulose staining

Granulose accumulation was monitored by iodine staining. Each *C. beijerinckii* strain was grown on agar mCGM agar plates and incubated anaerobically at 37 ^{0}C. After 24 h and 48 h incubation, one petri dish was opened and inverted over I_2 crystals for approximately 1 min. The colonies of granulose-negative mutants were identifiable by their being unstained by the sublimed I_2 vapor (90, 235).

Microscopy

A phase-contrast microscope (Olympus BX51) was used to determine the morphology and sporulation stages of cells. Fresh preparations were observed by phase-contrast microscopy at ×400 magnifications. Transmission electron microscopy (TEM) was used to have a better insight into the cell's internal structures. Samples from CM2 grown strains were taken at 12 and 24 h and prepared as described previously (271).

Fermentation

Fermentations were performed at 35ºC in CM2 medium (230), which contains per liter: yeast extract, 5.00 g; KH2PO4, 1.00 g; K2HPO4, 0.76 g; ammonium acetate,3.00 g; *p*-aminobenzoic acid, 0.10 g; MgSO4·7 H2O, 1.00 g; and FeSO4·7 H2O, 0.50 g, glucose, 60,00 g. Metabolites were determined in clear culture supernatants after removal of cells by centrifugation. Glucose, acetate, butyrate, lactate, acetone, butanol, and ethanol were determined by high-performance liquid chromatography (HPLC) as described previously (68, 180) using 4 methyl valeric acid (30 mM) acid as an internal standard.

4

Transcriptome analysis

Experimental design

To have 3 independent biological replicates, the same procedure was repeated 3 times in 3 different weeks. Each week, a fresh preculture was used to inoculate 2 identical bioreactors. Samples were taken over the early exponential, late exponential and stationary phases (samples at 4 h, 11 h and 23 h). Following the centrifugation of the samples, cell pellets were washed with chilled RNase free water and suspended in RNase free water to obtain a suspension having an optical density (measured at 600 nm) of approximately 1 mAU immediately after collection. A 3-mL diluted sample was centrifuged, the supernatant was discarded, and the cell pellet was stored at − 80 °C for subsequent isolation. Supernatants were used for HPLC analysis.

RNA isolation, sequencing protocol and bioinformatics analysis

Frozen samples were thawed on ice, and RNA was isolated using the high pure RNA isolation kit (Roche Diagnostics, Basel, Switzerland). Quality and concentration of RNA

samples were checked using a nanodrop machine (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The absence of DNA in the RNA samples was evaluated by qPCR analysis performed with BioRad CFX 96 Touch™. The RNA quality and integrity were determined using the Qsep 100 bioanalyzer (Bioptic Inc.).

After Illumina sequencing, the reads were mapped to the *C. beijerinckii* genome using Bowtie (237) then converted into BAM files with the Samtools (238). The differential analysis and statistical analysis were performed with the Sartools pipeline (239). The data were visualized in a strand-specific manner using COV2HTML (http://mmonot.eu/COV2HTML/)(240).

Results and discussion

Phenotype characterization

Metabolism

Recently, SpoIIE-deficient mutants of *C. beijerinckii* were generated at our lab using CRISPR-Cas9 to remove a 2,4 kb fragment within the gene coding sequence (119). The deletion was confirmed by whole-genome sequencing. Then their solvent producing capacity was assessed by small-scale fermentations in anaerobic bottles. Among the recombinants obtained, Δ*spoIIE*C2 did not produce solvents in contrast to the other four SpoIIE deficient strains isolated (Figure 1).
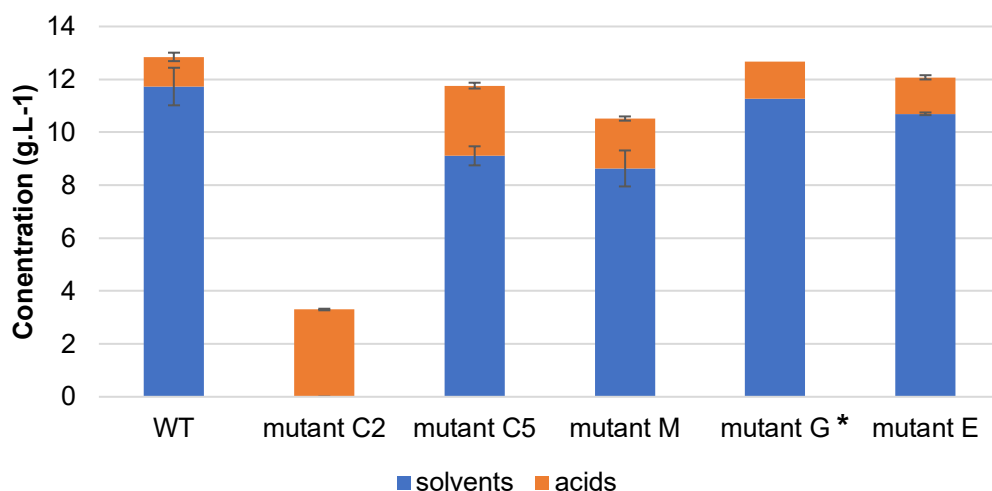
Figure 1: Final product titers after 75 h of cultivation of each isolated *spoIIE* deficient strain. Cultivation was done with 50 g.L^{-1} of glucose at 35°C. Except for mutant G, the standard deviation of the mean indicated was determined based on the data from biological duplicates (n=2).

To better monitor the cultivation parameters, fermentations were then performed in reactors. During growth on glucose, Δ*spoIIE*C2 cultures still accumulated acids (Figure 2A), but no solvents were formed (Table 2). Δ*spoIIE*C2 cultures reached a lower cell density than WT or other Δ*spoIIE* mutants cultures (Figure 2B). Indeed, the maximum absorbance measured was only 4,2 mAU at 30 h, while Δ*spoIIE*C5 cultures reached an absorbance of 14 mAU. This impaired growth of the degenerated mutant has been observed in previous research (122, 124, 272). Nevertheless, limited growth is not a strict characteristic of degenerated strains since degenerated strains with wild-type growth were previously described as well (122, 155, 273). The glucose consumption in Δ*spoIIE*C2 cultures was low, not exceeding 5 g.L^{-1} after 30 h of cultivation, and the pH of the Δ*spoIIE*C2 culture dropped quickly from 6,2 to 4,8 after 7 h of cultivation. The accumulation of acids and the absence of solvents in the medium during cultivation indicated that Δ*spoIIE*C2 was a degenerated Δ*spoIIE* mutant. The degeneration occurred probably after the *spoIIE* disruption with the CRISPR-Cas9 tool (119). The CRISPR-Cas9 method requires incubation of the cells in xylose-rich media for 48 h to 4 days, followed by several subcultures for plasmid curation. These conditions might have favored the generation of strain mutations.

Table 2: Product concentration after 30 h of cultivation of Δ*spoIIE* mutant strains. Cultivation was done in CM2 media in 0,4 L bioreactors at 35°C. One standard deviation of the mean was determined based on the data from independent biological and technical duplicates (n=4). *Acetate in the medium was consumed.

	WT	Δ*spoIIE*C5	Δ*spoIIE*C2
Product and substrates titers in g.L^{-1}			
Glucose consumed	31,50 ± 1,71	33,96 ± 1,44	4,38 ± 0,76
Acetate	-1,06 ± 0,12*	-0,85 ± 0,09*	0,72 ± 0,06
Butyrate	0,50 ± 0,06	0,67 ± 0,12	1,30 ± 0,05
Lactate	0,44 ± 0,10	0,06 ± 0,03	0,03 ± 0,09
Acetone	1,86 ± 0,05	2,66 ± 0,19	n.d
Butanol	5,80 ± 0,12	6,27 ± 0,30	n.d
Specific growth rate (h^{-1})	0,31 ± 0,03	0,32 ± 0,01	0,25 ± 0,02

4

A.

B.
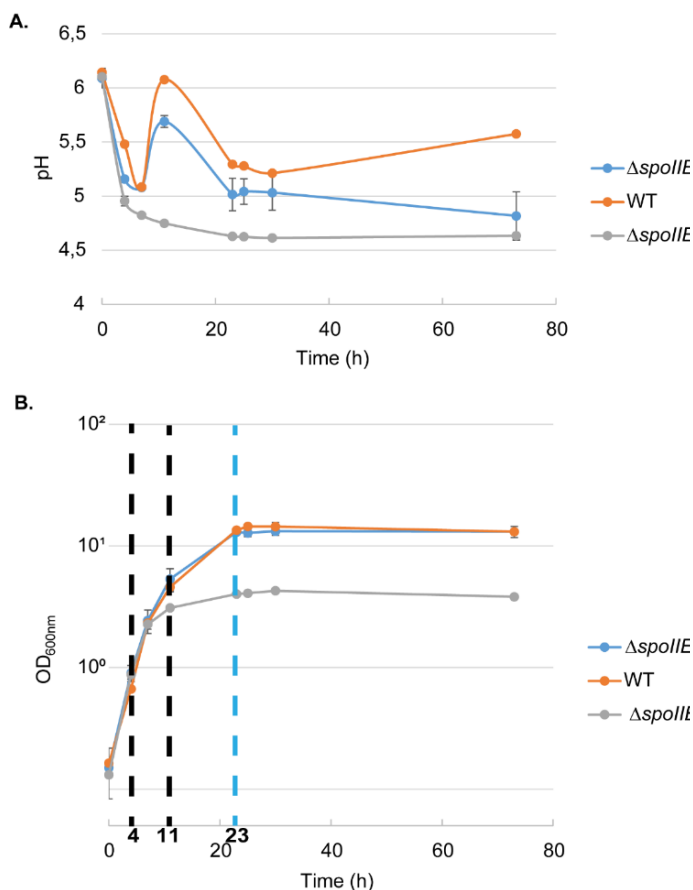
Figure 2: Comparison of fermentation profiles of *C. beijerinckii* NCIMB 8052 WT, Δ*spoIIE*C5 and Δ*spoIIE*C2 strains during 73 h of cultivation in 0,4 L bioreactors at 35ºC.
(A) Change in pH of each culture (B) Cell growth measured as optical density at 600 nm. Values represent the mean of the biological replicates, and error bars represent the standard deviations. Time-points for samples subjected to RNA expression analysis are indicated by black vertical dotted lines when taken in both Δ*spoIIE*C5 and Δ*spoIIE*C2 cultures and blue when taken only in the Δ*spoIIE*C5 cultures. One standard deviation of the mean was determined based on the data from independent biological replicates (n=4).

To determine the effect of pH control on the cultures, Δ*spoIIE*C2 cultivations with controlled pH at 5.1 were carried out in bioreactors. Again, no solvent production was observed (Figure 3), despite an increase in glucose consumption. The glucose consumed was used for the formation of acetate and butyrate. In previous work on degenerated *C. beijerinckii* strains, solvent-producing capacities were restored by lowering the incubation temperature (155). Indeed, Kashket and Cao (155) reported that lowering the temperature and thereby decreasing the acid formation rate at beneficial effects on solvent production. Thus, fermentation parameters were modified, and cultures were carried out at 30ºC. Still, lowering the temperature did not change the profile of the fermentation products of Δ*spoIIE*C2 cultures (not shown).
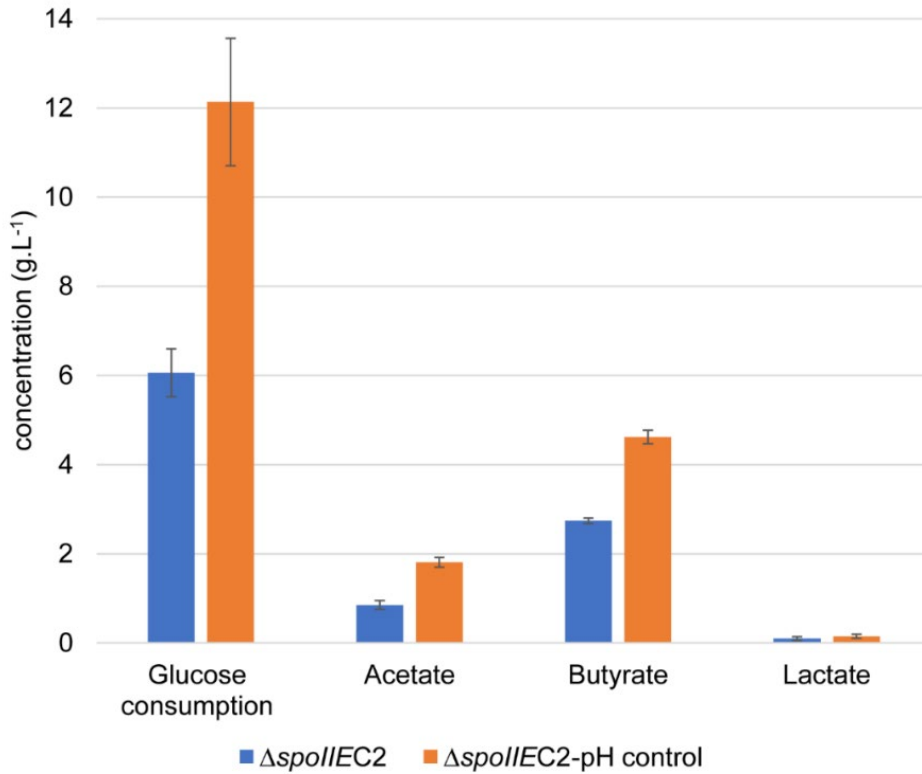
Figure 3: Glucose consumption and acid production after 72 h of cultivation of Δ*spoIIE*C2. Cultivation was done at 35ºC in 0,4 L bioreactors. One standard deviation of the mean was determined based on the data from biological duplicates (n=2).

Granulose formation is another pendant of carbohydrate metabolism in clostridia. Granulose is a polymer of glucose stored by the cells at the entry of the sporulation cycle (21). Granulose is reported to enable the storage of energy needed for the completion of sporulation. Even though the Δ*spoIIE* solventogenic mutants did not sporulate, they were all able to produce granulose (60). Granulose formation in Δ*spoIIE*C2 was assessed by iodine staining. Δ*spoIIE*C2 and WT strains were grown anaerobically on solid media, and after 24 and 48 h, the plates were exposed to I_2 vapor to stain the colonies producing granulose (Figure 4A). Δ*spoIIE*C2 colonies grew slower than the WT, as only a few colonies were seen after 24 h of incubation. After 48 h, more colonies were observed, which after I_2 staining became dark-colored. These were visible on the WT and Δ*spoIIE*C2 plates, confirming the production of granulose in both strains.

In addition, we observed a difference in colonial morphology (Figure 4B). While WT colonies had a circular form and a convex elevation, Δ*spoIIE*C2 colonies were irregular, flat with an undulated margin, just as reported previously for other degenerated strains (155, 274, 275). Colonial morphology is regulated by cell signaling pathways and results from

cell motility and exopolysaccharide synthesis (276, 277). Consequently, the differences in morphology of Δ*spoIIE*C2 colonies indicate a change in cell signaling, motility, exopolysaccharide secretion, or premature cell death.
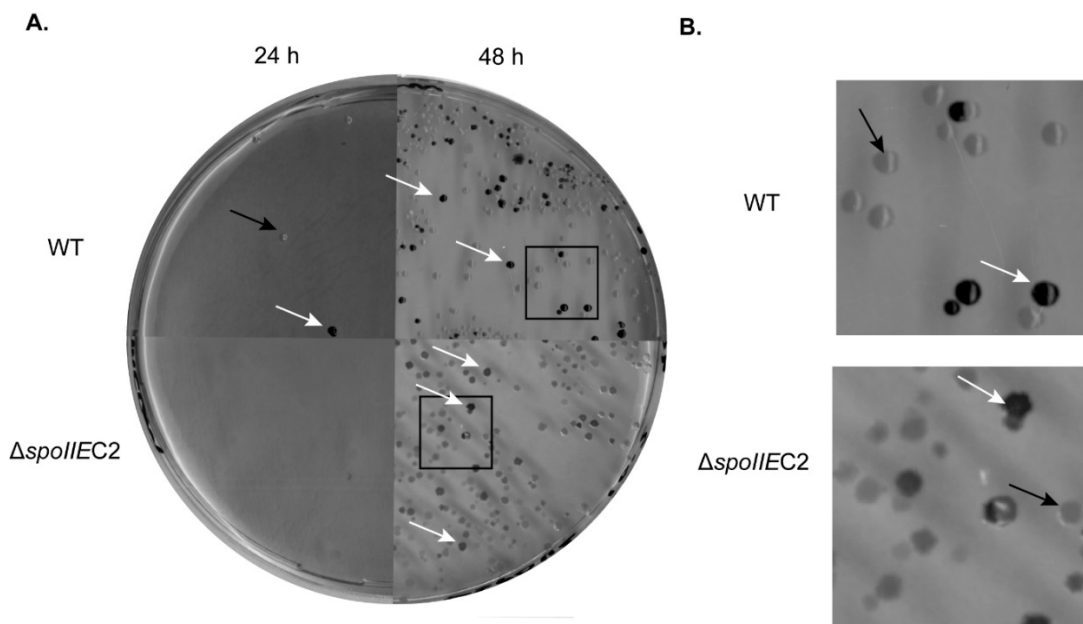
Figure 4: Granulose accumulation in wild-type and Δ*spoIIE*C2 colonies. (A) after 24 h and 48 h of growth on mCGM agar plates. Black arrows indicate colonies not producing granulose, and white arrows indicate colonies producing granulose. The areas in the black squares were magnified in (B) to observe the colony shapes.

Microscopy observations

The Δ*spoIIE*C2 cell morphology was studied by phase-contrast microscopy and TEM. As observed in the other Δ*spoIIE* mutants, we could not detect any spores in the cultures (Figure 5A). Moreover, no clostridial cells (cells swollen due to granulose accumulation) were observed in Δ*spoIIE*C2 cultures. This observation could be explained by a decrease and delay in the granulose production in Δ*spoIIE*C2 cells compared to the WT. As for the cell shape, both the Δ*spoIIE*C5 cells and the Δ*spoIIE*C2 cells were elongated after 24 h, as was observed before by other studies (60, 124). Like in Δ*spoIIE*C5 cells, Δ*spoIIE*C2 cells harbored several asymmetric septa and electron-dense granular structures. Still, these structures were visible earlier in Δ*spoIIE*C2 cultures, compared to Δ*spoIIE*C5 cultures, after 24 h or 72 h of cultivation, respectively. These structures, probably condensed DNA, were detected on both sides of the cells' septa and at the polar ends (Figure 5B). TEM micrographs showed in addition cell debris and a light cytosol in Δ*spoIIE*C2 cells, a sign of a loss of cytoplasmic material and damaged cells.
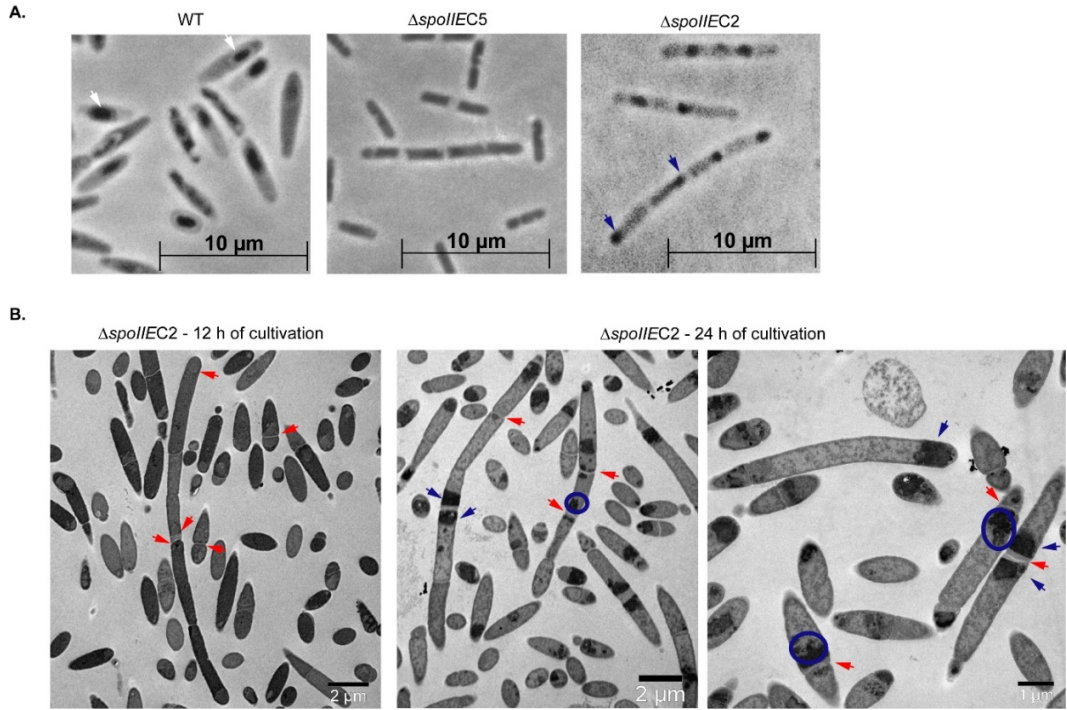
Figure 5: Microscopy pictures of *C. beijerinckii* WT, Δ*spoIIE*C2 and Δ*spoIIE*C5 cells. (A) Phase-contrast microscopy images x1000 magnification of *C. beijerinckii* WT, Δ*spoIIE*C2 and Δ*spoIIE*C5 cells after 25 h of cultivation. While pre-spores and mature spores were seen in the WT images after 25 h of cultivation, non were seen in Δ*spoIIE* cultures. (B) Transmission electron microscopy images of Δ*spoIIE*C2 cells after 12 h and 24 h of cultivation. White arrows indicate spores and pre-spores; red arrows indicate a septum; dark blue arrows indicate electron-dense masses, and blue circles highlight areas where this mass is unevenly distributed.

4

Genome analysis

Despite having the same *spoIIE* mutation, Δ*spoIIE*C2 had a different behavior and morphology than the other Δ*spoIIE* strains developed. Thus, we compared the genomes of all the isolated Δ*spoIIE* strains with our WT strain and did a single nucleotide polymorphism (SNP) analysis (60) to screen for mutations in the Δ*spoIIE* genomes. Only the nucleotide changes present in 100% of the reads were considered. Compared to the WT and the other Δ*spoIIE* mutants, Δ*spoIIE*C2 harbored seven unique SNPs, listed in Table 3. Out of these SNPs, six were located in coding sequences, five had an impact on the protein sequence, and none of these mutations have been reported to cause strain degeneration in solventogenic clostridia (122, 124, 278).

One mutation was detected in the gene encoding Cbei_0048, which is annotated as a transcriptional regulator of unknown function. The protein belongs to the TrmB transcriptional regulator family, a newly classified family of transcriptional regulators involved in regulating metabolic pathways and mainly studied in archaea (279). The mutation is located outside the DNA binding region and leads to an amino acid substitution close to the C-terminal end of the protein.

A nonsense mutation was detected in the coding sequence of Cbei_0201, a protein of unknown function. Cbei_0201 forms an operon with Cbei_0200 and belongs to the Ybbr family. Cbei_0201 is homologous to the CdaA regulatory protein, CdaR, characterized in *B. subtilis* (280). The synteny of this operon in *Bacillus* is conserved in *C. beijerinckii*. CdaA and CdaR enable the synthesis of c-di-AMP, a second messenger mediating growth and cell homeostasis (281), which can become toxic when present in excess in the cells (280). The SNP caused the truncation of CdaR at the 99 a.a., which usually is 407 a.a. long, and might affect the protein function and, as such, impact the c-di-AMP levels in Δ*spoIIE*C2 cells.

The Cbei_0699 gene harbored a missense mutation. Cbei_0699 is annotated as IIBC subunit of a PTS system specific for cellobiose/arbutin/salicin. The mutation caused a substitution in the conserved region of the IIB domain, which enables the opening of the IIC channel. Therefore, the mutation could impact the protein activity, hindering the transport of carbohydrates to the cell.

An SNP in Cbei_1193, a gene encoding a CDP-diglyceride synthetase, caused a transversion in a non-conserved region of the protein. Cbei_1193 is involved in the diacylglycerol biosynthesis pathway.

Cbei_1712, encoding the sporulation factor Spo0A, also harbored a mutation in the Δ*spoIIE*C2 genome. Spo0A is essential for the transition from exponential to stationary phase in Firmicutes. The mutation consists of the substitution of a valine by a leucine

(V172L) between the regulatory and the DNA binding domains (Figure 6) in a conserved region (136), which may impact Spo0A activity. Three studies recently studied the impact of SNPs in *spo0A* on the sporulation and solvent formation capacity of the strain (71, 107, 129). Depending on the location, solvent production could be impaired, as described in Chapter 1 and illustrated in Figure 6. The SNPs in the conserved linker, indicated in green in Figure 6, impaired solvent formation in *C. acetobutylicum* and *C. saccharobutylicum* (107). This mutation might be the cause for the degeneration phenotype of the Δ*spoIIE*C2 mutant.
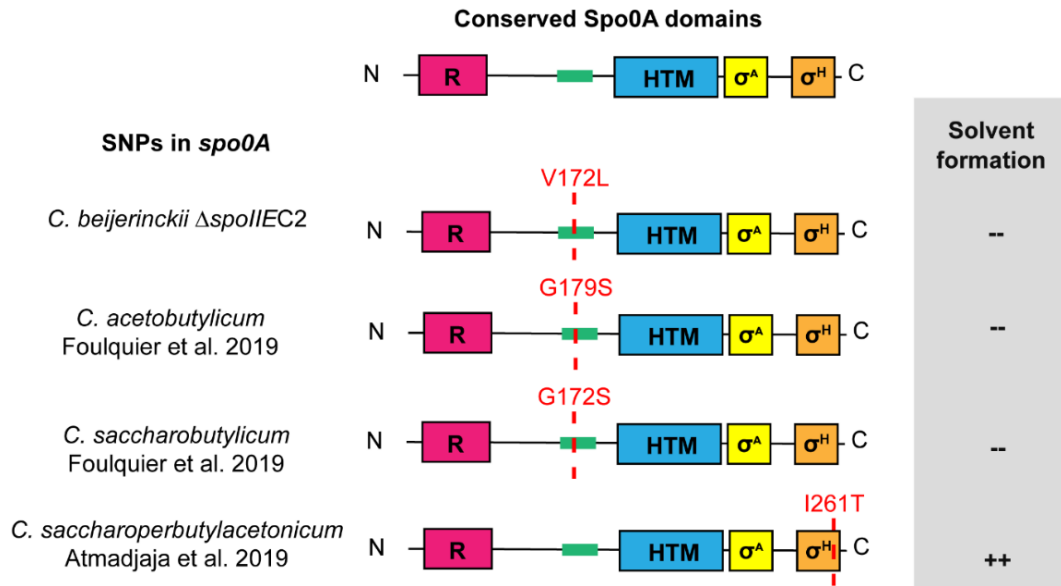
Figure 6: SNPs reported in *spo0A* and their impact on solvent formation. Five conserved regions are indicated by colored boxes; R: signal receiver domain, Green box: conserved region with no known function, HTM: helix turn motif, σ^A: putative σ^A activator region, σ^H: putative σ^H activator region. The ability to sporulate, to form heat resistant spores and to produce solvent is indicated next to the scheme of the *spo0A* mutation present in each mutant strain; -- : indicates the abolition of the feature in the mutant compared to the wild-type phenotype; ++ indicates an increase compared to the wild-type phenotype

Still, we cannot exclude that the phenotype of the Δ*spoIIE*C2 mutant results from the incremental effect of the different mutations detected in its genome. Despite several attempts to evaluate the role of the *spo0A* mutation in the Δ*spoIIE*C2 mutant by introducing a wild-type copy of *spo0A* into the strain, the Δ*spoIIE*C2 mutant could not be transformed.

Table 3: Unique SNPs detected in the Δ*spoIIE*C2 genome compared to the wild-type and the other Δ*spoIIE* mutants sequenced; a.a: amino acid

Nucleotide change	Locus tag	Encoded product	Position in CDS	Codon Change	Change in a.a	Protein Effect
G -> T	Cbei_0048	TrmB family transcriptional regulator	324	ATG -> ATT	M -> I	Substitution
G -> T	Cbei_0086	Transcription-repair coupling factor	3231	GTG -> GTT		None
G -> T	Cbei_0201	Putative transcription regulator, CdaR	298			Truncation
C -> A	Cbei_0699	PTS cellobiose transporter subunit IIBC	26	GCG -> GAG	A -> E	Substitution
C -> A	Cbei_1193	CDP-diglyceride synthetase	242	GCT -> GAT	A -> D	Substitution
G -> T	between Cbei_1235 and Cbei_1236	Non coding region				
G -> T	Cbei_1712	Sporulation transcription factor Spo0A	514	GTA -> TTA	V -> L	Substitution

RNA-seq analysis

Overview of the transcriptomic data

To further investigate the causes and consequences of Δ*spoIIE*C2*'s* degeneration phenotype, we analyzed and compared the transcriptome of Δ*spoIIE*C2 cultures to the Δ*spoIIE*C5 transcriptome (60) and to the recently published studies on two degenerated *C. beijerinckii* strains, BA105 (122) and DG-8052 (124). Samples for mRNA isolation were collected from cultures of the Δ*spoIIE*C5 and Δ*spoIIE*C2 strains grown in bioreactors. The transcriptome of Δ*spoIIE*C2 cultures was compared to the transcriptome of Δ*spoIIE*C5 cultures to study the effect of the loss of solvent production capacity exclusively. As described previously in (60), samples for mRNA isolation were harvested at three time points for Δ*spoIIE*C5 (4 h, 11 h and 23 h), corresponding to early exponential, mid-exponential and onset of the stationary phase. Samples at two time points were collected for Δ*spoIIE*C2, at 4 h and 11 h, corresponding to early exponential and entry into stationary phase. After RNA isolation, library construction and sequencing, the data were mapped against the published genome (NCBI). The expression profile of both Δ*spoIIE* mutant strains, Δ*spoIIE*C5 and Δ*spoIIE*C2, were compared through a differential gene expression analysis, calculated using the SARtools pipeline (239). Three differential analyses were carried out (Figure 7A), two comparing mRNA extracted from samples taken at the same time of cultivation, and one comparing mRNA samples taken at 11 h in Δ*spoIIE*C2 cultures

and 23 h in Δ*spoIIE*C5 cultures as both correspond to the onset of the stationary phase. Out of the 5597 coding genes annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), 5020 were detected in our transcriptomic data. These genes were then clustered according to the Clusters of Orthologous Groups (COGs) database (NCBI) to aid data interpretation.

Over 35% of the genes transcribed (1963 out of 5020) were differentially expressed at least once in Δ*spoIIE*C2 cultures compared to Δ*spoIIE*C5 cultures. The highest number of differentially expressed genes in Δ*spoIIE*C2 and the genes with the highest fold-change were observed at 11 h, whether we compared this sample to the 11 h or the 23 h Δ*spoIIE*C5 samples (Figure 7A).

Six functional groups represented over 70% of the differently expressed genes (Figure 7B): Energy production and conversion, Signal transduction mechanisms, Carbohydrate transport and metabolism, Amino acid transport and metabolism, Transcription and Unknown function. The genes encoding proteins of unknown functions represented 36% of the differentially expressed genes in Δ*spoIIE*C2, indicating that many actors of *C. beijerinckii'*s regulatory network remain to be discovered. Similarly, those functional groups concentrated the most differentially expressed genes in DG-8052 (124) and BA105 (122).

Out of these genes, 422 were differentially expressed at both time points. These 422 genes encode proteins involved in various cellular processes, but most of them (180 out of 422) belong to the 'unknown function' COG. Differentially expressed genes encoding proteins involved in signal transduction were mainly up-regulated at the studied time points. Most of the genes involved in translation, carbohydrate transport and metabolism, energy production and conversion, amino acid transport and metabolism were down-regulated (Figure 7C), indicating a decrease in cellular activity.
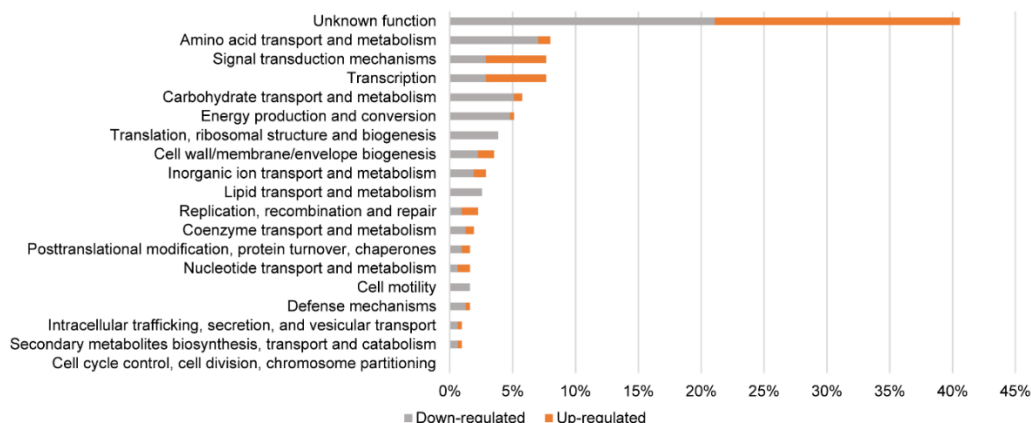
4

A.

B.

Figure 7: Overall differential expression dynamics. (A) Volcano plots of each comparison. Red dots outside of the blue area represent genes statistically differentially expressed with $|log_2 fold| > 1.5$ and $padj \leq 0.05$ (B) COG clustering of the genes differentially expressed at stationary phase.

Differential expression of the mutated genes

mRNA from all the genes harboring an SNP were detected in the analyzed samples (Table 4). However, only two of these genes were differentially expressed in $\Delta spoIIEC2$. Cbei_0201 was up-regulated at 4 h and 11 h (log_2 fold change of 1,6 and 1,5) as well as *cbei0200*, encoding for CdaA, with log_2 fold changes of 2,4 and 2,1, in $\Delta spoIIEC2$ cultures compared to $\Delta spoIIEC5$ cultures at the growth stages. The SNP found in Cbei_0201 probably impairs its regulating role and causes an up-regulation of the operon. An increase of CdaA in the cell might lead to an excess of c-di-AMP, which could partially explain $\Delta spoIIEC2$'s phenotype. Indeed, in *B. subtilis*, c-di-AMP is toxic and in excess leads to impaired cell growth (282), a change in colony morphology, and to cell lysis after 48 h of cultivation (280). Cbei_1712, encoding Spo0A, was differentially expressed in $\Delta spoIIEC2$ at stationary phase, compared to both $\Delta spoIIEC5$'s 11 h and 23 h samples, with a log_2 fold change of respectively -2,91 and -2,31. This result indicates that either the SNP detected in

spo0A might affect Spo0A activity. Still, one of the other mutations might have impacted the expression of *spo0A* since a decrease in *spo0A* expression was also observed in DG-8052 cultures (124), even though DG-8052 does not harbor any mutation in its *spo0A* gene.

Table 4: Differential expression of the genes harboring a mutation in Δ*spoIIE*C2. A log$_2$ fold change in italic indicates that the gene is not differentially expressed at that time point (log$_2$ fold <1,5 or padj > 0,05). The gene coverage indicates the average number of mRNAs per gene.

Gene Id	Predicted protein function	COG class description	Average gene coverage	Log$_2$ fold change in expression Δ*spoIIE*C2 relative to Δ*spoIIE*C5		
				4 h	11 h/ 11 h	11 h/ 23 h
Cbei_0048	Transcriptional regulator, TrmB	Transcription	2,8	0,41	-0,11	0,56
Cbei_0086	Transcription-repair coupling factor	Replication, recombination and repair	9,2	0,12	0,67	0,12
Cbei_0201	YbbR family protein, CdaR homolog	Function unknown	18,4	1,6	1,9	1,5
Cbei_0699	Phosphotransferase system, EIIC	Carbohydrate transport and metabolism	0,1	-0,44	-2,8	-0,59
Cbei_1193	Phosphatidate cytidylyltransferase	Lipid transport and metabolism	12,9	-0,42	0,32	0,45
Cbei_1712	Response regulator receiver protein, Spo0A	Signal transduction mechanisms	116,8	-0,6	-2,9	-2,3

Transcription

More than 40% of the genes belonging to the transcription COG were differentially expressed at either 4 h or 11 h, but most of them were differentially expressed at 11 h (Table S1). Genes encoding RNA polymerase subunits alpha, beta and beta' were down-regulated at 11 h in Δ*spoIIE*C2 cultures compared to Δ*spoIIE*C5 cultures at the same time point (Table 5). This result is coherent with the interruption of the exponential growth around 11 h observed in Δ*spoIIE*C2 cultures (Figure 2B), which would slow down the transcription machinery. Several sigma factors were also down-regulated in Δ*spoIIE*C2, including the sporulation specific sigma factors *sigF*, *sigE*, *sigG* and the motility modulating alternative sigma factor *sigI*. Compared to Δ*spoIIE*C5 cultures at 23 h, no significant difference in *sigI* expression was detected, which correlates to the earlier entrance of Δ*spoIIE*C2 cultures in the stationary phase.

Interestingly other genes encoding extracytoplasmic function (ECF) sigma factors were up-regulated. In bacteria, ECF sigma factors intervene in cells' stress response by redirecting transcription and coordinating the physiological changes required for the bacteria to adapt to its environmental conditions (283). In *Bacillus*, seven ECF were identified (283), whereas *C. beijerinckii* NCIMB 5052 harbors nineteen CDS encoding ECF sigma factors. No study on those sigma factors was realized in *C. beijerinckii*, but according to studies

4

conducted in *Bacillus,* these sigma factors are not vital for the cell in optimal growth conditions and intervene only in regulating the cell wall and membrane stress response (284, 285). Two ECF sigma factors, homologs to the σ^X and σ^Y proteins from *Bacillus*, seemed to be involved in the onset of sporulation and granulose degradation in *C. acetobutylicum* (286). And in *Clostridioides difficile*, three ECF sigma factors (σ^T, σ^U, σ^V) were responsible for the resistance to antimicrobials (287, 288). Among the ECF sigma factors up-regulated, we identified homologs of the σ^V, σ^X, σ^Y sigma factors. According to protein interaction prediction models (34), these ECF sigma factors interact with cell wall hydrolysis and chemotaxis proteins. Interestingly, we observed a concomitant up-regulation of genes involved in cell wall stress response, such as autolysins and efflux pumps (Table S5).

In contrast to the DG-8052 strain, neither *sigH* nor *sig54* were differentially expressed (124) in Δ*spoIIE*C2. Still, three σ^{54} specific transcriptional regulators (Cbei_1463, Cbei_2039 and Cbei_2180) were up-regulated in Δ*spoIIE*C2. Cbei_1463 and Cbei_2180 were up-regulated only at 11 h. Both genes were reported to regulate the expression of alcohol dehydrogenase encoding genes (126, 289), which were down-regulated in Δ*spoIIE*C2 at 11 h.

4

Table 5: Differential expression of genes RNA polymerase subunits (|log₂ fold|> 1,5 and p.adj<0,05) in Δ*spoIIE*C2 mutant relative to Δ*spoIIE*C5. If |log₂ fold| < 1,5 or p.adj>0,05, the log₂ fold change is in italic.

| Gene ID | Predicted protein function | Log₂ fold change in expression Δ*spoIIE*C2 relative to Δ*spoIIE*C5 | | |
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_0144	DNA-directed RNA polymerase, beta subunit	-1,2	-2,3	-2,0
Cbei_0145	DNA-directed RNA polymerase, beta' subunit	-1,4	-1,7	-1,7
Cbei_0180	DNA-directed RNA polymerase, alpha subunit	-1,5	-2,0	-1,8
Cbei_0313	RNA polymerase, ECF subfamily	-0,7	2,8	3,8
Cbei_0814	RNA polymerase sigma factor SigF	-1,6	-3,8	-1,7
Cbei_1120	RNA polymerase sigma factor SigE	-3,5	-6,7	-2,8
Cbei_1121	RNA polymerase sigma factor SigG	-3,0	-6,0	-3,6
Cbei_1736	RNA polymerase, ECF subfamily	1,1	-1,8	-3,2
Cbei_2829	RNA polymerase sigma factor, ECF subfamily, SigI	-2,3	-4,2	-0,9
Cbei_3102	RNA polymerase sigma factor, ECF subfamily, SigV	1,9	2,9	3,3
Cbei_3162	RNA polymerase, ECF subfamily, Sig Y homolog	-0,1	2,7	3,6
Cbei_3341	RNA polymerase, ECF subfamily	0,2	5,7	3,0
Cbei_3538	RNA polymerase, ECF subfamily	0,5	1,3	2,2
Cbei_3569	RNA polymerase, ECF subfamily	-0,4	7,1	5,1
Cbei_3576	RNA polymerase, ECF subfamily	0,7	7,2	6,3
Cbei_3675	RNA polymerase, ECF subfamily, SigX homolog	1,2	3,3	2,6

4

Genes encoding the transcriptional repressor LexA and DNA repair proteins RadC, RecF and RecN were up-regulated in Δ*spoIIE*C2 cultures at stationary phase, which indicates an activation of the SOS response. LexA is involved in DNA protection and repair in an unfavorable environment for the cells in bacteria. In *C. difficile,* a role in the regulation of other biological processes was described (290), and in *C. acetobutylicum*, LexA seems to be involved in the cell response to chemical stress (290), along with transcriptional regulators of amino acid metabolism. Next to *lexA*, homologs of *argR* and *cymR*, involved in the stress response mechanism in *C. acetobutylicum*, were up-regulated in Δ*spoIIE*C2.

Amino acids transport and metabolism

Genes associated with amino acid transport and metabolism were strongly differentially expressed in Δ*spoIIE*C2 cells compared to Δ*spoIIE*C5 cells. Over 40% of the genes belonging to the amino acid transport and metabolism cluster were differentially expressed at stationary phase. Genes encoding amino acid transporters were mainly down-regulated. At 4 h and 11 h, the genes encoding branched-chain amino acids (BCAA), arginine, glutamine, proline transporters were down-regulated (Table S2). Besides, the expression of genes coding for amino acid biosynthesis proteins was affected. BCAAs biosynthesis was

impacted as the genes encoding for enzymes involved in their biosynthesis were strongly down-regulated (log2 fold varying between -5 and -10). In Clostridia, BCAA are necessary for the shift to solventogenesis and play a role in stress response (59, 291). The decrease of BCAA biosynthesis and transport could lead to growth limitation and inadequate stress response, as observed in Δ*spoIIE*C2. Only methionine transport and biosynthesis were up-regulated at the stationary phase, which could be explained by methionine's putative role as stress protectant. Indeed, in *C. acetobutylicum*, methionine was proven to help the strain survival under stress conditions (59).

Carbohydrate transport and metabolism

Sugar transport

In this study, we identified 80 genes encoding sugar transporters that were differentially expressed (Table S3). These include genes coding for PTS and MFS transporters. The expression of five mannose/fructose/sorbose and five lactose/cellobiose PTS operons were down-regulated at 4 h and 11 h, which contrasts with the up-regulation of PTS encoding genes in the degenerated BA105 (122). Contrary to the DG-8052 cultures (124), σ^{54} regulated operons were not differentially expressed in Δ*spoIIE*C2 cultures.

Glycolysis

The genes involved in glycolysis were mainly down-regulated in Δ*spoIIE*C2 cultures compared to Δ*spoIIE*C5 cultures (Figure 8). Genes encoding key enzymes were down-regulated. Indeed, the *gap-pfk-tpi* operon was down-regulated in Δ*spoIIE*C2 at both time points. GAP is responsible for the oxidation of glyceraldehyde-3-phosphate coupled to the generation of NADH. A lack of GAP could hamper the glycolytic flux and cause a decrease in the NADH level in the cell. A decrease in NADH generation was reported in degenerated *C. saccharoperbutylacetonicum* strains, and the addition of benzyl viologen, a redox dye, could partially restore butanol production (272). The primary copy of the *pfk,* Cbei_0998, also encodes one of the rate-limiting enzymes of the glycolysis and was down-regulated at 11 h in Δ*spoIIE*C2. Reduction in the glycolytic flux because of limited amounts of GAP and PFK would decrease ATP and pyruvate yield. This decrease may explain the early interruption of the exponential growth phase and the limited biomass formation. Surprisingly, a decrease in the *gap-pfk-tpi* was only reported in DG-8052 (124) but not BA105 (122).

Acid and solvent formation

In Δ*spoIIE*C2 cells, the genes encoding enzymes required for acid production were not differentially expressed at 4 h (Figure 8). At stationary phase, the *ptb-buk1* was not differentially expressed Since the *ptb-buk1* operon belongs to the Spo0A regulon (36, 118),

Spo0A might be still partially active in Δ*spoIIE*C2 cells. An up-regulation of the *pta-ack* operon was observed in Δ*spoIIE*C2 compared to Δ*spoIIE*C5 at the same time point (Figure 8). In contrast, the genes encoding enzymes involved in converting the acids into solvents were clearly down-regulated in Δ*spoIIE*C2 in all studied time points compared to Δ*spoIIE*C5. In particular, the genes belonging to the sol operon (Cbei_3832-3835) were strongly down-regulated at 4 h and 11 h, with a $\log_2$ fold change ranging between -4,7 and -1,7. The expression profile in Δ*spoIIE*C2 of the genes involved in acid and solvent formation corresponds to the fermentation profile presented in Figure 2 and Table 2. These variations in gene expression have been described in other studies of degenerated *C. beijerinckii* strains (122–124), suggesting that these variations are typical for degenerated strains.

Granulose

In Δ*spoIIE*C2 cultures, granulose biosynthesis and degradation genes were differentially expressed compared to the Δ*spoIIE*C5 cultures. Genes involved in the granulose formation were up-regulated, while genes involved in granulose degradation were down-regulated. The same expression pattern was observed when compared to wild-type cultures (data not shown). These results explain the delay in granulose formation observed in Δ*spoIIE*C2 colonies. Curiously, other genes encoding enzymes from the granulose pathway were down-regulated in the degenerated *C. acetobutylicum* M5 strain compared to the wild-type (292).

4
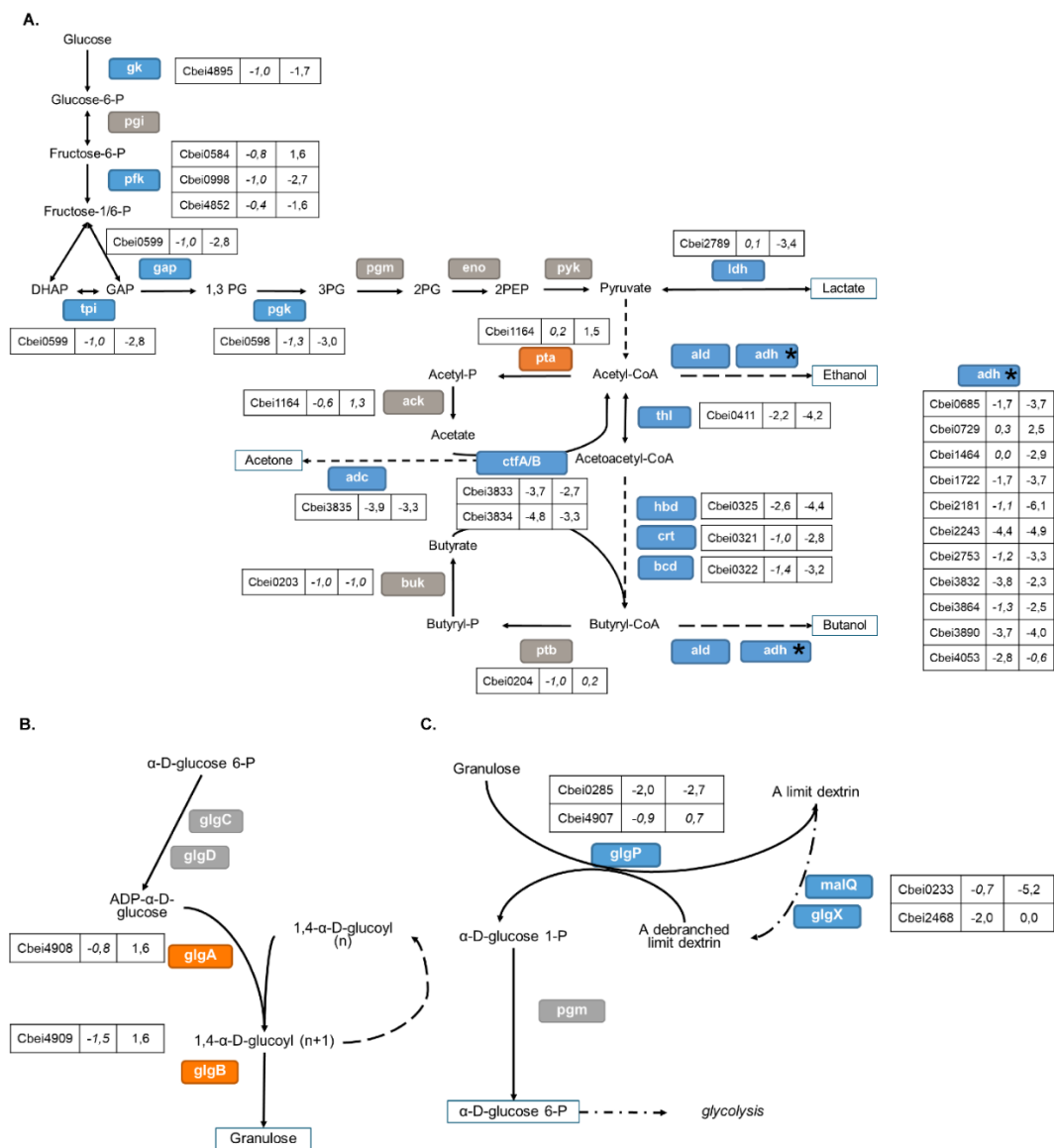
Figure 8: Differential expression in ΔspoIIEC2 at 4 and 11 h of genes involved in carbohydrate metabolism compared to ΔspoIIEC5 at 4 and 23 h. Figure adapted from Microcyc. (A) glycolysis and ABE metabolic pathway. The asterisk indicates the presence of several paralogues, which are listed on the side of the figure. (B) Granulose formation (C) Granulose degradation. Gene names are in colored boxes according to the differential expression; genes in grey boxes have no change in expression, genes in orange boxes are up-regulated and genes in blue are down-regulated. The $\log_2$fold change in expression is indicated to the differentially expressed genes ($|\log_2$ fold$|> 1,5$ and p.adj$<0,05$). If $|\log_2$ fold$| < 1,5$ or p.adj$>0,05$, the $\log_2$ fold change is in italic.

Signal transduction mechanisms and secondary metabolites

Clostridia initiate the sporulation cycle and solventogenesis in response to environmental changes. Signal recognition and transduction mechanisms sense and relay these changes throughout the cell to regulate cell processes. In *C. acetobutylicum*, three signal transduction mechanisms have been studied, an Agr system (90), an BtrK/BtrM system (91) and an RRNPP type system (252). Besides, secondary metabolites such as polyketides regulate solventogenesis and granulose formation in *C. acetobutylicum* (83) and *C. saccharoperbutylacetonicum* (82). Even though no signal transduction system has been studied in *C. beijerinckii* yet, several two-component systems are annotated on *C. beijerinckii's* genome, among which homologs to the Agr and BtrK operons were identified. Furthermore, a different polyketide cluster, a PKS-NRPS hybrid, was detected in the genome of *C. beijerinckii* (80). In our study previous of the *spoIIE* mutant (60), *spoIIE* disruption had no impact on the expression of Agr genes but impacted the expression of transport and biosynthetic genes of the PKS-NRPS gene cluster. In Δ*spoIIE*C2, an up-regulation of *agrB* was detected at stationary phase (Table S4), but no changes in expression were detected in *agrA* or *agrB* genes. AgrB has been identified as a membrane-associated protein processing the autoinducing peptide (AIP) signal molecule (90). The disruption of the Agr quorum did not have any impact on solvent formation in *C. acetobutylicum*. However, the impact of an over-expression of *agrB* has not been studied in *Clostridium,* and no reports about differentially expressed quorum-sensing genes were made in other degenerated *C. beijerinckii* strains.

Even though no change in the expression of the cell motility regulating sigma factor, *sigD*, was observed in Δ*spoIIE*C2, genes encoding chemotaxis proteins were strongly differentially expressed at 11 h. Genes coding for the flagellum regulation proteins CheW and CheA were up-regulated at 11 h (Table S4), as observed in the DG-8052 strain at 12 h (124), while genes for flagellum biosynthesis were down-regulated at 11 h.

Genes from the PKS-NRPS cluster were differentially expressed in the mutant (Table S4). Indeed, at 4 h, genes encoding transport and biosynthetic genes were down-regulated in Δ*spoIIE*C2. At 11 h, genes coding for transporters and the regulatory protein were up-regulated while most genes encoding biosynthetic proteins were down-regulated. No report on the differential expression of this cluster was made in other degenerated *C. beijerinckii* strains, but according to our results, this PKS-NRPS cluster might contribute to the regulation of solventogenesis.

4

Early sporulation genes and stress response

In Δ*spoIIE*C5 cells, genes involved in sporulation initiation and stage I of the sporulation circle were expressed at wild-type levels. In Δ*spoIIE*C2 cultures, several genes involved in sporulation initiation were differentially expressed compared to Δ*spoIIE*C5 cultures (Table S6). The activation of Spo0A by phosphorylation in *Clostridium* requires several transcriptional regulators such as CodY, AbrB and SinR (113, 293) and orphan kinases (143, 294). *C. beijerinckii* harbors several genes encoding orphan kinases in its genome (125), but only one copy was differentially expressed in Δ*spoIIE*C2 cultures. At both time points, Cbei_3169, encoding an orphan kinase, was strongly down-regulated (log$_2$ fold -3,2 and -3,8). The three AbrB homologs were also differentially expressed in *C. beijerinckii* Δ*spoIIE*C2 cultures; *cbei3375* was down-regulated at 4 h. Conversely, the two other copies, *cbei0088* and *cbei4885,* were strongly up-regulated at stationary phase (Table S6). Similarly, an up-regulation of *cbei4885* was observed in another degenerated *C. beijerinckii* strain (122). The gene coding for Spo0A was down-regulated at stationary phase in Δ*spoIIE*C2 cultures as well as the sporulation genes encoding operons, *spoIIAA-B-sigF* and *spoIIGA-sigE-sigG* operons. Moreover, most of σ^F and σ^E regulons were strongly down-regulated, with a log$_2$ fold change ranging between -5,5 and -1,7, as observed in DG-8052 (124). Surprisingly, three genes encoding sporulation proteins, *cfsB*, *ytxs,* and *ylmC,* were up-regulated.

Simultaneously, genes coding for the chaperonin proteins Cpn10, GrpE and Hsp20 were strongly up-regulated in Δ*spoIIE*C2 at stationary phase along with *lonA* and *sodA* genes with a log$_2$ fold change ranging from 1,5 to 3 (Table S6). In contrast, genes encoding rubrerythrins were down-regulated. The upregulation of genes coding for stress response proteins (chaperonins, proteases and the superoxide dismutase) was also reported in the *C. acetobutylicum* degenerated M5 strain (292) and the degenerated DG-8052 strain (124). Due to a fast pH drop and the low pH, Δ*spoIIE*C2 cells activate their stress response mechanism early during cultivation.

Conclusion

This study describes the phenotype and genotype of a Δ*spoIIE C. beijerinckii* degenerated strain, Δ*spoIIE*C2, and offers additional information on strain degeneration. Using fermentation and microscopy analysis, we distinguished between characteristics linked to sporulation interruption (cell morphology, absence of spores) and characteristics generally linked to the inability to form solvents (pH crash, limited growth and early cell lysis). Compared to Δ*spoIIE*C5 cultures, Δ*spoIIE*C2 cultures had a poor growth even in pH-controlled conditions and produced acids as major products as reported for other degenerated *Clostridium* strains (81, 155, 272, 292). Genome analysis revealed seven single-nucleotide mutations unique to Δ*spoIIE*C2, which probably caused the strain degeneration. However, none of these mutations were detected in the recently studied degenerated *C. beijerinckii* strains, DG-8052 and BA105 (124, 132). To study further the consequences of strain degeneration, we analyzed Δ*spoIIE*C2's transcriptome and compared it to the solventogenic Δ*spoIIE*C5 strain (119) and other degenerated *Clostridium* strains described in previous studies (122, 124, 278, 292). Degeneration concurred with a vast reorganization of the transcriptome, with one-third of Δ*spoIIE*C2's genome being differentially expressed at least at once, and gene expression mostly affected at stationary phase. The mutated *cdaR* and *spo0A* were both differentially expressed. While *cdaR* was up-regulated at both studied time points and *spo0A* was down-regulated at stationary phase. These observations suggest that those genes might be responsible for the degeneration phenotype. However, more studies would be necessary to assess the impact of the detected mutation on the encoded proteins and the metabolome. The most affected functional groups were the COG of unknown function, transcription, carbohydrate- and amino acid metabolism, and signal transduction. In accordance with previous reports on degenerated *Clostridium* strains, strain degeneration caused a decrease in cellular metabolism, and cellular activity was mainly limited to signaling and adaptation to environmental stressors (122, 124, 292). Besides, our results report the putative role of ECF sigma factors and c-di-AMP (through CdaR inactivity) in degeneration and cell envelope stress, actors that have not been investigated yet in solventogenic clostridia. This study gives thus clues on new targets for studying strain degeneration and stress response in solventogenic clostridia. A better understanding of the physiology of the degeneration will allow developing strategies to control this phenomenon to create stable commercial strains for the production of biochemicals.

4

Nucleotide sequence

The RNA-seq data described in this study have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7481 and E-MTAB-7481. The genome sequence of *Clostridium beijerinckii* NCIMB 8052 Δ*spoIIE*C5 and the Δ*spoIIE*C2 strains described in this study is available on the European Nucleotide Archive (ENA) under the accession number PRJEB39199 and PRJEB39200.

Acknowledgments

4

Supplementary files

Table S1: Most differentially expressed genes involved in transcription, with |log2 fold| > 2,5 and p.adj<0,05 at least once in Δ*spoIIE*C2 cultures relative to Δ*spoIIE*C5 cultures. If |log₂ fold| <1,5 or p.adj>0,05, the gene is not considered differentially expressed and the log₂ fold change is in italic.

Gene ID	Predicted protein function	log₂ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_0035	Helix-turn-helix domain protein	0,6	3,9	3,4
Cbei_0047	Alanine racemase	-0,4	-3,7	-1,6
Cbei_0088	Transcriptional regulator, AbrB family	0,7	4,4	4,9
Cbei_0120	Transcriptional repressor, CtsR	-0,1	6,1	3,2
Cbei_0243	Transcriptional antiterminator, BglG	1,0	3,0	2,6
Cbei_0259	transcriptional regulator, TetR family	-1,4	3,9	2,3
Cbei_0278	transcriptional regulator, MarR family	3,2	9,8	8,5
Cbei_0286	helix-turn-helix- domain containing protein, AraC type	0,7	1,9	4,0
Cbei_0302	transcriptional regulator, MarR family	2,4	2,7	3,3
Cbei_0453	transcriptional regulator, DeoR family	0,6	3,2	1,6
Cbei_0457	transcriptional regulator, DeoR family	1,9	4,8	2,6
Cbei_0466	transcriptional regulator, DeoR family	0,2	0,8	3,4
Cbei_0517	transcriptional regulator, LysR family	-1,1	-3,5	-3,7
Cbei_0633	ribonuclease R	0,3	3,2	1,9
Cbei_0678	transcriptional regulator, AraC family	1,2	-0,5	4,4
Cbei_0722	transcriptional regulator, LysR family	-1,3	6,1	1,5
Cbei_0731	Alanine racemase	-0,3	-2,5	-2,4
Cbei_0749	transcriptional regulator, RpiR family	-0,3	3,7	1,9
Cbei_0752	transcriptional antiterminator, BglG	0,0	-3,0	-2,8
Cbei_0779	transcriptional regulator, MarR family	1,5	3,3	2,0
Cbei_0969	helix-turn-helix domain protein	1,2	1,0	2,9
Cbei_0973	transcriptional regulator, RpiR family	0,0	3,1	2,5
Cbei_1067	transcriptional regulator, MarR family	0,7	-4,9	-1,6
Cbei_1199	NusA antitermination factor	0,4	-3,6	-1,6
Cbei_1200	protein of unknown function DUF448	-1,1	-4,3	-2,7
Cbei_1253	transcriptional regulator, TetR family	0,8	5,2	2,8
Cbei_1307	helix-turn-helix- domain containing protein, AraC type	3,0	1,9	1,5
Cbei_1448	regulatory protein, ArsR	0,9	2,8	2,2
Cbei_1450	putative transcriptional regulator, Crp/Fnr family	0,5	3,1	2,9
Cbei_1463	GAF modulated sigma54 specific transcriptional regulator, Fis family	0,7	5,2	2,9
Cbei_1500	transcriptional regulator, PadR-like family	0,5	2,7	3,9
Cbei_1678	putative transcriptional regulator, XRE family	0,5	3,1	3,5
Cbei_1709	arginine repressor, ArgR	1,4	2,7	2,9

Gene ID	Predicted protein function	log₂ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_1748	transcriptional regulator, BadM/Rrf2 family, CymR	0,2	2,4	2,8
Cbei_1987	regulatory protein, LysR	-2,5	-4,6	-3,3
Cbei_2010	putative transcriptional regulator, AsnC family	-1,0	-3,1	-1,9
Cbei_2039	sigma54 specific transcriptional regulator, Fis family	2,1	2,7	2,7
Cbei_2054	transcriptional regulator, LysR family	-0,5	-5,0	-3,6
Cbei_2108	regulatory protein, ArsR	-3,4	6,1	1,8
Cbei_2180	GAF modulated sigma54 specific transcriptional regulator, Fis family	0,5	3,9	2,4
Cbei_2215	helix-turn-helix domain protein	2,0	1,2	3,2
Cbei_2221	transcriptional regulator, BadM/Rrf2 family	-1,6	-1,9	-2,5
Cbei_2265	helix-turn-helix domain protein	1,1	0,2	3,3
Cbei_2362	helix-turn-helix- domain containing protein, AraC type	1,5	4,1	5,6
Cbei_2373	putative transcriptional regulator, MerR family	2,7	0,3	3,0
Cbei_2385	ROK family protein	1,9	4,0	2,7
Cbei_2390	transcriptional regulator, GntR family	0,4	-5,7	-3,2
Cbei_2475	transcriptional regulator, TetR family	4,2	-0,5	0,5
Cbei_2479	transcriptional regulator, TetR family	1,8	-2,7	-0,7
Cbei_2561	SOS-response transcriptional repressor, LexA	-0,4	3,7	2,2
Cbei_2834	transcriptional antiterminator, BglG	0,9	0,2	3,8
Cbei_2931	transcriptional regulator, TetR family	-0,8	-1,8	-3,0
Cbei_3005	ROK family protein	1,2	2,9	2,9
Cbei_3088	iron dependent repressor	-1,4	-3,1	-3,5
Cbei_3095	helix-turn-helix- domain containing protein, AraC type	0,2	3,9	1,7
Cbei_3104	transcriptional repressor, CopY family	1,5	4,0	1,9
Cbei_3110	putative transcriptional regulator, PucR family	2,8	0,4	1,4
Cbei_3181	transcriptional regulator, LysR family	-1,8	-3,3	-3,7
Cbei_3183	putative transcriptional regulator, MerR family	-0,6	-2,2	-2,7
Cbei_3221	transcriptional regulator, LysR family	-3,1	0,9	-1,0
Cbei_3291	GCN5-related N-acetyltransferase	0,8	3,0	2,7
Cbei_3301	transcriptional regulator, MarR family	2,1	3,0	0,6
Cbei_3315	transcriptional regulator, HxlR family	-1,8	-1,0	-3,0
Cbei_3318	transcriptional regulator, TetR family	-0,8	-3,2	-2,6
Cbei_3375	transcriptional regulator, AbrB family	-2,5	-0,3	-0,5
Cbei_3421	hypothetical protein	-0,6	2,8	1,9
Cbei_3479	transcriptional regulator, XRE family	0,7	4,1	3,6
Cbei_3480	transcriptional regulator, MarR family	2,5	3,0	2,4
Cbei_3483	transcriptional regulator, XRE family	0,0	2,6	1,8
Cbei_3485	transcriptional regulator, HxlR family	-2,7	-2,5	-2,3

4

Gene ID	Predicted protein function	log$_2$ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_3536	regulatory protein, ArsR	0,7	6,2	2,7
Cbei_3549	transcriptional regulator, XRE family	1,9	7,8	3,1
Cbei_3573	transcriptional regulator, PadR-like family	2,1	2,2	3,0
Cbei_3593	putative transcriptional regulator, TetR family	3,4	4,7	5,1
Cbei_3598	putative transcriptional regulator, MerR family	0,9	0,4	2,9
Cbei_3606	GCN5-related N-acetyltransferase	-2,0	-2,8	-1,5
Cbei_3616	transcriptional regulator, XRE family	0,5	3,6	1,9
Cbei_3627	conserved hypothetical protein	2,2	4,9	3,8
Cbei_3644	transcriptional regulator, TetR family	0,4	4,7	1,8
Cbei_3649	transcriptional regulator, MarR family	2,8	5,7	5,0
Cbei_3815	transcriptional regulator, GntR family	0,6	4,6	3,0
Cbei_3886	putative transcriptional regulator, TetR family	3,1	4,1	5,2
Cbei_3896	transcriptional regulator, PadR-like family	1,1	2,5	2,7
Cbei_3898	transcription activator, effector binding	0,1	1,4	3,1
Cbei_3910	transcriptional regulator, RpiR family	-1,4	-3,9	-2,4
Cbei_3913	pentapeptide repeat protein	0,2	0,6	2,8
Cbei_4085	transcriptional regulator, GntR family	0,4	3,9	2,4
Cbei_4108	transcriptional regulator, HxlR family	-2,7	0,2	-0,7
Cbei_4129	Transcriptional regulator IclR-like protein	-0,1	3,9	2,0
Cbei_4148	Alanine racemase	0,7	3,0	1,5
Cbei_4150	putative transcriptional regulator, AsnC family	0,5	-2,0	-2,9
Cbei_4422	Alanine racemase	-2,1	-2,9	-1,8
Cbei_4446	transcriptional regulator, AraC family	-1,3	-4,3	-3,9
Cbei_4501	helix-turn-helix- domain containing protein, AraC type	2,6	1,6	2,8
Cbei_4526	transcriptional regulator, GntR family	0,1	2,5	2,5
Cbei_4534	transcriptional antiterminator, BglG	-1,7	-2,8	-1,1
Cbei_4561	transcriptional regulator, GntR family	0,4	4,2	1,9
Cbei_4642	transcriptional regulator, RpiR family	0,7	3,0	3,0
Cbei_4650	regulatory protein, LacI	0,8	3,6	2,4
Cbei_4692	transcriptional regulator, TetR family	1,4	-0,8	3,3
Cbei_4702	transcriptional regulator, RpiR family	1,3	1,8	4,0
Cbei_4797	transcriptional regulator, GntR family	0,3	2,8	1,5
Cbei_4885	transcriptional regulator, AbrB family	0,2	0,6	2,8
Cbei_4935	regulatory protein, LuxR	-1,2	-4,2	-2,3
Cbei_4996	helix-turn-helix domain protein	0,7	4,3	5,3
Cbei_5094	parB-like partition protein	0,4	2,8	2,4
Cbei_5096	parB-like partition protein	0,8	1,9	2,5

4

Table S2: Most differentially expressed genes involved in amino acid metabolism, with |log2 fold| > 2,5 and p.adj<0,05 at least once in ΔspoIIEC2 cultures relative to ΔspoIIEC5 cultures. If |log₂ fold| <1,5 or p.adj>0,05, the gene is not considered differentially expressed and the log₂ fold change is in italic. NA: No protein assigned; n.d: the differential expression could not be calculated.

Gene ID	Predicted protein function	Putative encoded protein	log₂ fold change in expression		
			4 h	11 h/ 11 h	11h/ 23 h
Alanine biosynthesis & transport					
Cbei_1098	Aminotransferase, class V	NA	-0,3	4,7	2,1
Cbei_1851	Cysteine desulfurase, SufS subfamily	NA	-0,7	8,1	1,2
Cbei_2599	Cysteine desulfurase family protein	NA	-0,5	-1,0	-2,1
Cbei_2608	Aspartate 1-decarboxylase	PanD	2,9	2,0	0,8
Cbei_2870	Substrate-binding region of ABC-type glycine betaine transport system	NA	-3,4	-7,2	-7,2
Cbei_3972	Amino acid permease-associated region	NA	-1,6	-4,6	-4,3
Arginine biosynthesis & transport					
Cbei_4515	Argininosuccinate synthase	NA	0,0	-1,2	-6,0
Cbei_4516	Argininosuccinate lyase	NA	-2,0	-2,0	-6,5
Cbei_4517	N-acetyl-gamma-glutamyl-phosphate reductase	ArgC	-0,9	-1,4	-6,4
Cbei_4518	Arginine biosynthesis bifunctional protein ArgJ	ArgJ	-0,4	-1,2	-6,3
Cbei_4519	Acetylglutamate kinase	ArgB	-1,7	-1,3	-6,4
Cbei_4520	Acetylornithine and succinylornithine aminotransferase	NA	-1,0	-1,3	-6,5
Cbei_4521	Ornithine carbamoyltransferase	OtC	-0,8	-1,6	-6,8
Cbei_4781	Extracellular solute-binding protein, family 3	NA	-1,5	-2,6	-4,5
Asparagine biosynthesis					
Cbei_1034	Asparagine synthase (glutamine-hydrolyzing)	NA	-1,0	-2,9	-2,4
BCAA biosynthesis & transport					
Cbei_0212	Acetolactate synthase, small subunit	NA	0,5	-8,7	-8,1
Cbei_0213	2-isopropylmalate synthase/homocitrate synthase family protein	NA	0,1	-6,9	-5,8
Cbei_0214	3-isopropylmalate dehydratase, large subunit	LeuC	-0,8	-7,7	-7,2
Cbei_0215	3-isopropylmalate dehydratase, small subunit	LeuD	-1,7	-8,4	-7,6
Cbei_0216	3-isopropylmalate dehydrogenase	NA	-1,6	-7,9	-7,1
Cbei_0217	Acetolactate synthase, large subunit, biosynthetic type	NA	-0,6	-7,3	-6,2
Cbei_0218	Ketol-acid reductoisomerase	IlvC	-2,2	-8,8	-8,3
Cbei_1043	Branched-chain amino acid aminotransferase	NA	-1,1	-2,1	-2,8
Cbei_1510	Dihydroxy-acid dehydratase	IlvD	-2,6	-5,4	-5,4
Cbei_1562	Threonine dehydratase	NA	-0,3	-2,6	-2,4
Cbei_2175	2-isopropylmalate synthase	LeuA	-0,1	-2,6	-6,2
Cbei_2646	Acetolactate synthase, large subunit, biosynthetic type	NA	-1,3	-3,9	-4,1
Cbei_2647	Acetolactate synthase, small subunit	NA	-1,0	-5,0	-3,8
Cbei_4095	Dihydroxy-acid dehydratase	NA	-0,9	2,5	1,8

Gene ID	Predicted protein function	Putative encoded protein	log$_2$ fold change in expression		
			4 h	11 h/ 11 h	11h/ 23 h
Cbei_4118	Aminotransferase, class IV	NA	2,0	2,9	3,7
Cbei_4127	Dihydroxy-acid dehydratase	NA	-1,0	4,3	0,4
Cbei_1474	Branched-chain amino acid transport system carrier protein	BrnQ	-1,7	-6,7	-5,2
Cbei_1763	Inner-membrane translocator	NA	-2,2	-2,2	-3,2
Cbei_1764	Inner-membrane translocator	NA	-1,4	-1,6	-2,5
Cbei_5042	Extracellular ligand-binding receptor	NA	-1,3	-8,1	-10,6
Cbei_5043	Inner-membrane translocator	NA	-4,2	-6,8	-9,7
Cbei_5044	Inner-membrane translocator	NA	-2,5	-7,7	-9,5
Cbei_5045	ABC transporter related	NA	-1,0	-6,2	-7,8
Cysteine biosynthesis					
Cbei_0247	Serine O-acetyltransferase	NA	-1,2	-3,6	-1,2
Cbei_0577	Cysteine synthase	NA	0,1	5,4	3,3
Cbei_0578	Serine O-acetyltransferase	NA	-0,7	5,5	2,7
Cbei_4356	Cysteine synthase A	NA	0,0	6,0	2,6
Glutamine biosynthesis & transport					
Cbei_0444	Glutamine synthetase, catalytic region	NA	-1,4	-3,8	-3,4
Cbei_1524	Glutamate:proton symporter	GltP	-2,7	-5,4	-5,8
Cbei_2183	Glutamate synthase (NADPH), homotetrameric	NA	-0,8	-4,1	-5,7
Cbei_2391	Extracellular solute-binding protein, family 3	NA	0,4	-5,1	-2,9
Cbei_2510	Polar amino acid ABC transporter, inner membrane subunit	NA	-3,0	-2,1	-1,6
Cbei_4170	Polar amino acid ABC transporter, inner membrane subunit	NA	-0,3	-5,3	-3,1
Cbei_4171	Polar amino acid ABC transporter, inner membrane subunit	NA	1,7	-3,2	-0,7
Cbei_4172	Extracellular solute-binding protein, family 3	NA	0,3	-3,2	-0,4
Cbei_4173	ABC transporter related	NA	1,3	-4,1	-0,8
Cbei_4204	Glutamate synthase (ferredoxin)	NA	0,0	-4,2	-3,9
Cbei_4206	Glutamine synthetase, type I	NA	-0,5	-5,5	-5,3
Cbei_4779	ABC transporter related	NA	-1,1	-1,4	-2,8
Cbei_4780	Polar amino acid ABC transporter, inner membrane subunit	NA	-1,6	-3,5	-4,1
Histidine biosynthesis					
Cbei_1317	Histidinol dehydrogenase	NA	-0,3	-2,5	-4,6
Cbei_1318	Aminotransferase, class I and II	NA	0,1	-2,0	-3,9
Cbei_1319	Imidazoleglycerol-phosphate dehydratase	HisB	0,4	-2,3	-3,9
Cbei_1812	histidinol-phosphate aminotransferase	HisC	0,6	3,1	2,4
Cbei_2106	histidinol phosphate phosphatase HisJ family	NA	1,2	2,7	2,3
Lysine biosynthesis					
Cbei_3178	Lysine exporter protein (LYSE/YGGA)	NA	-1,1	-4,6	n.d
Cbei_3414	Lysine exporter protein (LYSE/YGGA)	NA	0,3	-2,4	0,4

4

Gene ID	Predicted protein function	Putative encoded protein	log$_2$ fold change in expression		
			4 h	11 h/ 11 h	11h/ 23 h
Cbei_4128	dihydrodipicolinate synthetase	NA	1,7	6,5	4,7
Methionine biosynthesis & transport					
Cbei_0625	Binding-protein-dependent transport systems inner membrane component	NA	0,8	2,9	2,1
Cbei_0629	Cystathionine gamma-synthase	NA	-0,6	2,7	1,1
Cbei_0637	Binding-protein-dependent transport systems inner membrane component	NA	0,1	3,0	1,7
Polar amino acids transport					
Cbei_1050	Polar amino acid ABC transporter, inner membrane subunit	NA	-0,6	-0,2	-3,0
Cbei_1051	ABC transporter related	NA	-0,7	0,3	-2,8
Proline biosynthesis & transport					
Cbei_0024	Glutamate 5-kinase	NA	0,7	3,1	1,7
Cbei_0025	Gamma-glutamyl phosphate reductase	NA	-0,3	2,6	0,8
Cbei_0876	Sodium/proline symporter	NA	-2,9	-5,5	-6,1
Cbei_2871	Glycine betaine/L-proline ABC transporter, ATPase subunit	NA	-2,1	-6,7	-6,5
Threonine biosynthesis					
Cbei_1293	Threonine synthase	NA	0,4	3,6	2,9
Cbei_1294	Homoserine kinase	ThrB	0,3	4,5	3,5
Tryptophan biosynthesis & transport					
Cbei_0652	Putative tryptophan transport protein	NA	0,9	-3,2	-5,5
Cbei_1750	Glutamine amidotransferase of anthranilate synthase	NA	-0,5	-5,3	-6,4
Cbei_1751	Anthranilate phosphoribosyltransferase	TrpD	-0,1	-4,8	-5,4
Cbei_1752	Indole-3-glycerol-phosphate synthase	NA	0,7	-5,1	-5,2
Cbei_1753	Phosphoribosylanthranilate isomerase	TrpF	1,2	-5,3	-5,8
Cbei_1754	Tryptophan synthase, beta subunit	TrpB	-0,3	-5,9	-6,5
Cbei_1755	Tryptophan synthase, alpha subunit	TrpA	0,5	-5,8	-6,2
Cbei_4119	Para-aminobenzoate synthase, subunit I	NA	0,9	2,5	3,4

4

Table S3: Most differentially expressed genes involved in the pentose phosphate pathway and sugar transport, with |log2 fold| > 2,5 and p.adj<0,05 at least once in Δ*spoIIEC*2 cultures relative to Δ*spoIIEC*5 cultures. If |log$_2$ fold| <1,5 or p.adj>0,05, the gene is not considered differentially expressed and the log$_2$ fold change is in italic. NA: No protein assigned; n.d: the differential expression could not be calculated.

Gene ID	Predicted protein function	Putative encoded protein	log$_2$ fold change in expression		
			4 h	11 h/ 11 h	11 h/ 23 h
Pentose Phosphate Pathway					
Cbei_0529	Ribulose-phosphate 3-epimerase	Rpe	0,2	-5,1	-2,7
Cbei_0224	Transketolase domain protein	Tkt	1,0	3,1	1,1
Cbei_4871	Transketolase domain protein	Tkt	-0,3	2,5	3,0
Cbei_2386	Putative transaldolase	Tkt	0,8	3,9	1,7
Cbei_0317	Transaldolase	TalB	0,6	1,0	-4,2
Cbei_4454	Putative transaldolase	TalB	0,1	2,7	0,7
Cbei_4645	Putative transaldolase	TalB	0,3	2,6	0,9
Sugar transport					
Cbei_0109	Sugar transporter	NA	-3,2	1,3	2,2
Cbei_0230	Extracellular solute-binding protein, family 1	NA	-1,6	-5,2	-7,8
Cbei_0244	Phosphoenolpyruvate-dependent sugar PTS system, EIIA 2	NA	1,0	3,9	2,2
Cbei_0274	Major facilitator superfamily MFS_1	NA	-2,8	1,4	0,4
Cbei_0337	PTS system, glucitol/sorbitol-specific, IIBC subunit	NA	-1,6	-2,6	-2,1
Cbei_0707	Major facilitator superfamily MFS_1	NA	-0,9	-2,0	-5,9
Cbei_0711	PTS system, mannose/fructose/sorbose family, IIA subunit	NA	-0,7	-4,3	-4,9
Cbei_0712	PTS system PTS, sorbose-specific IIC subunit	NA	-1,9	-4,9	-5,2
Cbei_0713	PTS system, mannose/fructose/sorbose family, IID subunit	NA	-1,9	-5,1	-5,0
Cbei_0723	Drug resistance transporter, EmrB/QacA subfamily	NA	-0,6	3,6	2,0
Cbei_0732	Extracellular solute-binding protein, family 1	NA	-1,6	-5,0	-4,6
Cbei_0751	PTS system, glucose subfamily, IIA subunit	NA	-0,8	-3,3	-2,7
Cbei_0758	PTS system, lactose/cellobiose-specific IIB subunit	NA	-0,7	-2,4	-3,3
Cbei_0939	D-galactose-binding periplasmic protein precursor	NA	-0,4	-3,3	-1,6
Cbei_0966	PTS system mannose/fructose/sorbose family IID component	NA	0,3	-2,3	-2,9
Cbei_0977	Major facilitator superfamily MFS_1	NA	-2,2	-3,4	-2,5
Cbei_1219	PTS system, phosphocarrier protein HPr	HpR	-2,8	-2,7	-2,8
Cbei_1440	Extracellular solute-binding protein, family 1	NA	-1,7	-3,0	-4,9
Cbei_1770	Major facilitator superfamily MFS_1	NA	-0,9	-4,7	-2,0
Cbei_1835	Sugar (Glycoside-Pentoside-Hexuronide) transporter	NA	-0,9	1,5	-3,0
Cbei_1840	PTS system, mannose/fructose/sorbose family, IIB subunit	NA	2,1	-1,4	-3,4

4

Gene ID	Predicted protein function	Putative encoded protein	log$_2$ fold change in expression		
			4 h	11 h/ 11 h	11 h/ 23 h
Cbei_1954	Major facilitator superfamily MFS_1	NA	-0,8	-3,1	-3,5
Cbei_2226	Protein of unknown function DUF6, transmembrane	NA	-1,0	-6,0	0,0
Cbei_2320	PTS system, trehalose-specific IIBC subunit	NA	-5,1	-1,1	-2,4
Cbei_2357	Extracellular solute-binding protein, family 1	NA	-0,1	-2,3	-2,8
Cbei_2499	PTS system, lactose/cellobiose-specific IIB subunit	NA	-1,4	-4,3	-2,4
Cbei_2500	PTS system Galactitol-specific IIC component	NA	-3,3	-3,5	-2,6
Cbei_2665	PTS system, lactose/cellobiose family IIC subunit	NA	-2,9	-2,2	-2,6
Cbei_2666	PTS system, lactose/cellobiose-specific IIB subunit	NA	NA	-3,1	NA
Cbei_2707	PTS system, lactose/cellobiose family IIC subunit	NA	-1,2	-2,6	-3,7
Cbei_2739	Putative sugar-specific permease, SgaT/UlaA	NA	-1,9	-3,3	-4,8
Cbei_3073	D-galactose-binding periplasmic protein precursor	NA	1,2	-3,5	-1,4
Cbei_3142	D-galactose-binding periplasmic protein precursor	NA	1,4	-3,6	-2,7
Cbei_3295	ABC-type sugar transport system, periplasmic component	NA	-0,8	-3,3	-2,6
Cbei_3350	Citrate transporter	NA	-1,2	-3,8	n.d
Cbei_3367	Putative sugar-specific permease, SgaT/UlaA	NA	-1,3	-3,3	-2,6
Cbei_3410	Protein of unknown function DUF6, transmembrane	NA	-1,3	-6,8	-6,1
Cbei_3651	Major facilitator superfamily MFS_1	NA	-1,8	4,9	5,1
Cbei_3813	PTS system, lactose/cellobiose family IIC subunit	NA	-1,8	1,1	-2,5
Cbei_3873	PTS system, mannose/fructose/sorbose family, IIB subunit	NA	-3,4	-1,1	-2,4
Cbei_3888	Major facilitator superfamily MFS_1	NA	-4,1	-3,6	-3,4
Cbei_3939	D-galactose-binding periplasmic protein precursor	NA	-0,5	-3,7	-1,9
Cbei_4427	Putative galactoside ABC transporter	NA	0,4	-2,7	-1,6
Cbei_4432	Monosaccharide-transporting ATPase	NA	-2,7	-2,1	-1,3
Cbei_4511	MIP family channel protein, putative glycerol uptake facilitator protein	GlpF	4,5	0,8	-1,1
Cbei_4532	PTS system, N-acetylglucosamine-specific IIBC subunit	NA	-2,1	-3,6	-3,1
Cbei_4533	PTS system, glucose subfamily, IIA subunit	NA	-0,9	-5,6	-3,7
Cbei_4537	PTS system PTS, lactose/cellobiose-specific IIA subunit	NA	3,4	0,3	1,4
Cbei_4558	PTS system PTS, sorbose-specific IIC subunit	NA	-2,0	-3,3	-2,7
Cbei_4634	PTS system, lactose/cellobiose family IIC subunit	NA	-1,3	-2,6	-4,6
Cbei_4635	PTS system, lactose/cellobiose-specific IIB subunit	NA	0,5	-2,0	-4,0
Cbei_4651	Major facilitator superfamily MFS_1	NA	-0,3	-4,4	-4,9
Cbei_4705	PTS system, alpha-glucoside-specific IIBC subunit	NA	-0,1	3,3	1,6
Cbei_4982	PTS system, glucose subfamily, IIA subunit	NA	1,2	1,9	2,8
Cbei_5012	PTS system, sucrose-specific IIBC subunit	NA	-2,1	-4,3	-2,7

4

Table S4: Most differentially expressed genes involved in quorum-sensing, chemotaxis and secondary metabolism, with |log2 fold| > 2,5 and p.adj<0,05 at least once in Δ*spoIIE*C2 cultures relative to Δ*spoIIE*C5 cultures. If |log$_2$ fold| <1,5 or p.adj>0,05, the gene is not considered differentially expressed and the log$_2$ fold change is in italic.

Gene ID	Predicted protein function	log$_2$ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
TCS quorum systems				
Cbei_0192	two component transcriptional regulator, winged helix family	1,2	3,3	2,1
Cbei_0534	two component transcriptional regulator, winged helix family	0,3	3,7	2,7
Cbei_0589	two component transcriptional regulator, winged helix family	0,5	-2,9	-0,8
Cbei_0590	integral membrane sensor signal transduction histidine kinase	0,8	-2,8	-0,4
Cbei_1125	two component transcriptional regulator, winged helix family	-0,3	5,4	2,7
Cbei_1126	multi-sensor signal transduction histidine kinase	0,7	4,9	3,0
Cbei_1480	two component transcriptional regulator, LytTR family	-0,8	-2,8	-1,6
Cbei_2117	two component transcriptional regulator, winged helix family	0,2	-3,0	-1,5
Cbei_2252	two component transcriptional regulator, winged helix family	1,3	-1,9	3,3
Cbei_2400	two component transcriptional regulator, winged helix family	0,6	2,7	1,7
Cbei_2845	two component transcriptional regulator, winged helix family	-0,7	-3,0	-1,8
Cbei_2951	two component transcriptional regulator, winged helix family	1,2	2,9	-1,3
Cbei_2953	methyl-accepting chemotaxis sensory transducer	1,8	3,4	0,5
Cbei_3321	integral membrane sensor signal transduction histidine kinase	1,4		2,9
Cbei_3322	two component transcriptional regulator, winged helix family	-0,1	3,2	1,8
Cbei_3358	integral membrane sensor signal transduction histidine kinase	-0,6	2,6	2,1
Cbei_3359	two component transcriptional regulator, winged helix family	0,4	4,0	3,2
Cbei_3565	two component transcriptional regulator, winged helix family	-0,2	2,9	2,1
Cbei_3566	UBA/Ts-N domain-containing protein	1,5	1,4	2,8
Cbei_3667	Hypothetical protein, BtrM1 homolog	1,2	-3,5	-2,0
Cbei_3669	Two component transcriptional regulator, winged helix family, BtrR	0,1	-3,4	-1,0
Cbei_3785	two component transcriptional regulator, LuxR family	0,5	0,3	2,8
Cbei_3816	two component, sigma54 specific, transcriptional regulator, Fis family	-0,4	4,7	1,7
Cbei_3866	integral membrane sensor signal transduction histidine kinase	-0,3	-2,9	-3,3
Cbei_3867	two component transcriptional regulator, LuxR family	-0,5	-3,1	1,0
Cbei_3966	signal transduction histidine kinase regulating citrate/malate metabolism	2,3	3,9	5,5
Cbei_3967	two component transcriptional regulator, LytTR family	0,8	3,2	6,0
Cbei_4578	Accessory gene regulator B	0,3	1,0	3,0
Cbei_5025	integral membrane sensor signal transduction histidine kinase	-1,1	4,6	-0,1
Cbei_5026	two component transcriptional regulator, winged helix family	-2,0	3,8	-0,6
Chemotaxis				
Cbei_3954	putative CheW protein	0,1	-5,8	-2,8

4

Gene ID	Predicted protein function	log$_2$ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_4183	CheA signal transduction histidine kinase	1,3	4,6	6
Cbei_4184	putative CheW protein	2,5	6,7	9
Cbei_4304	putative CheW protein	1,2	2,7	2,7
Secondary metabolite cluster				
Cbei_0233	4-alpha-glucanotransferase	-0,7	-2,2	-5,2
Cbei_0236	Efflux transporter, RND family, MFP subunit	-0,2	-4,6	-1,9
Cbei_0237	ABC transporter related	0,6	-3,7	-1,1
Cbei_0238	Protein of unknown function DUF214	-1,1	-5,5	-2,6
Cbei_0243	Transcriptional antiterminator, BglG	1,0	3,0	2,6
Cbei_0244	Phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 2	1,0	3,9	2,2
Cbei_0245	Mannitol dehydrogenase, C-terminal domain	0,2	2,8	1,4
Cbei_0247	serine O-acetyltransferase	-1,2	-3,6	-1,2
Cbei_0249	conserved hypothetical protein	-1,7	-1,4	2,5
Cbei_0259	transcriptional regulator, TetR family	-1,4	3,9	2,3
Cbei_0260	drug resistance transporter, EmrB/QacA subfamily	-5,5	2,3	-1,1
Cbei_0261	efflux transporter, RND family, MFP subunit	-3,0	3,7	0,5
Cbei_0262	ABC transporter related	-3,6	3,3	0,4
Cbei_0264	hypothetical protein	-2,8	2,2	-0,7
Cbei_0265	hypothetical protein	-3,9	0,2	0,3
Cbei_0267	two component transcriptional regulator, LytTR family	1,7	3,9	4,3
Cbei_0268	signal transduction histidine kinase regulating citrate/malate metabolism	1,8	2,8	1,8
Cbei_0269	Accessory gene regulator B homolog	0,6	-4,6	1,7
Cbei_0272	dipeptidyl aminopeptidase/acylaminoacyl-peptidase related protein	0,1	3,3	1,6
Cbei_0274	major facilitator superfamily MFS_1	-2,8	1,4	0,4
Cbei_0275	iron dependent repressor	-1,7	4,3	1,1
Cbei_0277	cyclase family protein	-0,1	5,2	0,5
Cbei_0278	transcriptional regulator, MarR family	3,2	9,8	8,5
Cbei_0279	methyl-accepting chemotaxis sensory transducer	0,7	7,2	2,8
Cbei_0280	hypothetical protein	-8,5	-6,0	-4,1
Cbei_0281	hypothetical protein	-3,7	-4,6	0,5
Cbei_0282	hypothetical protein	-3,7	-6,0	-0,3
Cbei_0283	ABC transporter related	-5,0	-5,1	-1,8

4

Table S5: Most differentially expressed genes involved in cell wall and membrane formation, with |log2 fold| > 2,5 and p.adj<0,05 at least once in Δ*spoIIE*C2 cultures relative to Δ*spoIIE*C5 cultures. If |$\log_2$ fold| <1,5 or p.adj>0,05, the gene is not considered differentially expressed and the $\log_2$ fold change is in italic.

Gene ID	Predicted protein function	$\log_2$ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_0032	glutamate racemase	1,1	3,7	3,1
Cbei_0057	3D domain protein	-1,3	-6,6	-4,5
Cbei_0089	polysaccharide biosynthesis protein	-0,1	2,6	4,6
Cbei_0107	UDP-N-acetylmuramoylalanine--D-glutamate ligase	-0,2	3,1	2,0
Cbei_0236	efflux transporter, RND family, MFP subunit	-0,2	-4,6	-1,9
Cbei_0261	efflux transporter, RND family, MFP subunit	-3,0	3,7	0,5
Cbei_0318	Mannose-1-phosphate guanylyltransferase (GDP)	0,6	1,0	-3,9
Cbei_0368	polysaccharide deacetylase	-0,8	3,9	3,0
Cbei_0386	glycosyl transferase, group 1	-1,2	-2,5	-4,0
Cbei_0400	Monogalactosyldiacylglycerol synthase	0,0	-0,6	2,8
Cbei_0487	NLP/P60 protein	-0,9	-5,3	-3,0
Cbei_0493	rod shape-determining protein MreD	2,4	0,0	2,0
Cbei_0519	Serine-type D-Ala-D-Ala carboxypeptidase	-1,0	-4,0	-1,9
Cbei_0591	glycoside hydrolase, family 25	-1,4	-4,3	-2,6
Cbei_0941	Serine-type D-Ala-D-Ala carboxypeptidase	-1,0	-3,1	-1,8
Cbei_0968	polysaccharide biosynthesis protein	5,4	-1,8	-2,0
Cbei_1024	beta-lactamase domain protein	-0,2	-2,5	-1,0
Cbei_1030	conserved hypothetical protein	1,1	-4,0	-2,4
Cbei_1114	penicillin-binding protein, transpeptidase	1,4	1,3	3,0
Cbei_1242	NLP/P60 protein	-0,2	3,7	1,7
Cbei_1308	UDP-N-acetylmuramyl-tripeptide synthetase	3,5	10,0	9,9
Cbei_1388	polysaccharide biosynthesis protein	-1,1	-2,9	-1,9
Cbei_1400	glycosyl transferase, family 4	1,3	3,3	2,9
Cbei_1493	cell wall hydrolase/autolysin	-0,8	-2,8	-1,4
Cbei_1501	Glycosyl transferase	-0,6	3,5	2,7
Cbei_1724	glycosyl transferase, family 2	1,2	6,9	3,4
Cbei_1725	glycosyl transferase, family 39	0,4	5,0	2,0
Cbei_1807	secretion protein HlyD family protein	1,1	2,9	1,2
Cbei_2057	protein of unknown function DUF1304	-0,5	5,6	5,9
Cbei_2261	Lytic transglycosylase, catalytic	-2,9	-6,1	-5,3
Cbei_2434	NLP/P60 protein	-1,9	-4,0	-3,9
Cbei_2476	efflux transporter, RND family, MFP subunit	3,1	-0,9	-1,0
Cbei_2509	3D domain protein	0,8	-3,6	-1,6
Cbei_2585	glycosyl transferase, group 1	0,1	-2,6	-4,9

4

Gene ID	Predicted protein function	log$_2$ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_2589	glycosyl transferase, family 2	-1,1	-1,1	-4,3
Cbei_2601	dTDP-4-dehydrorhamnose 3,5-epimerase	-0,8	-2,0	-2,7
Cbei_2662	cell wall hydrolase/autolysin	-0,7	-2,2	-2,6
Cbei_2825	beta and gamma crystallin	-4,7	-4,9	-4,3
Cbei_2831	Carbohydrate-binding family V/XII	-4,3	-6,4	-5,9
Cbei_2880	NAD-dependent epimerase/dehydratase	-2,3	-2,3	-2,6
Cbei_2923	secretion protein HlyD family protein	-0,6	3,4	-1,0
Cbei_2956	uncharacterized protein, ErfK family contains peptidoglycan-binding domain	-0,7	-3,5	-2,7
Cbei_3004	prolipoprotein diacylglyceryl transferase	1,5	2,7	2,3
Cbei_3148	type IV pilus assembly PilZ	2,0	2,7	0,7
Cbei_3267	glycoside hydrolase, family 25	-0,7	-1,8	-2,5
Cbei_3638	Lytic transglycosylase, catalytic	-2,6	-6,8	-5,9
Cbei_3725	Domain of unknown function DUF1906	-4,6	-5,7	-6,7
Cbei_3885	secretion protein HlyD family protein	4,9	5,9	3,6
Cbei_3964	CHAP domain containing protein	-5,9	-6,7	-4,6
Cbei_4023	cell wall hydrolase/autolysin	-1,0	-2,7	-2,7
Cbei_4121	Cyclopropane-fatty-acyl-phospholipid synthase	1,2	3,4	1,7
Cbei_4236	glycosyl transferase, family 2	2,5	2,0	2,7
Cbei_4277	glycosyltransferase	1,5	2,8	3,7
Cbei_4278	N-acylneuraminate-9-phosphate synthase	0,5	2,6	2,6
Cbei_4280	GCN5-related N-acetyltransferase	1,4	2,1	3,6
Cbei_4330	D-alanyl carrier protein	0,7	2,3	2,7
Cbei_4331	membrane bound O-acyl transferase, MBOAT family protein	0,6	1,5	2,6
Cbei_4339	Nucleotidyl transferase	3,4	5,7	1,9
Cbei_4556	sugar isomerase (SIS)	-2,4	-2,7	-3,1
Cbei_4618	UTP-glucose-1-phosphate uridylyltransferase	0,8	-1,3	-1,7
Cbei_4620	sugar transferase	2,1	0,4	0,1
Cbei_4632	glycosyl transferase, family 2	-0,9	-2,5	-0,3
Cbei_4664	polysaccharide deacetylase	0,5	2,9	2,2
Cbei_4665	glycoside hydrolase, family 25	0,0	2,5	1,8
Cbei_4666	Lytic transglycosylase, catalytic	-0,1	2,5	1,7
Cbei_4667	sulfatase	-0,1	2,5	1,3
Cbei_4694	Nucleotidyl transferase	2,9	2,6	0,6
Cbei_4695	Choline/ethanolamine kinase	2,5	2,2	0,9
Cbei_4699	cell wall hydrolase/autolysin	-0,8	3,5	2,0

4

Gene ID	Predicted protein function	log$_2$ fold change in expression		
		4 h	11 h/ 11 h	11 h/ 23 h
Cbei_4717	putative cell wall binding repeat-containing protein	0,0	-3,9	-2,9
Cbei_4718	putative cell wall binding repeat-containing protein	1,6	1,5	3,4
Cbei_4739	Undecaprenyl-phosphate galactose phosphotransferase	1,0	3,2	1,9
Cbei_4742	lipopolysaccharide biosynthesis protein	-0,4	-3,1	n.d
Cbei_4753	glycosyl transferase, family 2	1,4	2,1	2,7
Cbei_4755	dTDP-4-dehydrorhamnose 3,5-epimerase	-0,4	2,8	1,4
Cbei_4757	Undecaprenyl-phosphate galactose phosphotransferase	0,7	-2,6	-1,6
Cbei_4758	glycosyl transferase, family 2	-1,0	-3,3	-2,6
Cbei_4874	uncharacterized secreted protein	-1,8	-5,6	-2,4
Cbei_5006	LrgB family protein	1,2	3,1	0,5
Cbei_5088	MscS Mechanosensitive ion channel	1,4	3,4	3,0

4

Table S 6: Most differentially expressed genes involved in early sporulation regulation and stress response, with |log2 fold| > 2,5 and p.adj<0,05 at least once in Δ*spoIIE*C2 cultures relative to Δ*spoIIE*C5 cultures. If |$\log_2$ fold| <1,5 or p.adj>0,05, the gene is not considered differentially expressed and the $\log_2$ fold change is in italic. NA: No protein assigned.

Gene ID	Predicted protein function	Putative encoded protein	$\log_2$ fold change in expression		
			4 h	11 h/ 11 h	11 h/ 23 h
Early sporulation					
Cbei_1712	response regulator receiver protein	Spo0A	-0,6	-2,9	-2,3
Cbei_3169	multi-sensor signal transduction histidine kinase	NA	-3,2	-5,5	-3,8
Cbei_0088	transcriptional regulator, AbrB family	NA	0,75	4,4	4,9
Cbei_3375	transcriptional regulator, AbrB family	NA	-2,5	-0,3	-0,5
Cbei_4885	transcriptional regulator, AbrB family	NA	0,2	0,6	2,8
Cbei_0813	anti-sigma regulatory factor, serine/threonine protein kinase	SpoIIAB	-2,9	-3,5	-1,9
Cbei_0814	RNA polymerase sigma factor SigF	σF	-1,6	-3,8	-1,7
Cbei_1119	peptidase U4, sporulation factor SpoIIGA	SpoIIGA	-1,2	-5,4	-1,3
Cbei_1120	RNA polymerase sigma factor SigE	σE	-3,5	-6,7	-2,8
σF regulon					
Cbei_0112	hypothetical protein	CfsB	3,9	7,9	7,5
Cbei_0823	Stage II sporulation P family protein	SpoIIP	-1,1	-2,3	-2,6
Cbei_1121	RNA polymerase sigma factor SigG	σG	-3,0	-6,0	-3,6
σE regulon					
Cbei_0567	Sporulation protein YhbH	YhbH	-0,9	-3,2	-3,5
Cbei_1122	Sporulation protein YlmC/YmxH	YlmC	-2,2	-6,2	-3,4
Cbei_1136	Sporulation stage IV, protein A	SpoIVA	-3,3	-5,4	-5,5
Cbei_1210	Sporulation protein YlmC/YmxH	YlmC	2,3	0,2	3,7
Cbei_1389	Sporulation stage V, protein B	SpoVB	-1,6	-3,8	-2,7
Cbei_1568	Sporulation protein YtxC	YtxC	1,9	1,4	3,9
Cbei_1579	Stage V sporulation protein D	SpoVD	-0,5	-2,7	-1,2
Cbei_1692	Sporulation stage III, protein AA	SpoIIIAA	-0,4	-3,6	-4,7
Cbei_1693	Sporulation stage III, protein AB	SpoIIIAB	-1,2	-3,8	-2,5
Cbei_1694	Stage III sporulation AC family protein	SpoIIIAC	-0,5	-4,6	-3,8
Cbei_1695	Sporulation stage III, protein AD	SpoIIAD	-0,4	-3,8	-2,4
Cbei_1696	Sporulation stage III, protein AE	SpoIIIAE	-1,0	-4,6	-4,5
Cbei_1697	Sporulation stage III, protein AF	SpoIIIAF	0,0	-2,9	-1,7
Cbei_1962	Sporulation stage II, protein M	SpoIIM	0,1	-5,1	-2,2
Cbei_4229	Sporulation protein YunB	YnB	-0,5	-3,7	-2,7
Stress response					
Cbei_0123	ATPase AAA-2 domain protein	ClpC	-0,5	6,2	2,0
Cbei_0328	Chaperonin Cpn10	Cpn10	0,1	8,2	1,9
Cbei_0569	Rubrerythrin	NA	-0,9	-3,3	-2,9

4

Gene ID	Predicted protein function	Putative encoded protein	log$_2$ fold change in expression		
			4 h	11 h/ 11 h	11 h/ 23 h
Cbei_0645	ATPase AAA-2 domain protein	ClpC	-0,7	9,1	2,8
Cbei_0828	Heat-inducible transcription repressor HrcA	HrcA	-0,4	4,8	-0,3
Cbei_0829	GrpE protein	GrpE	1,3	7,9	2,2
Cbei_0830	Chaperone protein DnaK	DnaK	0,0	5,9	0,0
Cbei_0831	Chaperone protein DnaJ	DnaJ	0,0	5,5	0,3
Cbei_1254	ATP-dependent protease La	LonA	0,1	8,7	1,5
Cbei_1856	Superoxide dismutase	SodA	1,3	2,3	2,7
Cbei_2325	Rubrerythrin	NA	-4,6	-1,9	-1,7
Cbei_3069	Rubrerythrin	NA	-2,5	-2,1	-2,7
Cbei_4123	Heat shock protein Hsp20	Hsp20	0,6	9,3	2,1

4

Chapter 5-

L-Rhamnose metabolism in *Clostridium beijerinckii* DSM 6423

Mamou Diallo[1,2*], Andre D. Simons[1*]; Hetty van der Wal[1], Florent Collas[1], Bwee Houweling-Tan[1], Servé W. M. Kengen[2], Ana M. López-Contreras.[2 #]

[1] Wageningen Food and Biobased Research, Wageningen, The Netherlands
[2] Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands
* Both authors contributed equally

Abstract

Macroalgae (or seaweeds) are considered potential biomass feedstocks for the production of renewable fuels and chemicals. Their sugar composition is different from that of lignocellulosic biomasses, and in green species, including *Ulva lactuca*, the major sugars are L-rhamnose and D-glucose. *C. beijerinckii* DSM 6423 utilized these sugars in an *U. lactuca* hydrolysate to produce acetic acid, butyric acid, isopropanol, butanol and ethanol (IBE) and 1,2-propanediol. D-glucose was almost completely consumed in diluted hydrolysates, while L-rhamnose or D-xylose were only partially utilized. In this study, the metabolism of L-rhamnose by *C. beijerinckii* DSM 6423 was investigated to improve its utilization from natural resources. Fermentations on D-glucose, L-rhamnose and a mixture of D-glucose and L-rhamnose were performed. On L-rhamnose, the cultures showed low growth and sugar consumption and produced 1,2-propanediol, propionic acid and n-propanol in addition to acetic and butyric acids, whereas on D-glucose, IBE was the primary product. On a D-glucose/L-rhamnose mixture, both sugars were converted simultaneously and L-rhamnose consumption was higher, leading to high levels of 1,2-propanediol (78.4 mM), in addition to 59.4 mM of butanol and 31.9 mM isopropanol. Genome and transcriptomics analysis of D-glucose- and L-rhamnose-grown cells revealed the presence and transcription of genes involved in L-rhamnose utilization, but also in bacterial microcompartment (BMC) formation. These data provide useful insights into the metabolic pathways involved in L-rhamnose utilization and the effects on the general metabolism (glycolysis, early sporulation, stress-response) induced by growth on L-rhamnose.

Importance

A pre-requisite for a thriving biobased economy is the efficient conversion of biomass resources into useful products, such as biofuels, bulk- and specialty chemicals. In contrast to other industrial microorganisms, natural solvent-producing *Clostridia* utilize a wide range of sugars, including C5, C6 and deoxy-sugars, for the production of long-chain alcohols (butanol, 2,3-butanediol), isopropanol, acetone, n-propanol and organic acids. Butanol production by *Clostridia* from first-generation sugars is already a commercial process, but for the expansion and diversification of the A/IBE process to other substrates, more knowledge is needed on the regulation and physiology of fermentation of sugar mixtures. Green macroalgae, produced in aquaculture systems, harvested from the sea or from tides, can be processed into hydrolysates containing mixtures of D-glucose and L-rhamnose, which can be fermented. The knowledge generated in this study will contribute to the development of more efficient processes for macroalgae fermentation and of mixed sugar fermentation in general.

Introduction

The increasing worldwide demand for fuels and chemicals contradicts the diminishing availability of fossil resources, which are currently the main source of these compounds. In past decades, the concept of biorefinery has been established as an alternative to petroleum-based refineries, in which multiple products (energy, fuels and (high-value) chemicals) are produced from one biomass source (295, 296). Nowadays, the most established biorefineries are based on lignocellulosic biomasses. However, diversification of biomass resources is needed to ensure sufficient availability and flexibility of processes. Macroalgae have gained attention in recent years as feedstock for the production of fuels and chemicals due to the advantages that they show with respect to traditional terrestrial feedstocks for biorefinery: i) higher productivity (biomass produced per unit of surface) than terrestrial crops, ii) no competition for arable land, iii) lower freshwater consumption during cultivation, and iv) no requirement for fertilizers (297). In addition, macroalgae show a distinctive chemical composition compared to lignocelluloses and terrestrial crops, as some species are rich in carbohydrates, proteins, fatty acids and/or bioactive components that make them very suitable for biorefinery as sources of multiple valuable products (298, 299). Current developments in sustainable large-scale cultivation of macroalgae could result in increased availability of these biomasses at economic conditions (300).

In the green seaweed, *Ulva lactuca*, D-glucose and L-rhamnose are the main carbohydrates present in the ulvan polysaccharide, and it has been reported that these sugars could be extracted using mild pre-treatment conditions (230, 301). In fermentations performed with *Clostridium acetobutylicum* and *Clostridium beijerinckii,* using *U. lactuca* hydrolysates as substrate, the solvents acetone, butanol, and ethanol (ABE) were produced. Interestingly, *C. beijerinckii* was also able to produce 1,2-propanediol when grown on L-rhamnose, but not on D-glucose (230). In contrast, *C. acetobutylicum* did not show any production of 1,2-propanediol, and was unable to grow solely on L-rhamnose. *C. beijerinckii* DSM 6423 (also NRRL B-593, formerly known as *C. butylicum* NRRL B-593), was able to grow on L-rhamnose as a sole carbon source, producing 1,2-propanediol, but also propanol and propionate, in addition to acetic- and butyric acids (302). L-Rhamnose utilization is well-studied for some microorganisms, such as *Escherichia coli* or *Salmonella typhimurium* (303, 304), and has been described for *C. phytofermentans* (271). In the latter species, L-rhamnose was shown to be converted along a phosphorylated pathway, involving rhamnulose, rhamnulose-P and lactaldehyde as key intermediates. Lactaldehyde is the precursor of the main end-product 1,2-propanediol (271, 305). 1,2 Propanediol is an interesting chemical. Its production has been studied in different micro-organisms, including fungi, bacteria and yeasts (306, 307). The involvement of bacterial microcompartments (BMC) in the catabolism of 1,2-propanediol into n-propanol and propionic acid has been described in *C. phytofermentans* and other organisms as an exciting

5

feature (271, 307). L-Rhamnose metabolism by solventogenic *Clostridia*, however, is not well characterized. Production of 1,2-propanediol, propionate and propanol by *C. beijerinckii* suggests that in this solventogenic species, L-rhamnose is converted by a metabolic route similar to that reported for *C. phytofermentans*.

The aim of this study is to investigate the L-rhamnose metabolism in the solventogenic strain *C. beijerinckii* DSM 6423. Growth and product-formation on D-glucose or L-rhamnose were compared. The pathway for L-rhamnose conversion was reconstructed through analysis of the recently sequenced genome of this strain (102). The gene transcription profile in cultures grown on D-glucose and on L-rhamnose as sole carbon sources were determined using RNA sequencing and the differences observed were analyzed with respect to sugar metabolism, early sporulation and stress response. The results obtained contribute to enhance our knowledge about the unique capability of solventogenic *Clostridia* to ferment a variety of carbohydrates into a wide spectrum of products with commercial interest.

Results

Fermentation of *Ulva lactuca* hydrolysate by *C. beijerinckii*

The potential of *C. beijerinckii* for utilization of *U. lactuca* hydrolysate, containing D-glucose and L-rhamnose as major sugars, was investigated using a hydrolysate prepared according to Bikker et al. (308). Cultures on hydrolysate and on control media containing D-glucose, L-rhamnose or a mixture of sugars as main carbon- and energy sources were grown in serum flasks. The hydrolysate was rich in D-glucose (115 mM) and L-rhamnose (86 mM), and in addition, contained 28 mM of D-xylose. Very low growth was observed in cultures of pure hydrolysate or hydrolysate supplemented with nutrients as in CM2 medium. Only a small amount of the D-glucose in the hydrolysate was consumed after 144h of incubation at 37°C, leaving the D-xylose and the L-rhamnose unused in the medium. When the hydrolysates, both pure or supplemented with nutrients, were diluted 1:1 with demineralized water, their fermentability improved. In the cultures on the diluted hydrolysates (DH), the D-glucose was consumed almost completely, whereas approximately 50% of the D-xylose and approximately 20% of the L-rhamnose were consumed. The major products of these last fermentations were butyrate, acetate, isopropanol and butanol. Low levels of 1,2-propanediol were detected in the cultures on the diluted hydrolysate (DH) cultures (Table S1).

The major fermentation products in the L-rhamnose control cultures were butyrate and 1,2-propanediol, whereas an IBE fermentation took place on glucose cultures (Table S1). On all the cultures tested, L-rhamnose was only partially utilized. In the cultures with mixtures

of D-glucose and L-rhamnose, D-glucose was completely utilized, and the consumption of L-rhamnose was higher compared to that observed on L-rhamnose-only cultures (Table S1).

The yields of IB(E) produced per D-glucose or D-glucose/D-xylose consumed in the different cultures are shown in Table S1. In this table, it can be observed that in the non-diluted hydrolysate cultures, the yields obtained are higher than the yields of IBE produced from D-glucose in the control cultures (0.94 and 0.80 in H-*Ulva* and H-*Ulva*+N respectively, vs. 0.72 in D-glucose control cultures). This is also the case for DH-Ulva+N cultures, where the yield of IB produced is 0.82 mM of IB per mM of D-glucose and D-xylose consumed. This indicates that in the hydrolysates, most probably other carbon sources, such as oligo- or di-saccharides, are present that can be utilized to produce solvents by *C. beijerinckii*.

Fermentation of L-rhamnose and L-rhamnose/D-glucose mixture

To better characterize the fermentation of L-rhamnose and to obtain cell material for RNA-seq analysis, cultures were grown in bioreactors with 400-mL of working volume without pH control. Samples were taken at different time points for the determination of metabolites and for RNA-seq analysis. Fermentations on D-glucose and D-glucose/L-rhamnose mixtures were performed as reference. The results are shown in Figure 1 and Table 1. *C. beijerinckii* was able to grow on L-rhamnose as a carbon- and energy source, albeit growth and L-rhamnose consumption were lower than in D-glucose-grown cultures, with OD600 values of approximately 3 and 11 for L-rhamnose and D-glucose, respectively (Figure 1). Growth on L-rhamnose ceased as soon as the pH of the culture dropped below 5, and in contrast to the case of D-glucose-grown cultures, the pH did not increase anymore. To check if the growth stopped due to the low pH, a second fermentation on L-rhamnose only was performed, in which the pH was controlled to, or above, 5.2. The growth profile, L-rhamnose consumption, and product formation were not significantly different from those of the non-pH-controlled cultures (results not shown), indicating that the low pH was not the only cause of growth cessation. Interestingly, L-rhamnose fermentation does not lead to the generation of typical solvents like acetone, butanol and ethanol. Also, re-assimilation of acids apparently does not occur, as 19.2 mM of acetate and 11.7 mM of butyrate were produced in the L-rhamnose culture (Figure 1, Table 1).

Instead of producing IBE, as was seen in D-glucose-grown cultures, the strain produced 1,2-propanediol, propanol, and propionate in addition to acetate and butyrate, when L-rhamnose was provided as the carbon source. Propionate and propanol are known to be typical products of the catabolism of 1,2-propanediol in many microorganisms, including clostridial species (271).

For biomass determination from the L-rhamnose and on the L-rhamnose/D-glucose cultures, dry matter content was measured at the end of the fermentations. The calibration curve that relates biomass to OD_{600} values of the cultures obtained in D-glucose cultures was not applicable for L-rhamnose-grown cultures, as these show a very different cell morphology (Figure S1). The highest yields were found for cultures grown on D-glucose or the mixture, with yields of 0.12 and 0.059 mol biomass/mol of sugar consumed, respectively.

As L-rhamnose and D-glucose are both present in hydrolysates from green seaweeds, their co-metabolism was studied in cultures grown on a mixture of these sugars in a ratio 1:1. In Figure 1 and Table 1, it can be seen that in these cultures, all D-glucose was consumed and that the consumption of L-rhamnose was significantly higher compared to the rhamnose only cultures (86.7 mM and 46.4 mM on D-glucose/L-rhamnose cultures and on L-rhamnose, respectively). Interestingly, both sugars in the medium were consumed simultaneously, although glucose was consumed at a higher rate (Figure 1). L-rhamnose was only partially consumed, as observed in the L-rhamnose-only cultures, remaining approximately 31% of the initial content in the medium.

The primary fermentation products on the D-glucose/L-rhamnose mixture corresponded to those observed for the D-glucose- and L-rhamnose-only fermentations, IBE and 1,2-propanediol, respectively. The concentration of 1,2-propanediol reached 78 mM, approximately four times higher than that was seen in the L-rhamnose-only cultures, as a result of higher sugar consumption.

As mentioned above, in the L-rhamnose-only cultures, low concentrations of propionic acid and n-propanol were detected in the medium (Table 1, Figure 1). Remarkably, in cultures grown on mixtures of D-glucose and L-rhamnose, no propanol or propionate were detected (Table 1).
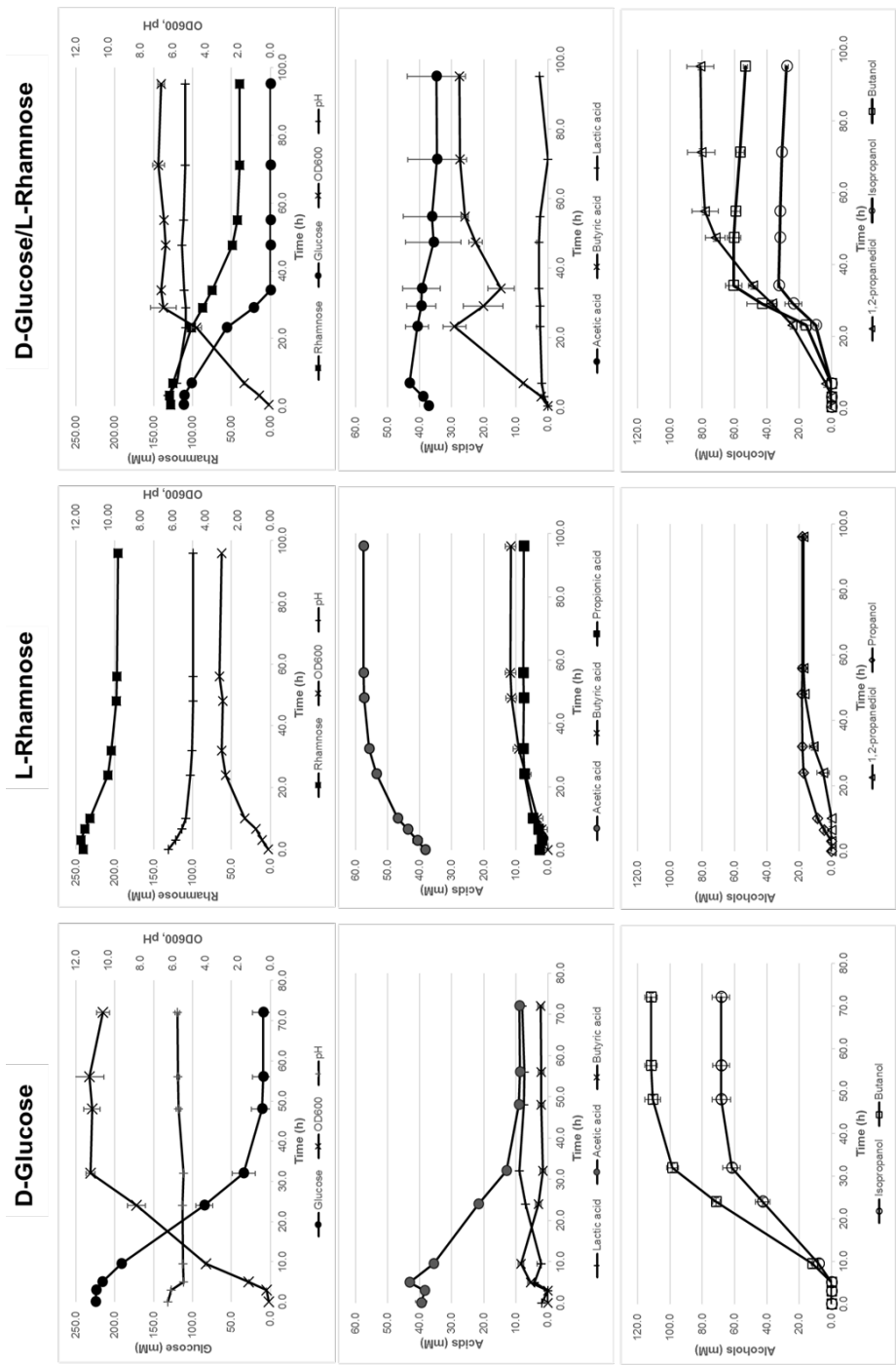
5

Figure 1: Fermentation profiles of *C. beijerinckii* grown on D-glucose (left), L-rhamnose (center) and D-glucose/L-rhamnose mixture (right). Fermentations were performed in duplicate in CM2 medium supplemented with the indicated sugars. Only products present at a concentration >1mM are shown. Standard deviation is shown with error bars, indicating one standard deviation of the mean (n=2).

5

Table 1: Fermentation data of cultures of *C. beijerinckii* grown on D-glucose, L-rhamnose and on a D-glucose/L-rhamnose mixture.

	D-Glucose (t=56h)	L-Rhamnose (t=56h)	D-Glucose/ L-Rhamnose (t=72h)
Substrates at t = 0 h (mM)			
D-Glucose	224.8		111.2
L-Rhamnose		243.4	129.3
Acetate	39.5	38.3	37.3
Substrates consumed at t = end (mM)			
D-Glucose	215.3	n.d	111.2
L-Rhamnose		46.4	86.7
Acetate*	30.7	0	1.2
Products at t = end (mM)			
Acetate*	8.8	57.5 (*19.2**)	36.2
Lactate	7.5	n.d	2.4
Butyrate	2.1	11.7	25.9
Acetone	3.2	n.d	7.9
Isopropanol	68.5	n.d	31.9
Ethanol	6	n.d	1.1
Butanol	111.8	n.d	59.4
1,2-Propanediol	n.d	17.2	78.4
Propanol	n.d	18.2	n.d
Propionate	n.d	7.8	n.d
Biomass	26.6	2.1	11.7
Yields and recovery			
Biomass (mol/mol sugar)	0.12	0.045	0.059
Yield Butanol (mol/mol D-glucose)	0.52		0.53
Yield 1,2-PD (mol/mol L-rhamnose)		0.37	0.904
Yield propanol (mol/mol L-rhamnose)		0.39	
Yield propionate (mol/mol L-rhamnose)		0.17	
Carbon recovery** (%)	95	96	88
Electron recovery (%)	98	96	88

n.d: not detected, * Acetate was produced in L-rhamnose cultures, ** The release of CO_2 was estimated and included in the calculations

L-Rhamnose pathway reconstruction

To further investigate the pathways of L-rhamnose metabolism in *C. beijerinckii*, bioinformatic analyses were performed on the genome sequence of the strain, as recently published (102, 309). Since the L-rhamnose catabolism was recently described for *C. phytofermentans* ISDg, this strain was used as the main source of genes for the query for BLASTp searches (310), but data on other organisms were used as well (307), as shown in Table 2. Genes encoding enzymes involved in all steps of the transport and metabolism of L-rhamnose into 1,2-propanediol could be identified (Table 2), with similarities ranging

from 51 % to 83 %. As in *C. phytofermentans*, most genes involved in the L-rhamnose catabolism were clustered within a genomic region (Figure 2). For two of the enzymes encoded in the cluster, the rhamnulose-phosphate aldolase enzyme (CIBE_0615) and the 1,2-propanediol oxidoreductase (CIBE_0617), gene duplications with high similarity were present (CIBE_3969 and CIBE_2890, respectively) elsewhere in the genome.

 Proteins putatively involved in L-rhamnose transport into the cell did not show similarity to those of *C. phytofermentans* but were most similar to those of an ABC transporter found in the soil bacterium *R. leguminosarum* (311) and a transporter of the Major Facilitator Superfamily (MFS) of the plant pathogen *Dickeya dadantii*.

A cluster of 21 genes contains the genes for the further conversion of 1,2-propanediol to propionic acid and *n*-propanol (Figure 2). This cluster is almost identical to the clusters found in other organisms, including *C. phytofermentans* (305). 1,2-Propanediol is expected to be converted to propionaldehyde by a propanediol dehydratase. Unlike *S. typhimurium*, but similar to *C.phytofermentans*, this is likely catalyzed by a B_{12}-independent type of dehydratase (CIBE_4900; PduCDE). Propionaldehyde is further converted to propanol or propionyl-CoA, catalyzed by a propanol dehydrogenase and a propionaldehyde dehydrogenase, respectively. A homolog for the propanol dehydrogenase is present (CIBE_4892), belonging to the Zn-dependent dehydrogenases. For the propionaldehyde dehydrogenase, two homologs are present in the cluster (CIBE_4884 and 4893), which is unlike *C. phytofermentans*, which only contains one. Propionyl-CoA is converted to propionate involving a phosphate propionyl transferase and a propionate kinase. The cluster contains a gene encoding for the transferase (CIBE_4886) but not for the kinase. In *C.phytofermentans*, the propionate dephosphorylating is catalyzed by a kinase that is not specific to proniate-P, an acetate kinase encoded by a gene outside the L-rhamnose cluster. In *C.beijerinckii,* the bioinformatics analysis does not provide enough proof to identify the gene encoding the propionate kinase.

5

The BMC cluster identified in our *C. beijerinckii* belongs to the Glycyl Radical Enzyme-Containing Microcompartment type (GRM) like the one described in *C. phytofermentans* (271). GRMs are found mainly in *Firmicutes* and some Deltaproteobacteria and *Olsenella* (312). This BMC locus type contains the metabolosome core enzymes and a glycyl radical enzyme, which is the pyruvate lyase (CIBE_4900) in *C. beijerinckii*.

Bioinformatic studies (312, 313) showed that the GRM can be divided into subgroups depending on the BMC shell proteins and the accessory genes that belong to the locus. The BMC cluster from *C. beijerinckii* belongs to the GRM.3 Group because it contains genes that encode a peptidase, a flavoprotein, a EutJ homolog, S-adenosylmethionine synthetase and signaling proteins. They are suspected to be involved in several metabolic pathways, such as vitamin B_{12} or S-adenosylmethionine synthesis.
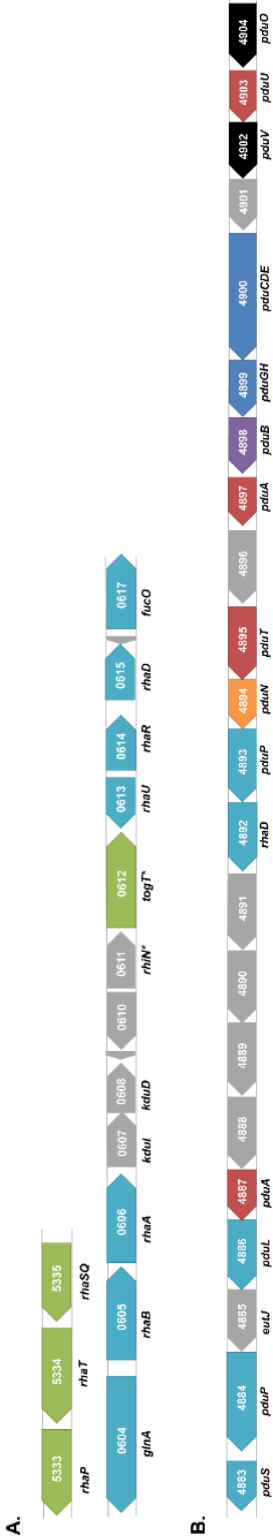
Figure 2: Schemes of the rhamnose utilization (A) and Bacterial Microcompartments (BMC) (B) clusters in *C. beijerinckii* DSM 6423. *2A)* The genes predicted to encode enzymes involved in the L-rhamnose metabolism are shown in blue, the genes in green encode for putative L-rhamnose transporters and the genes in grey are not reported to be involved in L-rhamnose metabolism. The functional homologue in *Rhizobium leguminosarum bv. Trifolii* is indicated for each gene, * Homologs involved in rhamno–galacturonan catabolism were identified in *Dickeya dadantii* 3937. *2B)* BMC superlocus, as compared to GRM3 (312). The functional equivalent in *S. typhimurium* of the genes predicted to be involved in the BMC are indicated below each gene. The genes are shown in different colors according to the function of the protein encoded: in blue, enzymes involved in the conversion of propanediol into propionate and propanol, in red, genes predicted to encode for BMC-H shell proteins, in purple for BMC-T shell proteins, in orange for BMC-P shell proteins and the genes in grey have an unknown function.

Table 2: Relative expression values from RNA-seq of *C. beijerinckii* DSM 6423 genes putatively involved in L-rhamnose uptake and conversion.

C. beijerinckii proteins	Proposed protein function	Closest homolog with experimental evidence [a]		Log$_2$ fold change in expression relative to D-glucose cultures		
		Organism	Protein (% similarity)	3 h	6.5 h	10 h
L-Rhamnose transport into the cell						
CIBE_5333	rhamnose ABC transporter, permease subunit	*Rhizobium leguminosarum bv. trifolii*	RhaP (51)	n.d	4.2	4.0
CIBE_5334	rhamnose ABC transporter, ATPase subunit	*R. leguminosarum bv. trifolii*	RhaT (62)	1.3	4.5	4.1
CIBE_5335	rhamnose ABC transporter, periplasmic solute binding subunit	-	-	3.1	5.9	4.5
CIBE_0612	MFS rhamnose cation symporter	*Dickeya dadantii (Erwinia chrysanthemi)*	TogT (63 %)	8.8	8.4	6.9
L-Rhamnose conversion to 1,2-propanediol						
CIBE_0605	rhamnulokinase	*C. phytofermentans*	Cphy_1146 (73)	7.8	7.8	7.1
CIBE_0606	L-rhamnose isomerase	*C. phytofermentans*	Cphy_1147 (78)	8.5	8.6	7.4
CIBE_0613	L-rhamnose mutarotase	*C. phytofermentans*	Cphy_1149 (83)	n.d	1.4	1.3
CIBE_0614	transcriptional regulator	*C. phytofermentans*	Cphy_1187 (66)	1.8	1.5	1.6
CIBE_0615	rhamnulose-1-phosphate aldolase	*Escherichia coli (strain K12)*	RhaD (69)	7.0	7.4	6.8
CIBE_0617	L-1,2-propanediol oxidoreductase	*E. coli (strain K12)*	FucO (79)	8.4	7.7	6.7

[a] All percent similarity values were determined using global alignments of protein sequences using the gapped BLAST algorithm (310). References for the characterized functional equivalent are as follows: *Rhizobium leguminosarum bv. trifolii* rhamnose transporters (307), *C. phytofermentans* rhamnose dissimilation enzymes (271), *Escherichia coli* (strain K12) rhamnulose-1-phosphate aldolase and L-1,2-propanediol oxidoreductase (314, 315), rhamnose MFS transporter (307, 316, 317)

5

Transcriptome analysis

Samples for mRNA isolation were taken from the fermentations on D-glucose and on L-rhamnose, as shown in Figure 1. Time points for sampling were chosen in such a way that the early exponential, acidogenic, and solventogenic growth phases were represented. For D-glucose-grown cultures, samples were taken after 3.0, 5.0, and 9.5 hours. For L-rhamnose-grown cultures, samples were taken after 3.0, 6.5, and 10 hours. After RNA isolation and sequencing, the data were mapped against the recently sequenced genome of this *C. beijerinckii* strain to quantify gene expression levels under each condition (102, 309, 318).

In summary, mapping of the RNA-seq reads against the published genome of *C. beijerinckii* resulted in reliable reads in a range of 93.17 % to 98.60 % (Table S3). For analyzing differentially expressed genes, the TAMARA tool on the MicroScope platform was used (29). Thresholds were selected at |log2 (fold change)| > 3 and adjusted p-value< 0.005, which resulted in a list of 671 significantly differentially expressed genes on L-rhamnose (11% of the genome). To see the impact of the L-rhamnose metabolism on selected functional clusters (Table 2, 3, S4, S5, S6) in all three time points, the log2 (fold change) was decreased to 0.5. Only fifty-nine of these significantly differentially expressed genes have a log_2-fold change above 3.0 at the three time points and 115 genes at two time points. Out of these fifty-nine genes, 25 correspond to the genes involved in L-rhamnose uptake and conversion and BMC formation. Moreover, except for the rhamnose mutarotase, the corresponding genes of the putative L-rhamnose degrading enzymes were among the highest expressed during growth on rhamnose compared to glucose (Table 2). They were upregulated between 6.84 and 8.58 times on a log_2-fold scale, depending on the enzyme and the sampling time point. For L-rhamnose transport, two different transport systems were identified, an ABC and an MFS transporter. The transcript data show that both systems are indeed upregulated during growth on L-rhamnose. Especially, the transporter of the MFS-type was highly upregulated (8.8-fold). As mentioned, for the rhamnulose-phosphate aldolase, two putative genes (*CIBE_0615* and *CIBE_3969*) were present. Of these, only *CIBE_0615* was highly expressed on L-rhamnose, suggesting that *CIBE_0615* encodes for the rhamnulose-phosphate aldolase, which cleaves rhamnulose-phosphate to lactaldehyde and DHAP. Likewise, two putative genes were proposed to encode for the 1,2-propanediol oxidoreductase, which converts lactaldehyde to 1,2-propanediol. However, only *CIBE_0617* was highly expressed on L-rhamnose, indicating that this gene encodes for the functional protein. Also, most putative genes for the conversion of 1,2-propanediol to propanol and propionate were highly expressed, confirming their role in L-rhamnose fermentation. For the propionaldehyde dehydrogenase, two genes were identified in the BMC-cluster, and both were highly upregulated. As for the propionate kinase, since no gene encoding for this enzyme is present in the BMC-cluster, we looked at the differential

expression of genes encoding for kinases outside of the BMC-cluster. We noticed the up-regulated expression of a putative butyrate kinase (2 fold on a log$_2$ scale), encoded by a gene present elsewhere in the genome (CIBE_5515). We can assume that this putative butyrate kinase acted as a propionate kinase *C. beijerinckii* was grown on L-rhamnose.

All the genes from the BMC locus were among the highest expressed when *C. beijerinckii* was grown on L-rhamnose (Table 3). A shallow expression of this cluster was detected in glucose-grown cultures from 4 to 55 reads per gene compared to 234 to 81,830 reads in L-rhamnose-grown cultures.

As the L-rhamnose cultures produced acetate and butyrate, but no IBE, the expression of the main genes associated with glycolysis, acidogenesis and solvent production were also analyzed. Most of the genes of the central metabolism to solvents were found to be less expressed in the L-rhamnose cultures (Figure 3, Table S5). Genes predicted to code for enzymes involved in acidogenesis by Máté de Gerando et al. (102) were also less expressed in L-rhamnose cultures, suggesting that the reactions for acetate or butyrate production from L-rhamnose might be catalyzed by different enzymes. Genes involved in solvent formation (butanol and ethanol), however, were slightly upregulated during the early exponential growth phase of the L-rhamnose cultures. In Table S5, a list of differentially expressed genes encoding for enzymes or other proteins predicted to be involved in glycolysis, acidogenesis and solventogenesis is shown, including the fold-change in expression between L-rhamnose and D-glucose.

The gene predicted to encode Spo0A, the global regulator of the metabolism in solventogenic clostridia (CIBE_2041), did not show a significant difference in expression levels when grown on D-glucose or on L-rhamnose. This indicates that stationary phase processes, including sporulation and stress-response mechanisms, might not be differently regulated under both conditions. In Table S6, a list of predicted proteins related to sporulation and to stress response and their fold-change in expression during growth on L-rhamnose compared to D-glucose is shown. Most genes encoding sporulation-related proteins or enzymes did not show important expression changes. However, data on the gene expression at stationary phase, after 20 hours of fermentation, is needed to have a better insight on differences in the regulation of stress-response mechanism between D-glucose or on L-rhamnose grown cultures.
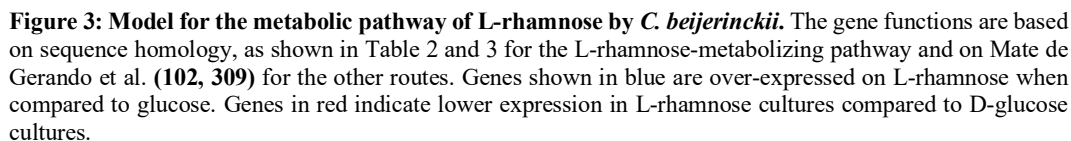
5

Figure 3: Model for the metabolic pathway of L-rhamnose by *C. beijerinckii*. The gene functions are based on sequence homology, as shown in Table 2 and 3 for the L-rhamnose-metabolizing pathway and on Mate de Gerando et al. **(102, 309)** for the other routes. Genes shown in blue are over-expressed on L-rhamnose when compared to glucose. Genes in red indicate lower expression in L-rhamnose cultures compared to D-glucose cultures.

Table 3: Composition and differential expression of the BMC locus in *C. beijerinckii* DSM 6423.

C. beijerinckii protein	Proposed protein function	Closest homolog with experimental evidence [a]		Log₂ fold change in expression relative to D-glucose cultures		
		Organism	Protein (% similarity)	3 h	6.5 h	10 h
CIBE_4883	propanediol oxidoreductase	*Salmonella typhimurium*	PduS (68)	5.8	5.4	3.8
CIBE_4884	propionaldehyde dehydrogenase	*C. phytofermentans*	Cphy_1178 (70)	9.5	8.3	6.2
CIBE_4885	EutJ, putative chaperonin, ethanolamine utilization protein	*S. typhimurium*	EutJ (63)	8.6	7.6	6.6
CIBE_4886	phosphate propanoyl transferase	*C. phytofermentans*	Cphy_1183 (67)	9.1	7.9	6.1
CIBE_4887	BMC-H shell protein	*C. phytofermentans*	Cphy_1182 (88)	8.7	7.9	5.6
CIBE_4888	conserved membrane protein of unknown function	-	-	8.8	7.8	5.9
CIBE_4889	MetK, S-adenosylmethionine synthetase	-	-	9.1	8.1	6.6
CIBE_4890	response regulator receiver protein	-	-	9.2	8.3	7.2
CIBE_4891	signal transduction histidine kinase, lyts	-	-	9.4	8.5	6.9
CIBE_4892	propanol dehydrogenase	*Klebsiella pneumoniae*	Dhat (43)	10.0	10.0	7.6
CIBE_4893	propionaldehyde dehydrogenase	*C. phytofermentans*	Cphy_1178 (63)	9.5	9.6	7.3
CIBE_4894	BMC-P shell protein	*C. phytofermentans*	Cphy_1184 (71)	9.5	9.5	7.5
CIBE_4895	BMC-H shell protein	*C. phytofermentans*	Cphy_1186 (61)	10.0	10.0	7.8
CIBE_4896	conserved protein of unknown function	-	-	11.0	10.0	8.2
CIBE_4897	BMC-H shell protein	*C. phytofermentans*	Cphy_1182 (89)	11.0	10.0	7.5
CIBE_4898	BMC-T shell protein	*S. typhimurium*	PduB (67)	10.0	9.0	7.0
CIBE_4899	propanediol dehydratase activator	*Clostridium butyricum*	DhaB2 (56)	11.0	11.0	7.7
CIBE_4900	propanediol dehydratase	*C. butyricum*	DhaB1 (58)	12.0	10.0	6.9
CIBE_4901	glutamine amidotransferase	-	-	6.3	5.9	5.4
CIBE_4902	propanediol utilization protein	*S. typhimurium*	PduV (62)	6.6	6.3	5.6

5

C. beijerinckii protein	Proposed protein function	Closest homolog with experimental evidence [a]		Log$_2$ fold change in expression relative to D-glucose cultures		
		Organism	Protein (% similarity)	3 h	6.5 h	10 h
CIBE_4903	BMC-H shell protein	*C. phytofermentans*	Cphy_1176 (76)	7.7	6.7	5.7
CIBE_4904	Propanediol utilization protein	*S. typhimurium*	PduO (62)	7.1	6.1	5.0
CIBE _4905	hypothetical protein	-	-	2.2	n.d[b]	n.d
CIBE _4906	xanthine/uracil/vitamin C permease	-	-	-2.3	-3.3	1.7

[a] All percent similarity values were determined using global alignments of protein sequences using the gapped BLAST algorithm [b] n.d for not detected. References for the characterized functional equivalent are as follows: *S. typhimurium* (304), *C. phytofermentans* (271), *K.pneumoniae* (202, 319), *C. butyricum* (320), for the BMC shell protein types (321).

Discussion

Next to lignocellulosic biomasses, aquatic biomasses such as seaweeds are a promising source for various industries (299, 308). It has been shown that clostridial species are able to grow on hydrolysates from the green seaweed *U. lactuca* and that the main products are acetone, butanol, and ethanol (ABE), which can be used in the biofuel industry (230). In addition, it was observed that 1,2-propanediol was produced as a result of L-rhamnose utilization. The metabolism of L-rhamnose was studied in more detail in various organisms, including *E. coli* and *S. typhimurium* (304, 322). In addition, Forsberg et al. (302) showed already in the 1980s that several clostridial species were able to ferment L-rhamnose, and a fermentation pathway homologous to the one in *E. coli* and *S. typhimurium* was suggested (302). Recently, the production of propionate and propanol from L-rhamnose was demonstrated in *C. phythofermentans,* and a fermentation model was proposed, which also included a specific organelle, the BMC (271). In the frame of this work, the L-rhamnose metabolism in the solventogenic *C. beijerinckii* DSM 6423 was studied by genome analysis, fermentation studies and transcriptomics.

C. beijerinckii DSM 6423 was tested for growth on *U. lactuca* hydrolysate, in a similar approach as earlier described by our laboratory for the strain *C. beijerinckii* NCIMB 8052 (308). In contrast to the latter strain, *C. beijerinckii* DSM 6423 did not grow well on the pure hydrolysate. When the hydrolysate was diluted, the growth and sugar consumption improved, indicating that this strain could be inhibited by components of the hydrolysate. The content of elements that take part in salts, as potential inhibitors of growth, in the hydrolysate was estimated based on the data from Bikker and co-workers (308). The calculated content of elements the *U. lactuca* hydrolysate is shown in Table S2, being S the most abundant with a concentration of 304 mM. The effect of salts on the growth of solventogenic *Clostridia* is not well characterized yet, and only few reports can be found

on this topic. Ezeji et al. showed that levels of S corresponding to 93 mM in the form of sodium sulfate resulted in inhibitory effects in the growth of *C. beijerinckii* on control media (44). The supplementation of microalgae-derived hydrolysates with sodium chloride (NaCl) at 342 mM and higher resulted in inhibition of growth in *C. pasteurianum* (323). In a different study, it was shown that the removal of S, among other elements, from a wood hydrolysate increased fermentability by *C. beijerinckii* (324). The strain used in this study showed higher sensitivity to inhibitors in the hydrolysate as compared to strain NCIMB 8052, although they are both genetically very similar. Interestingly, the hydrolysate was rich enough in nutrients, and supplementation with nutrients was not required for growth.

On control medium, *C. beijerinckii* DSM 6423 was capable of growth on L-rhamnose as the sole source of carbon and energy. L-rhamnose was converted into acetate, butyrate and the typical L-rhamnose-derived products 1,2-propanediol, propanol and propionate. Remarkably, typical solvents like isopropanol, butanol and ethanol were not produced. Possibly, solvent production may not be necessary during L-rhamnose fermentation, as all reducing equivalents are required for the production of 1,2-propanediol and propanol (Figure S2). Indeed DHAP (46.4 mM) conversion to acetyl-CoA leads to the formation of 46.4 mM of NADH and 46.4 mM of reduced ferredoxin. Reduced ferredoxin cannot directly donate electrons for solvent production, but should first transfer its electrons to NAD. Assuming that all ferredoxin is converted to NADH, this would yield 92.8 mM NADH in total. This NADH is then used for the production of 46.4 mM 1,2 propanediol and 10,4 mM propanol (7.8 mM NADH is derived from propionaldehyde conversion) in the lactaldehyde-branch, and the production of 15.5 mM hydroxybutyryl-CoA and 15.5 mM butyryl-CoA in the DHAP branch (Figure S2; Figure 3). Thus, this leaves no reducing equivalents for solvent formation, which agrees with the absence of IBE production during L-rhamnose fermentation. Inside the BMC, propionaldehyde is either oxidized or reduced to propanol and propionyl-CoA, respectively. If both products were produced in equimolar amounts, no net NADH would be produced/consumed. However, the non-equal production of propanol (18.2 mM) and propionate (7.8 mM) indicates that some NADH must come from outside the BMC, suggesting that reductant (NADH) is able to pass the BMC shell. It has been proposed before that NAD(H) is able to cross the BMC via specific pores (325).

From the metabolic pathway shown in Figure 3, based on Mate de Gerando et al. (102), it can be estimated that from 1 mol of L-rhamnose; 1 mol of 1,2-propanediol and 1 mol of DHAP should be formed. The 1,2-propanediol is then further metabolized into n-propanol and propionic acid. Table 1 shows that from 46.4 mM of L-rhamnose consumed, 43.2 mM of total products derived from 1,2-propanediol are formed, *viz.* 17.2 mM 1,2 propanediol, 18.2 mM of n-propanol and 7.8 mM of propionic acid. DHAP is converted along the EMP pathway to pyruvate, which is further metabolized into acetic acid (19.2 mM) and butyric acid (11.7 mM). These numbers are in agreement with the expected stoichiometry of this

5

part of the pathway, as 19.2 mM acetate and 11.7 mM butyrate are derived from 42,6 mM DHAP (19.2 mM + 2 x 11.7 mM). However, some carbon that is not included in these calculations should end up in biomass. Thus, from 46.4 mM of L-rhamnose (-2.1 mM of biomass), 44.3 mM of lactaldehyde and DHAP are produced. Based on the fermentation data, the following equations can be composed for both branches:

1 lactaldehyde → 0.39 1,2-propanediol + 0.41 propanol + 0.18 propionate

1 DHAP → 0.43 acetate + 0.26 butyrate (derived from 0.52 acetyl-CoA)

In accordance, the carbon- and electron recovery both reached 96% for the L-rhamnose cultures. For the D-glucose and the D-glucose/L-rhamnose mixture, the recoveries were also high, with 95% and 88%, respectively.

Growth on L-rhamnose, however, stops before all L-rhamnose is converted. The reason for this is not apparent. Possibly, too much acids are produced, which may become toxic. Commonly, during growth on D-glucose, solvents are produced to prevent excessive production of (undissociated) weak acids. However, running the fermentation under pH-controlled conditions did not improve the L-rhamnose conversion. During co-fermentation of L-rhamnose and D-glucose, substantially more L-rhamnose is fermented, suggesting that there might be an energetic reason for the growth retardation on pure L-rhamnose. Theoretically, solventogenic D-glucose fermentation yields ~2 moles of ATP per mole of sugar (326), which is more than twice the amount that can be obtained on L-rhamnose (0.9 mole ATP/mole sugar), assuming that L-rhamnose uptake requires 1 ATP/mole sugar (Figure S2). The OD_{600} data show that growth is best on D-glucose (OD_{600} =11.1), followed by the sugar mixture (OD_{600}=6.8) and the L-rhamnose culture (OD_{600}=2.8). The lower growth yield on L-rhamnose correlates with the calculated ATP yield, which is approximately 38% of the yield on D-glucose (Table 1, Figure S2). Moreover, L-rhamnose induces explicitly the formation of the BMC (*vide infra*), whose protein shell may impose an extra biosynthetic energetic burden for the cell. The lower biomass yield on L-rhamnose compared to on D-glucose was observed earlier by Forsberg et al. (302), but its origin was not further studied then. However, despite this apparent difference in growth yields, the reason for the premature growth stop on L-rhamnose remains obscure.

The genome analysis revealed the presence of all necessary genes that are specifically needed for the anticipated enzymes of the L-rhamnose pathway. These include genes for L-rhamnose uptake and subsequent conversion to 1,2-propanediol, propanol and propionate, and from which many are clustered and probably organized in several operons. Also, various genes coding for shell proteins of the BMC were identified (Table 3). Sequence analysis of the different operons shows that the organization of the genes involved in L-rhamnose metabolism is similar to what was found in *C. phytofermentans*. Indeed, the genes responsible for L-rhamnose uptake and conversion are located in a

different region than the BMC cluster. However, we observed significant differences in the genes involved in rhamnose transport and the size of the BMC cluster. In our strain, a gene coding for an L-rhamnose-specific ABC-type transporter was present in the genome, but another L-rhamnose-specific transporter gene belonging to the L-rhamnose conversion cluster (CIBE_0612, Table 2) was highly upregulated during growth on L-rhamnose. Thus, the latter transporter is most likely responsible for L-rhamnose uptake. This transporter belongs to the Major Facilitator Superfamily type (MFS), which uses an H+ gradient to transport the sugar, described in *Rhizobium leguminosarum bv. trifolii*. In solventogenic clostridia, this type of transporter has not been studied in detail yet.

The BMC cluster identified in *C. beijerinckii* DSM 6423 shows some differences compared to the one described in *C. phytofermentans* ISDg. It harbors 21 genes organized in nine operons in one locus, whereas a recent study shows that the *C. phytofermentans* genome harbors three BMC clusters, but only one was experimentally studied (271). The BMC gene cluster found in *C. beijerinckii* DSM 6423 is more related to the cluster found in other Clostridia such as *C. saccharolyticum* K10 or *C. ljungdahlii* DSM 13528 and α-proteobacteria, such as *Rhodobacter capsulatus* SB 1003. Homologs of the L-rhamnose utilization clusters found in *C. beijerinckii* DSM 6423 were also found in the genome of *C. beijerinckii* strain NCIMB 8052, which utilizes L-rhamnose as well (308).

Transcriptome analysis confirmed the involvement of the predicted genes in L-rhamnose conversion. Most metabolic proteins, except the propionate kinase, were highly upregulated (up to eight-fold). In fact, no specific propionate kinase gene could be identified by bioinformatic analysis. However, the up-regulation (2 fold) of one of the copies of the butyrate kinase gene CIBE_5515 in L-rhamnose-grown cells suggests that this gene may have activity towards propionyl-P. The various BMC-shell proteins were also highly upregulated (9-12 fold compared to D-glucose-grown cells). Thus, the BMC is specifically induced during growth on L-rhamnose, as has also been described for L-rhamnose conversion in *C. phytofermentans* (271). On the L-rhamnose/D-glucose mixture, we observed production of 1,2-propanediol but not of propanol and propionate. This suggests that D-glucose prevents induction of the BMC even when L-rhamnose is present.

It is assumed that the polyhedral shell prevents leakage of volatile metabolites or that it protects the cell against toxic intermediates; in this case, propionaldehyde (327) or radicals of the 1,2-propanediol dehydratase reaction (271). As mentioned above, synthesis of the protein–shell may exert a heavy burden on the protein-synthesis machinery and may, therefore, also affect the growth rate and energetics of the cell.

In this study, we show that *C. beijerinckii* is able to ferment L-rhamnose as a sole carbon- and energy source to produce acetic and butyric acids, 1,2-propanediol, propionic acid and n-propanol, which are products of commercial interest. The metabolism of L-rhamnose in

5

this strain shows similarities to pathways described in other Clostridia, but also presents exciting novelties, such as the presence of an MFS transporter for L-rhamnose. Co-fermentation of L-rhamnose with D-glucose leads to higher L-rhamnose utilization, with shows potential for the use of this strain for fermentation of *U. lactuca* hydrolysates, or other L-rhamnose-containing streams, provided that salt toxicity can be reduced. The results in this study serve as a basis for further developments towards efficient biomass utilization for the production of chemicals.

Materials and Methods

Bacterial strains and culture conditions

C. beijerinckii DSM 6423 was stored at -20 °C as spore suspension in 20% glycerol. The spore suspension was heat-shocked for 1 min at 95 °C before inoculation. Fermentations were performed in CM2 medium containing (in $g.L^{-1}$): yeast extract, 1.00; KH_2PO_4, 1.00; K_2HPO_4, 0.61; $MgSO_4 \times 7\ H_2O$, 1.00; $FeSO_4 \times 7\ H_2O$, 0.0066; *para*-aminobenzoic acid, 0.10; and ammonium acetate, 2.90. Stock solutions of D-glucose and L-rhamnose were autoclaved separately and added after autoclaving of the medium to a final concentration of 40 $g.L^{-1}$. All liquid media were made anaerobic by flushing with nitrogen gas. Fermentations in 400-mL working volume were performed in Infors HT Multifors bioreactors at 37 °C and a stirrer speed of 150 rpm. Bacterial growth was monitored by measuring the optical density at 600 nm (OD_{600}).

Product analysis

Fermentation substrates and products were measured by HPLC. Glucose, rhamnose, acetate, butyrate, lactate, acetone, ethanol, butanol, propanol, and isopropanol were measured in a Waters HPLC system equipped with a refractive index detector (Waters model 2414) and a Shodex KC-811 300 × 8 mm column at 80 °C with 3 mM H_2SO_4 as mobile phase and a flow rate of 1.00 $mL.min^{-1}$. As internal standard, 3 mM valeric acid in 1 M H_2SO_4 was used. Propionate and 1,2-propanediol were measured in a Dionex UltiMate3000 HPLC system equipped with a refractive index detector (Waters model 2414) and a Biorad Aminex HPX 87H 300 × 8 mm column at 30 °C with 3.7 mM H_3PO_4 as mobile phase and a flow rate of 0.60 $mL.min^{-1}$. As internal standard, 2.5 mM phthalic acid in water was used.

Carbon recovery

For the calculation of the carbon recovery, the total number of moles of carbon present in the products and biomass was divided by the total number of C-moles of the substrates. Since acetate was present at the start and end of the fermentation, it was considered as substrate and product. For simplicity, acetate was considered as substrate when its final

concentration was lower than at the start of the fermentation, whereas it was considered as product when its final concentration was higher than at the start. For D-glucose-grown cultures, the cell dry weight (cdw) was calculated from the optical density at 600 nm using the following formula:

$$cdw\ [g\ L^{-1}] = OD600 * 0.28 + 0.13$$

For the cultures grown on L-rhamnose and on D-glucose/L-rhamnose mix, the cell dry weight was determined by filtration of 10 mL of culture on a 0.22 μm porous filter, drying the biomass on the filter in an oven at 50°C overnight and weighing.

The carbon content of biomass was calculated using the standard elemental biomass formula ($CH_{1.8}O_{0.5}N_{0.2}$) given by Von Stockar et al. (328). CO_2 production during fermentation was taken into account. It was assumed that for the production of one mole of acetate or ethanol, butyrate or butanol, and acetone or isopropanol, one, two, and three moles of CO_2 are produced, respectively.

Electron recovery

The electron recovery was determined by calculating the degree of reduction per mole of all compounds produced divided by all compounds produced. For simplicity, acetate was again considered as substrate when its final concentration was lower than at the start of the fermentation, whereas it was considered as product when its final concentration was higher than at the start. Since the production of H_2 could not be accurately quantified during the fermentation, it was calculated from the stoichiometry of the reactions. The degree of reduction per mole of substrate or product are (in brackets): D-glucose (24), L-rhamnose (26), acetate (8), butyrate (20), lactate (12), acetone (16), isopropanol (18), butanol (24), ethanol (12), 1,2-propanediol (16), propanol (18), propionate (14), and H_2 (2) The degree of reduction of biomass was calculated from the standard elemental biomass composition of $CH_{1.8}O_{0.5}N_{0.2}$, which corresponds to a degree of reduction of biomass of 21 electrons per mole.

RNA sequencing

Total RNA was isolated from *C. beijerinckii* DSM 6423 for transcriptome studies. Samples were taken from duplicate 400-mL fermentations from cells in early exponential, acetogenic, and solventogenic phases. Cells were pelleted for 15 min at 3000 × g and 4 °C and stored at -80 °C until further use. RNA was isolated using TRIzol reagent (Thermo Fisher Scientific) and the PureLink RNA Mini Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. In short, the cell pellet was thawed on ice and resuspended in 5 mL TRIzol reagent for cell lysis. Next, 1 mL chloroform was added, and after centrifugation for 15 min at 13,000 × g, the upper aqueous phase was mixed with an equal

volume of 70 % ethanol. The solution was loaded on a spin cartridge, washed once, and treated with 30 U DNase I. After two additional washing steps, the RNA was eluted in RNase-free water. Quality and quantity of the isolated RNA were checked by gel electrophoresis and NanoDrop, respectively. Afterward, the samples were stored at -80 °C before being sent for sequencing. Library construction and sequencing were performed by Novogene Co. Ltd. Messenger-RNA was depleted with the Ribo-Zero Magnetic Kit, and a 250-300 bp insert cDNA library was constructed. Pair-ended 150 bp fragments were sequenced using the Illumina HiSeq platform. After sequencing, data were uploaded and analyzed with the MicroScope platform (29). Reads were mapped against the recently sequenced *C. beijerinckii* DSM 6423 genome (102, 309).

Nucleotide sequences

The DSM 6423 full genome sequence is available on the European Nucleotide Archive (ENA) under the accession number PRJEB11626 (309). Link: https://www.ebi.ac.uk/ena/data/view/PRJEB11626

The DSM 6423 RNA-seq data described by Mate de Gerando et al. (102) were deposited on the NCBI BioProject Database under the Accession Number GSE100024.Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100024 (318).

The DSM 6423 RNA-seq data described in this study have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7487.
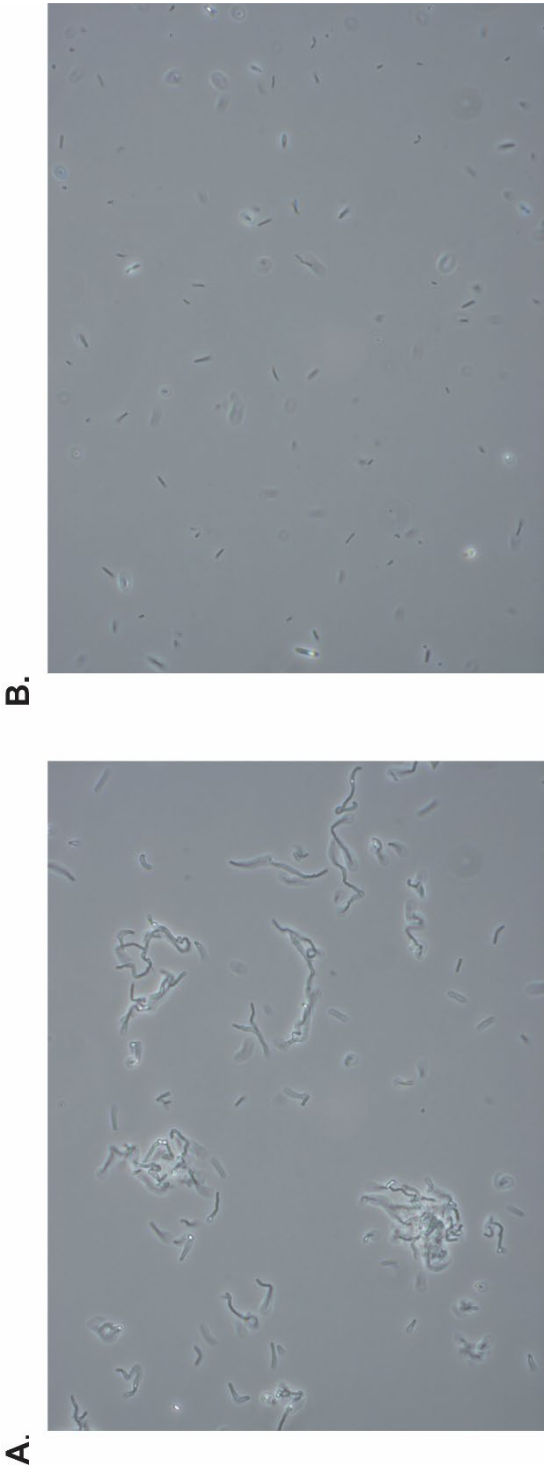
Acknowledgment

Supplementary files

Figure S1: Phase-contrast photographs (1000x) of cultures grown on a glucose/rhamnose mix (A) and on rhamnose (B) after 30 h. of fermentation.
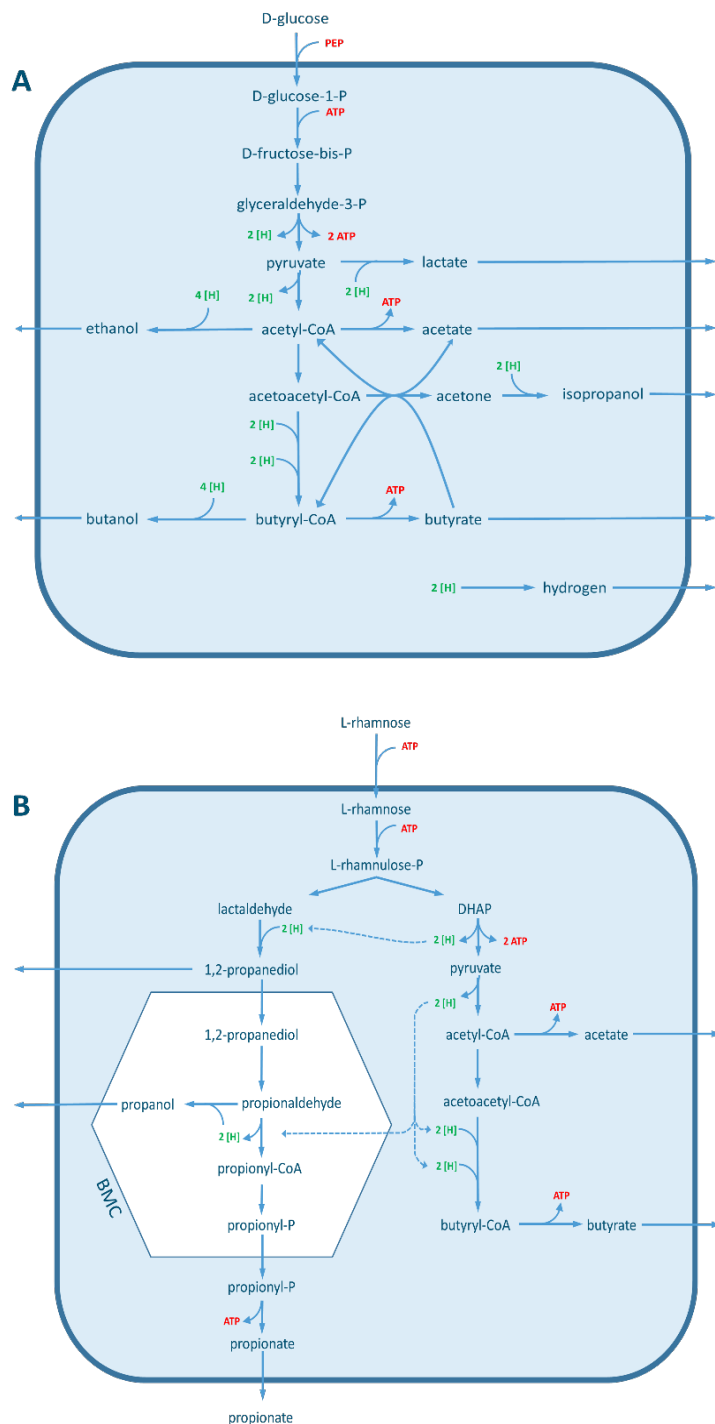
5

Figure S2. Redox balance and energy metabolism in *C. beijerinckii*. A. growth on D-glucose and B. growth on L-rhamnose.

Table S1: Fermentation of *Ulva lactuca* hydrolysates by *C. beijerinckii*. As controls, cultures on D-glucose, L-rhamnose, and a mix of D-glucose, L-rhamnose and D-xylose (G/R/X) were carried out. Codes: H-Ulva, hydrolysate of *Ulva lactuca* prepared as previously described ; H-Ulva + N, indicates H-Ulva supplemented with nutrients as in CM2 medium; DH-Ulva indicates H-Ulva diluted 1:1 with demiwater; DH-Ulva+ N indicates H-Ulva+N diluted 1:1 with demiwater. In brackets, the times at which the last sample of the cultures was taken.

	Glucose (t=144h)	Rhamnose (t=144 h)	G/R/X (t=167h)	H-Ulva (t=144h)	H-*Ulva* +N (t=144h)	DH-*Ulva* (t=167h)	DH-*Ulva*+ N (t=167h)
Substrates at t=0 (mM)							
D-Glucose	209.2		141.0	114.7	102.9	57.3	51.4
L-Rhamnose		238.7	156.6	85.8	74.5	42.9	37.3
D-Xylose			36.6	28.2	24.9	14.1	12.5
Acetate	36.5	35.0	36.1	9.8	46.5	4.9	23.2
Substrates consumed (mM)							
D-Glucose	164.3		141.0	21.9	21.3	52.9	50.3
L-Rhamnose		42.9	43.3			8.2	5.0
D-Xylose			34.6			6.8	7.1
Acetate*	23.6		5.5				10.2
Products at the end (mM)							
Acetate*		17.7		27.1	60.9	17.5	
Butyrate	6.9	13.8	17.3	6.2	6.8	27.7	47.6
Isopropanol	31.9	n.d	23.0	7.5	7.0	6.2	10.5
Ethanol	2.6	n.d	6.3	n.d	n.d	n.d	n.d
Butanol	84.1	n.d	72.7	13.2	10.1	21.3	36.8
1,2-Propanediol	n.d	15.0	22.6	n.d	n.d	n.d	0.8
Propanol	n.d	19.0	n.d	n.d	n.d	n.d	n.d
Propionate	n.d	7.0	9.6	n.d	n.d	n.d	n.d
Yields							
mM IBE/mM (D-Glucose+D-xylose) consumed	0.72		0.58	0.94	0.80	0.46	0.82

*Acetate was present in the media as component, and in case it was utilized, it is considered as a substrate. When acetate is produced, it is considered as a product, and the end concentration in the cultures is indicated. "n.d" indicates not detected

5

Table S2: Salt content in the *U. lactuca* hydrolysate, calculated from the analysis data shown in Bikker et al. (308)**.**

Element	Concentration (mM)
K	73
Mg	185
Na	67
Cl	32
S	304

Table S3: Mapping overview of the performed RNA sequencing.

Sample	Time point [h]	Total number of reads	Number of unmapped reads	Number of reliable reads [%]
Glc_1	3	5,354,497	33,708	96.17
Glc_1	5	4,331,116	20,562	96.94
Glc_1	9.5	4,721,967	19,719	96.71
Glc_2	3	4,466,679	54,964	94.88
Glc_2	5	4,700,384	23,416	95.86
Glc_2	9.5	5,175,736	23,026	95.71
Rha_1	3	4,505,619	17,259	98.60
Rha_1	6.5	4,718,680	43,935	97.52
Rha_1	10	4,284,034	105,542	95.84
Rha_2	3	4,525,859	28,847	98.20
Rha_2	6.5	4,987,878	53,998	97.19
Rha_2	10	4,320,185	191,739	93.71

Table S4: Relative expression value of genes belonging to the L-rhamnose conversion cluster but involved in other pathways.

C. beijerinckii protein	Predicted function	Log_2 fold change in expression relative to D-glucose		
		3 h	6.5 h	10 h
chondroitin sulfate degradation / rhamnogalacturonan degradation				
CIBE_0607	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase, 5-keto-4-deoxyuronate isomerase	7.7	7.7	6.6
CIBE_0608	2-keto-3-deoxygluconate oxidoreductase	6.9	7.0	6.2
CIBE_0610	glycosyl hydrolase, family 88, rhiN	6.7	6.6	6.4
unknown function				
CIBE_0609	protein of unknown function	6.2	6.7	6.9
CIBE_0611	xylose isomerase domain protein TIM barrel	4.0	4.1	4.0

Table S5: Differential expression during growth on L-rhamnose relative to D-glucose of the genes involved in the central metabolism. The clusters and gene assignments were based on the transcriptome study of Mate de Gerando et al. (102).

C. beijerinckii protein	Predicted function	Log$_2$ fold change in expression relative to D-glucose cultures		
		3 h	6.5 h	10 h
Glycolysis				
CIBE_0769	glyceraldehyde-3-phosphate dehydrogenase	-1.2	-2.6	-1.5
CIBE_2315	fructose-1,6-bisphosphate aldolase	-1.1	-2.0	-1.2
CIBE_0770	phosphoglycerate kinase	n.d	-2.5	-1.6
CIBE_0414	glucose-6-phosphate isomerase	n.d	-1.4	-1.1
CIBE_0771	triose phosphate isomerase	n.d	-2.4	-1.3
CIBE_5856	6-phosphofructokinase	-2.8	-2.2	n.d
CIBE_0774	enolase	1.3	n.d	n.d
CIBE_0772	phosphoglycerate mutase	n.d	-1.1	n.d
CIBE_1193	6-phosphofructokinase 2	5.0	2.3	n.d
CIBE_0655	pyruvate kinase	n.d	-0.91	n.d
CIBE_0755	phosphofructokinase	-2.3	n.d	n.d
CIBE_3600	fructose-bisphosphate aldolase class 1	2.4	3.7	n.d
CIBE_5451	fructose-1,6-bisphosphatase	n.d	n.d	-0.82
CIBE_2174	phosphoglycerate mutase	n.d	n.d	-0.96
CIBE_5855	pyruvate kinase	-3.4	-2.6	n.d
CIBE_5461	putative fructose-bisphosphate aldolase	1.2	1.6	1.1
CIBE_5876	putative transketolase N-terminal section	-1.5	-2.0	-1.4
Pentose phosphate pathway				
CIBE_5875	putative transketolase C-terminal section	-1.3	-1.9	-1.1
CIBE_0561	ribose 5-phosphate epimerase	n.d	-1.1	-1.0
CIBE_1389	ribulose-5-phosphate 3-epimerase	n.d	-1.4	-0.89
CIBE_3260	transaldolase	2.8	n.d	n.d
CIBE_5350	putative transaldolase	1.9	n.d	n.d
CIBE_0719	ribose 5-phosphate epimerase	-1.5	n.d	1.8
CIBE_2821	putative transaldolase	2.0	1.3	-0.77
Acidogenesis				
CIBE_2950	histidine kinase	1.1	n.d	n.d
CIBE_5186	pyruvate-flavodoxin oxidoreductase	n.d	-2.7	-3.1
CIBE_0565	degradative acetoacetyl-CoA thiolase	-2.7	-3.3	-4.0
CIBE_0339	putative Methylglutaconyl-CoA hydratase	-2.3	-2.8	-3.8
CIBE_6076	acyl-CoA dehydrogenase	2.3	n.d	n.d
CIBE_0343	3-hydroxybutyryl-CoA dehydrogenase	-2.6	-3.1	-3.9
CIBE_1400	phosphate acetyltransferase	n.d	-2.4	-1.7
CIBE_0224	butyrate kinase	-2.7	-3.4	-2.8

5

C. beijerinckii protein	Predicted function	Log₂ fold change in expression relative to D-glucose cultures		
		3 h	6.5 h	10 h
CIBE_1401	acetate kinase	n.d	-2.3	-1.4
CIBE_0223	phosphate butyryl coenzyme A transferase	-2.6	-3.3	-2.7
CIBE_1204	pyruvate formate lyase I	-1.4	-1.8	-1.7
CIBE_1205	pyruvate formate lyase activating enzyme 1	n.d	n.d	-1.6
CIBE_4880	pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	2.8	3.4	-0.78
CIBE_4879	pyruvate ferredoxin/flavodoxin oxidoreductase, beta subunit	2.9	3.6	-0.94
CIBE_3432	acyl-CoA dehydrogenase	3.3	2.4	-2.3
CIBE_4352	pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	2.0	1	n.d
CIBE_4218	acyl-CoA dehydrogenase	1.5	3.3	n.d
CIBE_0327	electron transfer flavoprotein subunit beta	-2	n.d	-4.0
CIBE_0328	electron transfer flavoprotein subunit alpha	-1.4	n.d	-3.7
CIBE_1564	putative beta-hydroxyacid dehydrogenase	n.d	-1.0	n.d
CIBE_4841	butyrate kinase	1.7	1.4	1
CIBE_4329	acetyl-CoA acetyltransferase	n.d	-1.8	-1.5
CIBE_5515	butyrate kinase	2.7	2.4	2.4
CIBE_1684	pyruvate-flavodoxin oxidoreductase	1.2	n.d	2.8
CIBE_5454	putative Methylglutaconyl-CoA hydratase	1.8	3.9	1.4
CIBE_0340	acyl-CoA dehydrogenase	-2.6	-3.2	-4.0
CIBE_2634	putative hydroxyacid dehydrogenase	2.1	n.d	n.d
CIBE_5833	3-oxoacyl-(Acyl-carrier-protein) synthase III	-1.5	n.d	n.d
CIBE_5834	acetoacetyl-CoA reductase	-1.5	n.d	n.d
A/IBE production				
CIBE_0565	degradative acetoacetyl-CoA thiolase	-2.7	-3.3	-4.0
CIBE_6076	acyl-CoA dehydrogenase	2.3	n.d	n.d
CIBE_4607	putative aminoacyloate CoA-transferase	-3.7	-2.0	-4.9
CIBE_4609	acetoacetate decarboxylase	-3.6	-1.3	-4.2
CIBE_3470	NADP-dependent isopropanol dehydrogenase	n.d	n.d	-4.6
CIBE_2050	NADPH-dependent butanol dehydrogenase	3.9	n.d	-1.8
CIBE_3149	putative acyloate-acetoacetate CoA-transferase	2.8	4.1	n.d
CIBE_3150	acetoacetyl CoA-transferase (subunit A)	4.0	4.3	n.d
CIBE_3970	acetate CoA-transferase YdiF	n.d	-1.2	-2.5
CIBE_5462	iron-containing alcohol dehydrogenase	2.7	3.1	n.d
CIBE_2622	NADPH-dependent butanol dehydrogenase	4.5	3.5	1.4
CIBE_2354	alcohol dehydrogenase	-1.3	n.d	n.d
CIBE_5453	short chain acyl-CoA transferase: fused alpha subunit ; beta subunit	1.1	3.1	1.7

5

Table S6: Differential expression of the genes involved in other stationary phase processes. For stress response genes, a comparison was made with data from Saujet et al. (245)

C. beijerinckii protein	Predicted function	Log$_2$ fold change in expression relative to D-glucose		
		3 h	6.5 h	10 h
Sporulation				
CIBE_0044	putative 2-phosphosulfolactate phosphatase	-1.6	n.d	n.d
CIBE_0055	spore maturation protein	n.d	2.3	-0.8
CIBE_0056	spore maturation protein	n.d	2.4	n.d
CIBE_0064	conserved protein of unknown function	1.8	3.8	n.d
CIBE_0089	regulator required for spore cortex synthesis (stage V sporulation)	1.6	n.d	-0.9
CIBE_0098	transcriptional regulator	2.4	1.2	2.6
CIBE_0099	stage V sporulation protein B	n.d	3.0	n.d
CIBE_0103	YabP family protein	n.d	2.7	n.d
CIBE_0104	spore cortex biosynthesis protein YabQ	n.d	3.1	n.d
CIBE_0107	sporulation protein, spoIIE	n.d	4.1	-2.8
CIBE_0132	subunit of a sporulation, competence and biofilm formation regulatory complex	n.d	-1.0	-1.1
CIBE_0218	cell wall hydrolase, SleB	n.d	4.0	n.d
CIBE_0504	histidine kinase	1.7	n.d	n.d
CIBE_0535	spore coat protein CotS	1.4	2.7	2.2
CIBE_0537	conserved protein of unknown function	2.7	5.2	n.d
CIBE_0549	sporulation stage II, protein R	-1.81	n.d	n.d
CIBE_0578	peptidase M23B	1.3	5.6	n.d
CIBE_0579	transcriptional regulator	n.d	5.57	n.d
CIBE_0641	small, acid-soluble spore protein 1	-2.0	n.d	1.9
CIBE_0672	peptidase M50	n.d	3.7	0.8
CIBE_0730	conserved protein of unknown function	1.4	n.d	n.d
CIBE_0739	stage V sporulation protein R	3.4	4.7	n.d
CIBE_0761	histidine kinase	n.d	n.d	-0.8
CIBE_0992	anti-sigma F factor antagonist	n.d	n.d	-3.1
CIBE_0994	RNA polymerase sporulation-specific sigma factor (sigma-F)	n.d	n.d	-2.4
CIBE_1003	stage II sporulation protein P	n.d	4.2	-1.0
CIBE_1023	putative stage IV sporulation YqfD	n.d	n.d	-0.9
CIBE_1036	spore coat protein CotS	2.2	5.7	n.d
CIBE_1351	RNA polymerase sporulation-specific sigma-29 factor (sigma-K)	2.9	3.1	1.7
CIBE_1355	peptidase U4, Sporulation sigma-E factor-processing peptidase	n.d	3.0	-2.1
CIBE_1356	RNA polymerase sporulation-specific sigma-29 factor (sigma-E)	n.d	3.5	-2.2
CIBE_1357	RNA polymerase sporulation-specific sigma factor (sigma-G)	1.4	3.9	-1.9
CIBE_1358	essential sporulation protein	2.1	4.1	-1.5
CIBE_1362	multi-sensor signal transduction histidine kinase	n.d	n.d	-0.8
CIBE_1373	morphogenetic stage IV sporulation protein	3.2	5.1	-1.1
CIBE_1376	essential sporulation DNA binding protein; regulator of biofilm formation	n.d	n.d	-1,0
CIBE_1398	sporulation integral membrane protein YlbJ	n.d	4.0	n.d
CIBE_1448	sporulation protein YlmC/YmxH	n.d	1.9	n.d
CIBE_1451	spore DNA translocase	1.2	n.d	-0.86

5

C. beijerinckii protein	Predicted function	Log₂ fold change in expression relative to D-glucose		
		3 h	6.5 h	10 h
CIBE_1456	regulator required for dehydratation of the spore core and assembly of the coat (stage V sporulation)	6.0	2.0	2.0
CIBE_1525	sporulation protein	2.4	5.1	3.7
CIBE_1586	conserved protein of unknown function	2.7	5.3	5.4
CIBE_1605	sporulation protein SpoVB	n.d	4.0	n.d
CIBE_1674	small, acid-soluble spore protein beta	4.7	1.6	4.8
CIBE_1733	putative spore coat protein	n.d	n.d	1.8
CIBE_1735	spore coat protein	1.6	n.d	1.8
CIBE_1736	putative spore coat protein	n.d	n.d	1.7
CIBE_1789	multi-sensor signal transduction histidine kinase	2.1	n.d	n.d
CIBE_1807	sporulation protein	1.6	2.5	3.0
CIBE_1819	stage V sporulation protein D	1.6	n.d	-1.4
CIBE_1821	stage V sporulation protein D	n.d	n.d	-1.4
CIBE_1825	factor for spore cortex peptidoglycan synthesis (stage V sporulation)	n.d	n.d	-0.94
CIBE_1845	acetyltransferase	-1.3	n.d	n.d
CIBE_2020	ATP-binding stage III sporulation protein	n.d	4.0	-1.8
CIBE_2021	stage III sporulation protein AB	n.d	4.4	-1.7
CIBE_2022	stage III sporulation AC family protein	n.d	4.8	-1.4
CIBE_2023	stage III sporulation protein	n.d	5.5	-1.3
CIBE_2024	sporulation stage III, protein AE	1.2	6.4	-1.1
CIBE_2025	sporulation stage III, protein AF	n.d	6.0	-1.4
CIBE_2026	stage III sporulation protein AG	n.d	5.5	-1.9
CIBE_2027	stage III sporulation protein AH	1.9	5.7	-1.6
CIBE_2041	sporulation protein spo0A	1.2	n.d	-1.4
CIBE_2218	D-alanyl-D-alanine carboxypeptidase DacF	1.6	n.d	3.8
CIBE_2221	uncharacterized spore protein YtfJ	1.4	2.7	1.5
CIBE_2227	transcriptional regulator, MerR family	n.d	-1.2	-0.92
CIBE_2390	sporulation stage II, protein M	3.6	7.3	0.99
CIBE_2490	component of the inner spore coat	3.6	6.6	5.2
CIBE_2491	spore coat peptide assembly protein CotJB	4.4	7.6	5.8
CIBE_2492	conserved protein of unknown function	5.6	8.1	5.7
CIBE_2495	histidine kinase	n.d	n.d	-0.99
CIBE_2581	dihydroxyacetone kinase, N-terminal domain	1.8	2.6	1.1
CIBE_2775	small, acid-soluble spore protein beta	4.3	1.4	1.7
CIBE_2875	PAS/PAC sensor signal transduction histidine kinase	2.0	n.d	1.2
CIBE_2913	small, acid-soluble spore protein beta	3.5	3.4	4.0
CIBE_2950	histidine kinase	1.1	n.d	n.d
CIBE_3040	conserved protein of unknown function	2.1	5.2	5.2
CIBE_3507	conserved protein of unknown function	3.6	5.3	1.5
CIBE_3532	transcriptional regulator, AbrB family	2.0	1.8	-1.1
CIBE_3643	histidine kinase	1.8	n.d	n.d
CIBE_3647	sensor histidine kinase	1.2	n.d	n.d
CIBE_3649	small acid-soluble spore protein, alpha/beta type	n.d	n.d	4.7
CIBE_3698	small acid-soluble spore protein, alpha/beta type	4.7	2.1	5.4
CIBE_3923	spore cortex-lytic protein	3.7	3.8	0.91
CIBE_3935	small, acid-soluble spore protein beta	2.2	n.d	4.1
CIBE_3951	spore protein	2.9	n.d	3.7
CIBE_3965	spore protein	2.8	1.9	5.9
CIBE_4004	conserved protein of unknown function	1.4	n.d	n.d
CIBE_4085	transcriptional regulator, AbrB family	2.9	n.d	1.6

5

C. beijerinckii protein	Predicted function	Log$_2$ fold change in expression relative to D-glucose		
		3 h	**6.5 h**	**10 h**
CIBE_4156	conserved exported protein of unknown function	3.8	n.d	1.2
CIBE_4576	phosphatase	1.0	n.d	n.d
CIBE_4975	18 kDa heat shock protein	1.5	n.d	1.0
CIBE_5092	sporulation protein YunB	n.d	2.48	n.d
CIBE_5119	component of the flagellar export machinery	1.41	n.d	n.d
CIBE_5516	protein Tlp homolog	3.1	n.d	5.3
CIBE_5860	putative morphogen	n.d	-0.92	-1.0
CIBE_5891	transcriptional regulator, AbrB family	-1.8	n.d	1.4
CIBE_6123	chromosome partitioning protein; transcriptional regulator	n.d	n.d	-0.75
CIBE_6124	DNA-binding protein Spo0J-like	n.d	n.d	-1.1
CIBE_b0010	protein of unknown function	5.2	2.3	2.4
CIBE_0993	anti-sigma factor (antagonist of sigma(F)) and serine kinase	n.d	n.d	-2.9
Stress response				
CIBE_0139	class III stress response-related ATPase, AAA+ superfamily	1.0	n.d	n.d
CIBE_0289	rubrerythrin	3.0	n.d	-0.85
CIBE_0347	chaperonin small subunit	2.4	n.d	n.d
CIBE_0348	chaperonin large subunit	2.4	n.d	n.d
CIBE_0740	rubrerythrin	3.1	4.87	n.d
CIBE_2142	33 kDa chaperonin	-1.2	-1.06	n.d
CIBE_1008	heat-inducible transcription repressor HrcA	1.4	n.d	n.d
CIBE_1009	protein GrpE	1.3	n.d	n.d
CIBE_1010	molecular chaperone	1.6	n.d	n.d
CIBE_1502	Lon protease	1.2	n.d	n.d
CIBE_1581	ATP-dependent Clp protease proteolytic subunit	n.d	n.d	-0.96
CIBE_1740	superoxide dismutase [Fe]	1.7	n.d	n.d
CIBE_2199	superoxide dismutase [Mn/Fe]	1.8	3.48	0.89
CIBE_2745	reverse rubrerythrin-1	1.6	2.08	2.6
CIBE_2934	reverse rubrerythrin-1	4.4	n.d	n.d
CIBE_3517	rubrerythrin	4.5	5.08	0.77
CIBE_3635	rubrerythrin	-1.4	n.d	n.d
CIBE_4975	18 kDa heat shock protein	1.5	n.d	0.98
CIBE_3943	rubrerythrin	2.4	3.1	2.4

5

Chapter 6-
General discussion and conclusions

Introduction

Clostridia are widespread in the environment and can be found in diverse ecological niches (3) (soil, gastrointestinal tracts of humans and animals, sewage, etc.). Through their complex and diversified response to stress factors, clostridia can represent up to 20 % (329) of the total bacterial population in extreme environments such as geothermal and mineral environments. Over 200 species belong to the *Clostridium* genus, and amongst them, species of importance in different fields. Most studied species are pathogenic (such as *C. botulinum* or *C. tetani*) or probiotic strains (such as *C.butyricum* or *C. tyrobutyricum*) since they are of significant interest in human health. Others like solventogenic and acetogenic clostridia have been investigated for their biotechnological application. Their adaptability to adverse conditions and diverse metabolism are assets that fuel the enthusiasm displayed by researchers and the biotechnological industry alike. Despite their potential, little is known about the regulatory networks governing their adaptation to stress. However, the generation of microbial platforms for biochemical production requires accurate *in sillico* models, which integrate the knowledge acquired on regulatory and metabolic networks and a large set of potent synthetic biology tools (330).

Functional genomics is a branch of genomics that studies the relationship between the genome and the phenotype of an organism (331). Through this approach, regulation networks and metabolic pathways are studied at a genome-scale level. Data on genome, transcriptome, proteome, and even metabolome are gathered and crossed with other phenotyping techniques like microscopy to understand gene and protein function and generate models to predict cell behavior and response to environmental changes. This approach can help discover and annotate regulation and metabolic pathways, thereby giving some clues for strain engineering towards desired traits. Throughout this thesis, we aimed 1) to contribute to the advances in *Clostridium* gene engineering and 2) to use a functional genomic approach to expand our knowledge on the gene circuits governing stationary-phase events in solventogenic clostridia.

This chapter will:

i.   address the trends and improvements in the tools used to study solventogenic and acetogenic clostridia

ii. give insights into the morphological changes observed in clostridial cultures

iii. present issues raised by this thesis that have to be tackled to generate accurate genome-scale models and promote the industrial use of clostridia.

6

Trends and advances in the study tools for solventogenic clostridia

Functional genomics aims to integrate information on the genome and the phenotype to apprehend the regulation of cellular events. Functional profiling, along with protein localization and interaction studies, is crucial to obtain this information. This approach focuses on the targeted generation and characterization of mutants. Thus, it requires a broad set of tools from synthetic biology to engineer genome and transcriptome and assess the impact on the phenotype in-depth.

Hurdles and progress in genome engineering of *Clostridium*

Compared to other microorganisms used in the biotechnical industry, engineering clostridia comes with several hurdles; some are listed in Table 1. Furthermore, because of differences in cell machinery and growth conditions, systems developed for model organisms like *Escherichia coli* or *Bacillus subtilis* cannot be directly applied to *Clostridium*. For example, fluorescent reporters from the GFP family cannot be used in clostridia to monitor expression or protein localization in real-time because these molecules require exposure to oxygen for maturation (234).

The transformation efficiency of clostridia with current methods is at least 100 times lower than the transformation efficiency of well-established platform organisms like bacilli. Moreover, the introduction of linear DNA in *Clostridium* is close to impossible. Therefore cloning has to be done in intermediate strains such as yeast (see Chapter 2), *Escherichia coli* (332–334) with shuttle vectors as depicted in Figure 1. The addition of a replication origin and resistance marker for the intermediate strain increases the vector's size, affecting the transformation efficiency of clostridia. Furthermore, clostridial genomes are AT-rich (genomes containing 30% GC on average), and the introduction of sequences with less than 30% GC can be toxic for the cloning strain. Indeed, Raghavan et al. (335) showed that the horizontal transfer of AT-rich genes in *E. coli* impaired strain fitness. This toxicity is caused by the overheating of the transcription machinery and the generation of spurious RNA. In *E. coli*, the housekeeping σ^{70} promoter recognizes AT-rich motifs, and in the presence of AT-rich horizontally transferred genes, the σ^{70} promoter binds aleatory on the gene sequence (336). These findings explain the difference in promoters' stringency in *Clostridium* compared to *E. coli* highlighted by Mordaka et al. (337) recently. In their study, three synthetic promoter libraries were generated with different percentages in identity (25%, 58% and 79%) with the native *thl* promoter. Their efficiency in *E. coli* and *Clostridium* were compared and despite being all functional in *E. coli,* only the 58 or 79% identity libraries were active in *Clostridium*, with a larger number of active promoters in the 79% library. A similar event might take place in *B. subtilis,* but no study was made on that subject mainly because cloning in *B. subtilis* is not widespread. A higher success rate

6

in construction can be achieved using *S. cerevisiae* as cloning strains, which has a genome consisting of 38 % GC. However, this method is not ideal because the plasmid yield from *S. cerevisiae* is too low to transform *Clostridium,* so an additional amplification by *E. coli* is required.
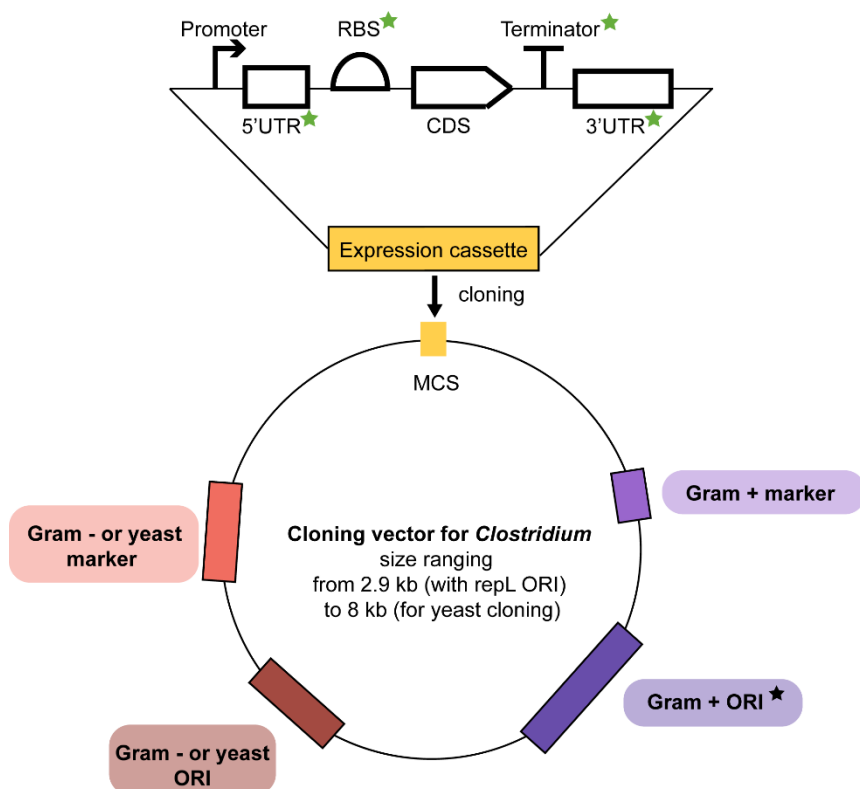
Figure 1: Scheme of genetic constructs used for genome engineering in *Clostridium*. The stars next to some genetic elements indicate the need for standardization; the green stars indicate the absence of libraries. UTR: untranslated region, RBS: ribosome binding site, CDS: coding sequence, MCS: multiple cloning site, ORI: origin of replication

Besides, variability in the defense mechanisms and antibiotic resistance was observed within the *Clostridium* genus (338). To date, two restriction-modification systems have been reported amongst solventogenic clostridia (213, 216). While *C. beijerinckii* DSM 6423 and NRLB 598 only accept dam-/dcm- DNA, the transformation of *C. acetobutylicum* with exogenous DNA requires dam+/dcm- methylation. Thus each tool needs to be tailored for the strain of interest, as shown in Chapter 2. Despite those hurdles, the toolkit for genome engineering in clostridia is expanding (19, 339, 340), and new genome engineering techniques and building blocks for DNA constructs such as promoters and reporters were added to the engineering toolbox since 2015.

Table 1: Reasons for some of the hurdles encountered in *Clostridium* genome engineering and comparison to other microorganisms used in the biochemical industry. * GC content according to (341) and the genome sequences available on NCBI. DBSR: double-strand break repair; SDSA: synthesis-dependent strand annealing; BIR: Break induced replication

	S. cerevisiae	*E. coli*	*B. subtilis*	*C. beijerinckii*	Consequences for synthetic biology in *Clostridium*
GC content*	38 %	50 %	43 %	30%	Cloning difficulties, synthetizing genes
Homologous recombination pathways	DBSR, SDSA and BIR (342)	RecBCD and RecFOR (343)	RecBCD and RecFOR (343)	RecFOR but incomplete RecBCD RecU missing (344)	Rare homologous recombination and double-crossover events
Transformation efficiency (cfu/µg)	10^8 (209, 345)	10^{10} (346, 347)	10^5-10^7 (348–350)	10^3-10^5 (119, 216, 334, 351, 352)	Low transformation efficiency
Cloning possible	Yes	Yes	Yes	Not reported	Use of an intermediate cloning strain required

Progress in genome engineering

Targeted mutagenesis

Before 2015, few methods enabled a targeted and scarless gene-editing in clostridia, and most mutants were generated using the Clostron (190, 191) and Targetron (353) systems. These transposon-based techniques enable a targeted insertion of an antibiotic maker in the genome. However, the introduction of selection markers causes polar effects (354), affecting the evaluation of the mutation on the phenotype. And with these techniques, engineering the same strain more than twice was not possible because only two antibiotic resistance genes are functional in *C. beijerinckii* or *C. acetobutylicum*. Other methods relied on homologous recombination (194, 355), but since the clostridial homologous recombination pathway is not complete (Table 1), that approach is less efficient than in other organisms. Double-crossover events are rare, making the generation of mutants with these systems very laborious.

In 2015, the first CRISPR-based genome engineering tools for *Clostridium* were described, and since then, several systems were developed and adapted for solventogenic clostridia. In the last five years, targeted gene deletion, insertion, replacements and even single nucleotide modifications using CRISPR-Cas9 and Cas12-based systems were devised for

6

C. acetobutylicum and *C. beijerinckii* (198, 199, 356, 357). The work presented in Chapter 2 is an example of a system developed for *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* DSM 6423. Chapter 2 also illustrates the necessity for tailored tools for each target strain; despite belonging to the same genus, the system developed earlier in *C. acetobutylicum* (203) could not be directly used in *C. beijerinckii* because of differences in the sensitivity to antibiotic markers. The efficiency of CRISPR-based methods is still being improved by controlling the lethal double-strand break activity caused by Cas9 (358). The Cas protein needs to be inactive to introduce the editing machinery into the cells. Xue et al. (196) developed a CRISPR-Cas9n tool, a nickase variant of the initial Cas9 gene, to circumvent this issue. However, in contrast to the Cas9 protein, the Cas9n is not a lethal counter-selection marker, decreasing genome editing efficiency. We used inducible promoters, one sensitive to xylose and a second to anhydrotetracycline, to preserve the efficiency of the counter selection marker, as described in Chapter 2.

Despite using inducible promoters to reduce Cas9 toxicity, a basal expression of *cas9* remains. And obtaining transformants harboring the Cas protein remained challenging; as observed with *C. beijerinckii* DSM 6423 (Chapter 2), 20 µg of pDNA was required to obtain one transformant. Recently, Wasel et al. (359) added an anti-CRISPR protein, AcrIIA4, to improve the tool, which significantly increased transformation efficiency. Cañadas et al. (357) tightened the expression of *cas9* by using a theophylline-dependent riboswitch. This tool named RiboCas is interesting because of its universality; no adaptation was required for other *Clostridium* species. Another drawback of those CRISPR-Cas systems is the requirement of homologous arms for gene-editing. As described in Chapter 2, CRISPR-Cas gene editing relies on homologous recombination and the selection of the recombinants by activating the Cas protein. An editing template consisting of homologous regions flanking the desired location of the mutation has to be provided to the cells on a plasmid. Since the homologous recombination is not very efficient in *Clostridium*, a minimum length of 300 bp per homology arm is required. The editing template increases the plasmid size and decreases the transformation efficiency, making the insertion of large sequences challenging. To bypass this issue, we developed with Wasel et al. a two plasmid CRISPR-Cas system (Chapter 2) that enabled the insertion of a 3.6 kb fragment into the genome of *C. acetobutylicum* (203). Huang et al. (360) addressed this matter differently by relying on a serine integrase-based site-directed recombination rather than homologous recombination. With their system, the butyric acid production pathway of *C. acetobutylicum* (8.5 kb) was inserted into the genome of *C. ljungdahlii*.

Next to gene-editing techniques, transcriptome engineering methods have been used to study gene function, especially when the genes cannot be disrupted. Antisense RNA (asRNA) was, for a long time, the only tool to downregulate gene expression in *Clostridium*

(66, 113, 361). But since 2015, more powerful tools have been developed. Changhee Cho and Sang Yup Lee adapted for *C. acetobutylicum* (362) a small regulatory RNA (sRNA) based method which has been successfully implemented previously in *E. coli* for multiplex gene knockdown (363). This sRNA-based method proved to be more efficient than asRNA, but multiplex editing with this tool remains demonstrated in *C. acetobutylicum*. Concurrently, CRISPR interference (CRISPRi) systems were devised. In these systems, the Cas protein cannot cut the DNA but still recognizes and binds to the DNA with the same fidelity as the native CRISPR-Cas protein (105, 356). While asRNA, sRNA and CRISPRi prevent transcription by binding to a target sequence, Li et al. (364) investigated another approach by editing the 5' untranslated regions (5'UTR) of the gene of interest and demonstrated that variation in the 5'UTR sequences greatly impacted gene expression in *C. acetobutylicum*.

Random mutagenesis

Random mutagenesis enables the creation of mutants libraries (217) and selection of strains according to their phenotype before studying their genotype. In a functional genomic approach, random mutagenesis can help solve gene function when complemented by targeted gene-editing. For example, the role of *sig54* in *C. beijerinckii* DSM 6423 (126) and *spo0A* in solvent production in *C. pasteurianum* (127) were discovered through this approach. Mutants libraries can be obtained using directed evolution with chemical or physical agents (126, 365), genome shuffling (185, 188, 366), or through aleatory insertions of transposons in the genome (367).

Improvement of building blocks in synthetic biology

As exposed in several reviews (19, 338, 368), besides the difficulties of genome editing in clostridia, a lack of standardized building blocks, illustrated in Figure 1, to accurately monitor or quantify gene expression and protein activity hinders efficient synthetic biology in *Clostridium*, slowing down the study of their physiology.

To control transcription

A reduced set of promoters are used in constructs designed for clostridia, as discussed in several reviews (19, 368). Most of them are native to clostridia and have been detected through genomic and transcriptomic analysis. Their strength can vary according to the growth phase or the species (337, 364, 369). The regulation of cellular events is very delicate; thus, successful strain engineering requires fine control of the expression levels of the cloned gene, as illustrated for the development of CRISPR genome engineering tools. Therefore, in addition to inducible promoters, there was a need for promoter libraries with standardized strength. In 2017 and 2018, two studies developed synthetic promoter libraries by mutation of the *thl* promoter from *C. acetobutylicum* ATCC 824 (337, 369).

6

In addition to promoters, 5' untranslated regions (UTRs) were edited to modulate gene expression in solventogenic clostridia. Lee et al. (364) revealed the importance of these UTRs by generating several synthetic UTRs and evaluating their impact on gene expression, while Cañadas et al. (357) used riboswitches to modulate Cas expression.

To monitor expression and translation

The presence of efficient reporters is crucial for the characterization of transcriptional and translational modulators. They can give information on transcription and translation levels in different conditions.

Several enzymatic reporters were developed for *Clostridium* (368), but few potent fluorescent reporters were successfully applied. In aerobic bacteria, fluorescent reporters (mainly GFP derivatives) have been successfully used only *in vitro* since they require oxygen for maturation. In Chapter 3, the mCherry protein was used to localize the *spoIIE* protein in the cell. The cells were fixed aerobically, and the signal detected was rapidly bleached. Thus enzymatic reporters have been used to evaluate promoter strength (337, 364, 369). Three types of oxygen-independent fluorescent reporters have been developed for *Clostridium*: flavin mononucleotide-based fluorescent proteins (FbFP) (370, 371), fluorescence-activating and absorption-shifting tag (FAST) proteins, Halo tag proteins and SNAP-tag proteins (372, 373). Their size varies from 14 kDa to 33 kDa, and they cover a broad fluorescent spectrum from blue ($\lambda = 362$ nm) to far-red ($\lambda = 690$ nm). Over 25 different fluorescent reporters can now be used in solventogenic clostridia. FbFPs were introduced into the *Clostridium* engineering toolbox in 2015 (374) and have been used in several species, including *C. ljungdahlii* (375), *C. acetobutylicum* (362) and *C. beijerinckii* (371). Yet, only one color is available, and low *in vivo* fluorescent intensity was reported when used in *C. ljungdahlii* and *C. acetobutylicum* (337, 375). The recently developed fluorescent systems FAST, Halo tag, and SNAP-tag cover most of the fluorescent spectrum, and their *in vivo* fluorescent activity in *C. ljungdahlii* and *C. acetobutylicum* is 100 times higher than the intensity reached with FbFPs (373). These new reporters will indeed significantly contribute to more prolific studies on the physiology of clostridia.

6

The rise in "omics" approaches to study solventogenic clostridia
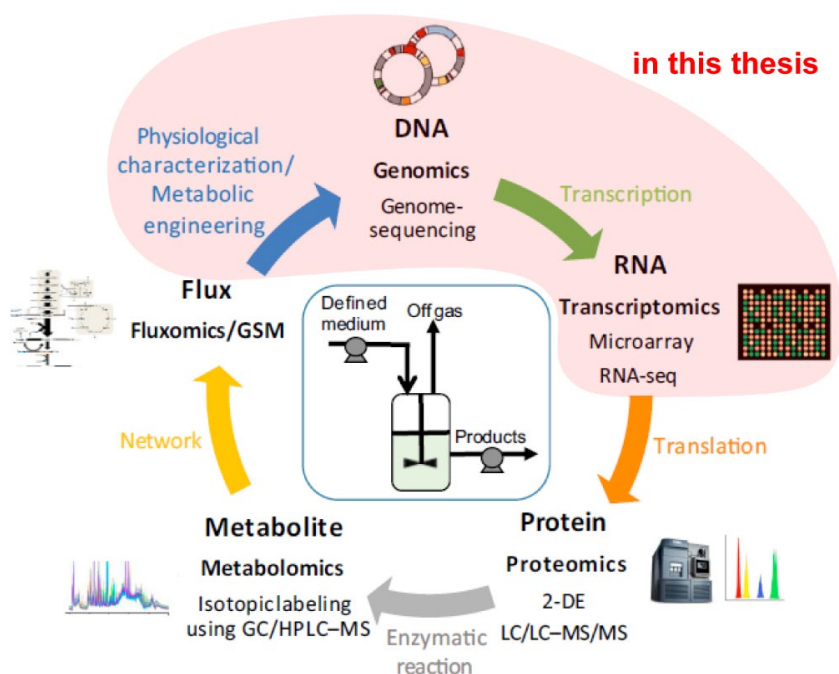
Figure 2: Toward combinatory "omics", a high throughput analysis of genotype and phenotype. Abbreviations: 2-DE, two-dimensional electrophoresis; GC, gas chromatography; GSM, genome-scale metabolic model; HPLC–MS, high-performance liquid chromatography-mass spectrometry; LC/LC-MS/MS, two-dimensional reversed-phase liquid chromatography-tandem mass spectrometry. Adapted from (376) with the permission of Trends in Biology.

Genomics & Transcriptomics

Genome and transcriptome-based studies have become increasingly popular in the last ten years. The decrease in the cost of high throughput sequencing led to a significant increase in sequenced genomes and transcriptome studies. Between 2014 and 2020, the number of genomes of solventogenic clostridia available on the NCBI database rose from 10 to 333, while the transcriptome profiling data doubled (Figure 3). In this thesis, genome and transcriptome analysis complemented with gene engineering was used to study *C. beijerinckii*.

The data generated by genome sequencing represent a massive source of information that can be exploited through comparative genomics: to identify and compare regulatory and metabolic pathways, detect single nucleotide polymorphisms (SNP), and study the epigenome with long-read sequencing technologies. The rise in annotated genomes has improved gene annotation algorithms, which facilitates functional gene annotation. Most functional genomic studies in solventogenic clostridia start by identifying, in the target organism, homologs from the model anaerobes *C. acetobutylicum* or *B. subtilis*. Similarly,

we identified the *spoIIE* gene and most sporulation-related genes in Chapter 3. The same procedure was used in Chapter 5 to detect the genes involved in rhamnose metabolism. However, a more exhaustive genome comparison analysis was required because neither *C. acetobutylicum* nor *B. subtilis* can convert L-rhamnose to propanol and propionate.

Genome sequencing itself is also used to confirm the correct edition of the genome and, in the best cases, the absence of SNPs in genomes. Genome sequencing has become compulsory in most journals to publish articles on engineered strains. As illustrated in Chapters 3 and 4, scarless genome engineering with a powerful tool such as CRISPR-Cas9 does not yield 100% identical mutants since the strain can intrinsically be prone to mutations. Information on SNP number and location helps evaluate the relationship between the genotype and the observed phenotype. Unfortunately, in Chapter 4, we could not pinpoint which mutation(s) led to the observed phenotype.

Similar to genomic analysis, transcriptome studies in solventogenic clostridia have risen in popularity (Figure 3B), but to a lesser extent because design, conduction of transcriptome experiments and data analysis are laborious. Transcriptomics give useful information on the active portion of the genome in organisms that harbor various gene duplications, like *C. beijerinckii*. Besides, it provides an insight into the modulation of gene expression in response to environmental changes; that is why we chose this approach to investigate the interruption of the sporulation process (Chapter 3), the loss of solvent production capacity (Chapter 4) and the formation of bacterial microcompartments (BMC) in *C. beijerinckii* (Chapter 5). We studied the interdependence of regulatory networks governing sporulation with other cellular events (amino acid and secondary metabolism, iron transport) through this approach. Depending on the angle chosen to analyze the data, different information can be gathered, as displayed by the work of Patakova and her team. Both the regulation of the transcriptome during fermentation (75) and the role of amino acids, vitamins and ions in the transition from exponential to stationary growth (365) were assessed in the *C. beijerinckii* NRRL B598 using the same transcriptomic data.

Two techniques can generate transcriptomic data: microarrays and high throughput RNA sequencing (RNA-seq). As observed in Figure 3B, microarray-based is preferred when working with *C. acetobutylicum;* the groups generating these data have been the first to perform transcriptome analysis in clostridia and thus have a set of microarrays available. However, when dealing with other species and in other groups, the material required for transcriptome analysis has to be purchased. It seems that in that case, high throughput sequencing-based transcriptome analysis is preferred. High throughput sequencing enables a more comprehensive investigation. Compared to microarrays, RNA-seq is more sensitive to rare mRNA, or mRNA from mutated genes, and can detect mRNA from unmapped genes. When dealing with non-conventional organisms with numerous genes of unknown functions like in *C. beijerinckii*, this represents a considerable advantage.

6

A.

B.
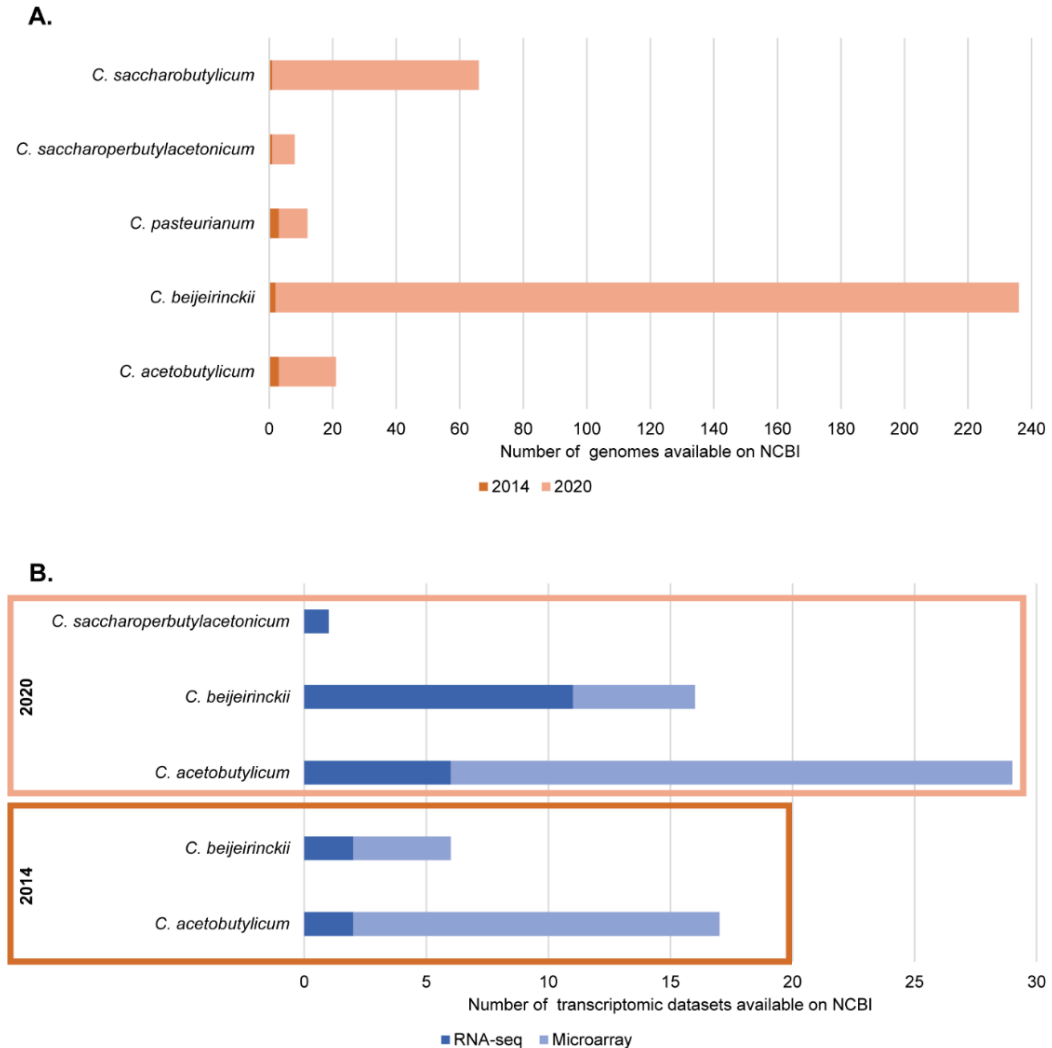
Figure 3: Publicly available (A) genomic and (B) transcriptomic data of solventogenic clostridia in 2014 and 2020 on the NCBI database.

Proteomics, metabolomics and fluxomics

Compared to transcriptomics, proteomics gives additional information on the post-transcriptional regulation and thus a better estimate of protein expression. For example, UTR regions can increase the translation of mRNA present at low levels in the cell (377), and abundant transcript can be degraded quickly, yielding lower protein levels than predicted by the transcriptome analysis. In solventogenic clostridia, proteomic studies have been mainly done in *C. acetobutylicum* and *C. beijerinckii* using mass spectrometry (MS) and 2D electrophoresis (71, 145, 377–379). Using those methodologies, the focus was made on detecting the protein involved in acidogenesis, solventogenesis and degeneration processes. Still, determining protein levels is insufficient to have a correct estimation of

187

protein activity because of post-translational modification events. Several post-translational modifications affect the protein activity in bacteria (62), with phosphorylation, acylation, succinylation, glycosylation, and pupylation being the most common. In solventogenic clostridia, post-translational modifications were reported several times in *C. acetobutylicum* (145, 380–383). Furthermore, studies investigated mainly the phosphoproteome; however, histidine phosphorylation, a central player in signal transduction, could not be detected because of the limitations of the techniques used. Recently, a second PTM, protein acylation, was investigated in *C. acetobutylicum* by Xu et al. (71), who showed the importance of butyrylation for Spo0A activation and Buk activity.

With the advancement in chromatography and MS techniques in the last years, high throughput analysis metabolites and flux have become more reliable. Metabolomics data have been mainly used in combination with fluxomics to study the central metabolism during the fermentation and to generate genome-scale models of the metabolic flux (384–387). More recently, metabolomics data have been used to study the response to changes in the culture (addition of trace elements, butanol stress) and the secondary metabolism (51, 82, 83, 388, 389).

Flow cytometry and imaging

Flow cytometry has been used in biomedical research for a long time and is now an established technique to screen for different cell types in tissues and cultures. Still, its use is not as systematic for studying microorganisms, and in the *Clostridium* field, flow cytometry is mainly used to detect changes in cell morphology during fermentation (109, 138, 390–393). Flow cytometry could not be used to its full potential because of the lack of appropriate fluorescent labels for multidimensional analysis. With the development of new fluorescent reporter systems, Papoutsakis and his team expanded the number of studies possible with flow cytometry (373). They further demonstrated in a recent study (394), how they could use flow cytometry to investigate the exchange of RNA between *C. acetobutylicum* and *C. ljungdahlii* cells in co-cultures. This study also shows how the new fluorescent reporters can be used with sophisticated imaging techniques, such as super-resolution microscopy and multiplex imaging.

Future perspectives

Significant advances have been made in the field of genome engineering of clostridia since 2015. CRISPR-based systems now enable scarless and efficient gene-editing. CRISPR-systems could also be used for random mutagenesis. Mutant libraries could be generated by providing the CRISPRi machinery with aleatory generated target sequences as demonstrated in *E. coli* (395), or with a modified CRISPR base editing tool, which binds to several spots in the genome.

To ease the design and realization of functional genomic more tools are required. Synthetic promoter libraries and fluorescent reporters are now available, but more research is required on other building blocks. Determining the relative stability of the origin of replications used for cloning, plasmid copy number, and compatibility for each strain would be useful, especially to study the impact of gene overexpression on the phenotype or create new pathways. The relative vector stability was determined for *C. acetobutylicum* (396) but not in other solventogenic clostridia. Synthetic libraries of RBS and terminators would improve genome engineering in clostridia. Moreover, a broader knowledge of the UTR in clostridia would give a useful insight into transcriptional regulation in clostridia. If 3'UTR regulation is found to be universal in all clostridia, it would be a great leap towards the design of a universal synthetic toolbox for *Clostridium*.

"Omics" have become standard approaches to study *Clostridium* as the price of high throughput sequencing has been decreasing, and the "omic" data have become more available. With the increasing requirements for researchers to make their data publicly available, it is now possible to perform analysis with data generated by another research group, as illustrated by Sandoval-Espinola et al. (397), who used the transcriptome analysis from Wang et al. (398). Still, some progress has to be made because most available data is raw and not suitable for quick inquiries. It started to evolve; EMBL-EBI developed the Expression Atlas database (https://www.ebi.ac.uk/gxa/home) that enables the visualization of transcriptomic experiments deposited on the Array Express database. But unfortunately, studies on prokaryotes cannot be found on the Expression platform. Similarly, the same service is offered on the Mage Microscope platform (https://mage.genoscope.cns.fr/microscope/home/index.php), where depositors can ask the platform to make their transcriptional analysis publicly available. In both cases, the data analysis is performed by an algorithm developed by the platform to simplify user access to the data.

Despite the rise in "omic" analysis, few studies combine more than two approaches to study the physiology of solventogenic clostridia (376), partially because of the complex data analysis. Those studies generate a large amount of data, and their analysis requires bioinformatics skills that are not available in every research group. Nevertheless, this is set to change with the development of cheap and user-friendly data analysis platforms and software. In parallel, bioinformatics is being more integrated into study programs for life sciences, resulting in a widening of the knowledge for the new graduates.

"Omics" are bound to become standard in *Clostridium* research, and soon more reliable genome-scale models will be generated thanks to combinatorial "omics" and the advances in the technologies used in "omic" studies. With the progress in sequencing, chromatography and MS technologies (399, 400), data generated give more detailed information on the epigenome or PTMs. Furthermore, more powerful techniques are being

6

developed by combining them with other high throughput techniques such as flow cytometry for single-cell "omics" (391, 401, 402) or imaging (403, 404). In batch fermentation conditions, *Clostridium* cultures present heterogeneity of populations, thus applied to the study of solventogenic clostridia, single-cell transcriptomics/metabolomics could enable the identification of high butanol producing cells, a knowledge that could be used for strain engineering. Single-cell analysis could also help to solve the mystery of the role of cell autolysis in sporulation. Liu et al. (108) showed that preventing autolysis in *C. acetobutylicum* impaired sporulation, and they suggested that cell autolysis provided sporulating cells with crucial metabolites. Another novel high throughput method is imaging mass spectrometry (IMS), which combines imaging and mass spectrometry; applied to *C. beijerinckii*, it could be used to study the protein turnover in microcompartments.

Insight into the morphological changes of solventogenic and acetogenic clostridia

As described in several studies on solventogenic clostridia (38, 155, 275, 394, 405, 406), several changes in the cell morphology can be observed during the cultivation of *Clostridium* (Figure 4). Depending on the conditions, sporulation, strain degeneration, granulose accumulation, biofilm, or bacterial microcompartment (BMC) formation can be observed in the culture. The triggers of these physiological changes and their impact on solvent formation are still being investigated. And the knowledge of these morphological changes is expanding with the advances in clostridial engineering and comprehensive phenotyping approaches, shedding some light on the regulation and connections between cellular events. Three cellular events (sporulation, strain degeneration, and BMC formation) observed in *C. beijerinckii* cultures were investigated in this thesis, combining gene engineering, genome comparison and transcriptome analysis.

The following section gives an overview of the current knowledge on the triggers and regulation of these physiological changes in solventogenic and acetogenic clostridia. Acetogenic clostridia are C-1 compounds utilizing microorganisms. They can fix CO_2, CO, or formate to produce acetate through the Wood-Ljungdahl pathway. Because of their unique metabolism and the development of biobased processes for chemical production, acetogens caught the interest of the biotechnological industry. Anaerobic acetogens are particularly interesting for industrial use since their growth in anaerobic conditions reduces flammability concerns linked to the use of CO and H_2 and reduces contamination risks (407). Fourteen acetogenic species have been identified within the *Clostridium* genera, and among which, six (*C. ljungdahlii, C. coskatii, C. drakei, C. carboxidivorans, C. ragsdalei, C. autoethanogenum*) are also natural solvent producers. Numerous studies on these clostridial acetogens have been done in the last ten years. Lanzatech currently managed to

6

scale up its gas fermenting process based on CO conversion to ethanol and operates a commercial plant (408).
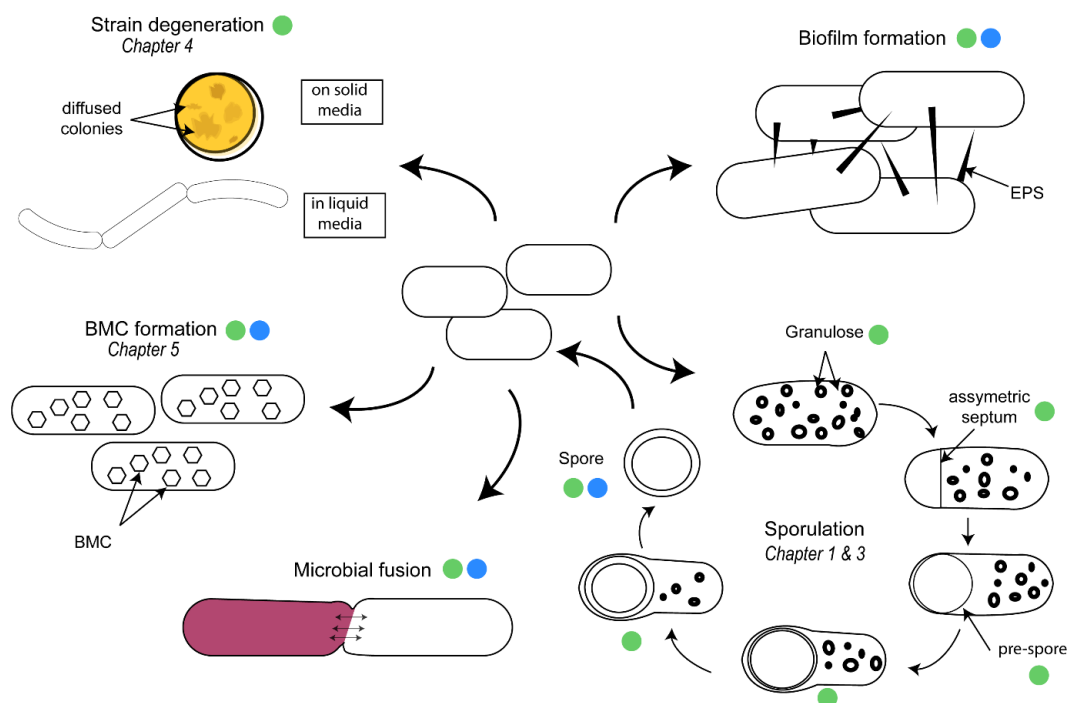
Figure 4: Changes in clostridial cell morphology. The detection of degenerated cells can be done on solid media. Degenerated colonies were reported to be thin and have an irregular shape. While strain degeneration and sporulation were observed with most solventogenic clostridia, BMCs and microbial fusion were not. Indeed, no BMC encoding cluster was seen in the genome of *C. acetobutylicum*, and microbial fusion was reported only in *C. acetobutylicum* co-cultures. BMC stands for bacterial microcompartment, and the purple cell represents a bacterium from a different species. Green full circles indicate changes observed in solventogenic clostridia, and blue full circles indicate changes reported in acetogenic clostridia.

Sporulation

In solventogenic clostridia

Sporulation induces multiple changes in cell morphology, and as described in Chapter 1, sporulation is observed mainly during batch fermentation. Studies on solventogenic clostridia showed that sporulation frequency and intensity vary depending on media composition (carbohydrate, metal content, amino acids) and cell density. Sporulation mechanisms in solventogenic clostridia were described in Chapter 1 of this thesis. The following section will focus on granulose formation, which is suggested to be linked with the initiation sporulation.

The first change observed at the onset of sporulation is the accumulation in the cells of granulose, a starch-like polymer of glucose, in the form of granules. Similar to bacterial

6

glycogen, granulose consists of α-1,4-linked glucose subunits but without branched α-1,6-linked glucose (235, 266). A variation in the composition of granulose has been observed between clostridial species. Its metabolic pathway, schematized in Figure 5, is predicted to be similar to the one described for glycogen-producing bacteria such as *E. coli* and *Salmonella* (257). Homologs to the *glg* genes encoding the enzymes enabling glycogen formations were identified in several *Clostridium* species. In *C. pasteurianum*, inactivity of GlgC/D, an ADP-glucose pyrophosphorylase, or GlgA, the granulose synthase hindered granulose formation (235). Similarly, in *C. acetobutylicum*, a *glgA* deficient mutant could not produce granulose (104).
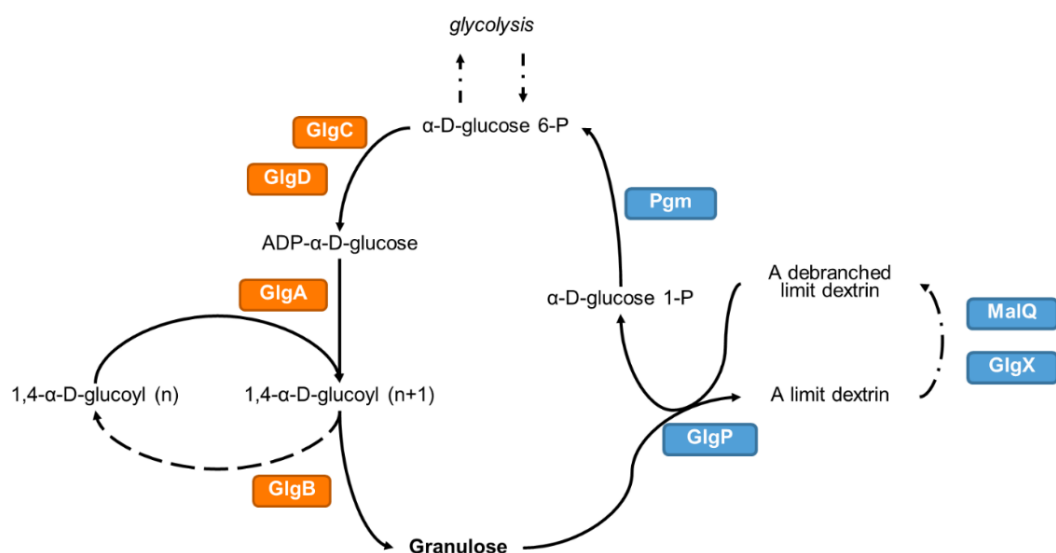
Figure 5: Putative metabolic pathways of granulose formation in orange and degradation in blue Figures adapted from Microcyc.

The onset of granulose accumulation precedes solvent formation and depends on the nature and the concentration of carbohydrates present in the medium (36, 235, 266). In *C. acetobutylicum*, the generation of *ccpA* and *csrA* deficient strains showed that both CcpA and CsrA, transcriptional regulators involved in the carbohydrate metabolism, promote the expression of *glgA* (42, 409).

The link between granulose formation and sporulation, and solvent production is not fully elucidated. As for solvent production and sporulation, Spo0A and its regulators affect granulose formation. In *C. pasteurianum* and *C. beijerinckii, spo0A* null mutants did not accumulate granulose (118, 127). In *C. acetobutylicum* ATCC 824, overexpressing *spo0A* increased granulose formation (103), but strains lacking the orphan histidine kinases activating Spo0A or σ^K, which promotes the transcription of *spo0A*, had an impaired granulose accumulation (17, 111). However, decoupling granulose formation from solvent formation and sporulation is possible. In *C. acetobutylicum*, disruption of polyketide

synthesis (83), the *agr* operon encoding the Agr quorum-sensing system (90), or the sporulation-specific sigma factors *sigE* and *sigF* (109) all prevented granulose accumulation but not solvent formation. Besides, sporulation and granulose formation were decoupled in *C. acetobutylicum* and *C. beijerinckii*. While the disruption of any RRNPP systems did not affect granulose formation in *C. acetobutylicum,* solvent formation and sporulation were impaired in the *qsrG* mutant (92). In Chapter 3, the *C. beijerinckiiΔspoIIE* mutant did not sporulate but accumulated granulose.

After granulose accumulation, the cell is divided by an asymmetric septum in two compartments, the forespore (future spore) and mother cell (with the granulose), and a network of around 120 genes regulates the completion of each sporulation stages leading to the release of the spore in the culture (Chapter 1). To assess the impact of an early interruption of the sporulation process, we disrupted *spoIIE* in *C. beijerinckii* (Chapter 2) and conducted a transcriptome analysis using RNA-seq (Chapter 3). Highlights of the work presented in Chapter 3 include the differences in the regulation of sporulation in *C. beijerinckii* and *C. acetobutylicum* and a link between sporulation and amino acid metabolism and secondary metabolism.

In acetogenic clostridia

Few data are available on the sporulation mechanism in acetogens. In fact, sporulation was rarely observed in the cultures of acetogenic clostridia. Still, some brief descriptions of the morphology of sporulating cells can be found for *C. coskatii, C. drakei*, *C. carboxidivorans, C. ragsdalei* and *C. autoethanogenum* (410–413). As described for solventogenic clostridia, these species form sub-terminal and terminal spores. During sporulation, swelling of the cells was reported in *C. coskatii* (413) and *C. carboxidivorans* (411)*,* as described for other solventogenic clostridia. In contrast, no swelling of *C. autoethanogenum* (410) and *C. ragsdalei* (412) cells was observed. A significant difference to most solventogenic clostridia is the absence of granulose in sporulating cells of acetogens.

To verify whether the reduced sporulation was due to a difference in the set of sporulation genes, we performed a blastp analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi) based on the minimal set of sporulation genes defined by Galperin et al. (98). Homologs to the 52 core sporulation genes were found in the above-mentioned solvent producing acetogens. Further, we searched for homologs of genes known to be involved in the initiation of sporulation in solventogenic clostridia. In all the genomes analyzed, homologs of the Agr and RRNPP quorum-sensing systems, as well as homologs of orphan kinases, were identified. As for secondary metabolites, no polyketide encoding gene cluster (80) was found; still, all strains harbor ranthipeptide producing enzymes. A recent study on the ranthipeptides produced by *C. beijerinckii* and *C. ljungdahlii* highlighted their importance in regulating cellular events (84); still, no link with sporulation in *C. ljungdahlii* could be made. Interestingly, homologs

6

to the enzymes involved in granulose formation were found in *C. drakei* and *C. carboxidivorans*, but not *C. ljungdahlii*, *C. autoethanogenum* or *C. coskatii*.

The low sporulation frequency observed in acetogens seems to be due to another factor than the absence of sporulation genes in the genome. Differences in triggers, transcriptional or post-transcriptional regulations of sporulation proteins during fermentation could explain this low frequency. Using publicly available transcriptomic data of *C. ljungdahlii* or *C. autoethanogenum* (414–417), we studied the expression profile of sporulation gene homologs under various conditions. When grown under salt stress conditions, *C. ljungdahlii* did not form spores despite the up-regulation of *spo0A* (414). An interesting study on the variation of the transcriptome of *C. ljungdahlii* under lithotrophic and organotrophic growth conditions (CO_2/H_2 vs. fructose) reported the up-regulation of an *agr* homolog and nine σ^E regulated sporulation genes during lithotrophic growth (415). Surprisingly, no change in *spo0A* expression was observed. A similar change was observed when *C. ljungdahlii* was exposed to oxygen (416). The homologs of *abrB*, *sigE,* and three late-stage sporulation genes were up-regulated in the O_2 challenged cultures. These results show that sporulation genes are expressed during the fermentation in *C. ljungdahlii* and that their expression varied in response to the environmental changes (changes in carbon source and presence of oxygen) as observed in other clostridia. This observation was confirmed by the transcriptome analysis done on *C. autoethanogenum* (417). Diender et al. (417) studied variations of the transcriptome of *C. autoethanogenum* in CO/H_2 compared to CO grown-cultures and in mono-culture compared to co-culture with *C. kluyveri*. In CO/H_2 grown-cultures, several stage II and III sporulation genes were up-regulated, and a slight increase in the expression of the sporulation-specific sigma factors was detected. In the synthetic co-cultures, an up-regulation of *spo0A* and *spoIIE* homologs was observed, while *sigF* and some sporulation genes involved in cortex formation were down-regulated. According to these results, the regulation of the sporulation in *C. ljungdahlii* and *C. autoethanogenum* seems to have common characteristics with the systems described in other solventogenic clostridia. Like in solventogenic clostridia, carbon sources impact sporulation regulation in *C. ljungdahlii* and *C. autoethanogenum.* Indeed, their growth in lithotrophic conditions led to an up-regulation of sporulation genes compared to organotrophic conditions. Moreover, the *sigE* and *sigF* regulons that have been described in *C. acetobutylicum and C. beijerinckii* seem to be conserved in *C. ljungdahlii* and *C. autoethanogenum.* Even so, these results are still not sufficient to claim that the regulatory network governing sporulation is identical in both acetogens and solventogenic clostridia.

Degeneration

In Chapter 4, strain degeneration was addressed by characterizing a degenerated Δ*spoIIE* mutant. Strain degeneration is defined as the loss of desirable traits (270) and characterized

in solventogenic clostridia by the inability to sporulate and produce solvent (155). Despite their inability to form solvents, recent studies proved that degenerated strains can still be attractive for biotechnological applications. Indeed, several studies highlighted the added value of some degenerated strains for organic acid, ester production and H_2 production (132, 273, 418).

As illustrated in Figure 5, colonies of degenerated cells are distinguishable by their irregular and diffuse morphology, and no spores are visible in cultures of degenerated strains. In contrast to other studies on strain degeneration, which studied strains that lost both the capacity to sporulate and produce solvent due to a degeneration event, degeneration only led to the loss of its ability to produce solvent. In contrast to previous reports, cell morphology of degenerated strains seems to be solely linked to the interruption of sporulation before asymmetric division, as shown in Chapter 4. Still, as reported in another study on strain degeneration in *C. beijerinckii* (124), degeneration is reflected on the transcriptome level by a significant change in the expression of genes of unknown function and the genes involved in transcription. In contrast, the phenotype of this degenerated Δ*spoIIE* mutant was not concomitant with changes in the expression of *sigH* and *sig54*, major players in the transition of the cultures from exponential to stationary phase. Other sigma factors, extracytoplasmic function (ECF) sigma factors seemed to be involved in the reorganization of the transcriptome in the degenerated Δ*spoIIE* strain.

Strain degeneration seems to be a phenotype caused by various genotypes; still, those genotypes might correspond to the same proteome or metabolome. Thus it would be interesting to investigate strain degeneration at these levels through a library of degenerated mutants. Also, reports on the genetic causes and ways to reverse strain degeneration were made (81, 123); but no study tested these methods on degenerated strains harboring different genotypes. Finally, studies investigating strain degeneration should characterize in-depth strains reported to be resistant to strain degeneration (419, 420); a common genotype might cause this phenotype.

BMC formation

6

The formation of BMCs was investigated through comparative genomics and transcriptomics in Chapter 5. Formation of BMCs can be observed when *C. beijerinckii* cultures are grown in media containing rhamnose as carbon source. Concomitantly, rhamnose induces the formation of 1,2 propanediol and the enzymes present in the BMCs catalyze the conversion of 1,2 propanediol into propionate and propanol. This reaction generates a toxic intermediate, propionaldehyde, expected to be contained inside the BMC, contrary to propionate and propionyl phosphate. The genome of *C. beijerinckii* was compared to that of *Escherichia coli*, *Salmonella typhimurium* and *C. phytofermentans* (271, 304, 319), microorganisms known to form BMC when grown on rhamnose, to

identify the genes responsible for this phenotype. The gene cluster encoding BMC proteins was identified, and its role in BMC formation was confirmed by the up-regulation of this cluster when the cultures consumed rhamnose as substrate.

Interestingly, comparative genomics showed that a gene cluster similar to the one described in *C. beijerinckii* was found only in *C. tetanomorphum* but not in other solventogenic clostridia. In contrast, BMC clusters were identified in the genome of three solventogenic clostridia, namely *C. ljungdahlii* (312)*, C. autoethanogenum* and *C. carboxidivorans*. However, the BMC clusters detected in *C. ljungdahlii* differ in size and composition compared to the one described in *C. beijerinckii* (312). Moreover, according to a recent study, it seems differently regulated. A transcriptomics study of *C. ljungdahlii* cultures grown in media containing fructose displayed the same profile described in Chapter 5 for rhamnose grown *C. beijerinckii* cultures. Similarly, the BMCs seem to sequester propionaldehyde and enable fructose conversion into propanol and propionate (415).

Biofilm formation

A biofilm is a microbial community consisting of cells lodged in a matrix of extracellular polymeric substances (EPS). Biofilms enable the cells to adhere to various surfaces or interfaces and promote colonization. In several bacterial species, a biofilm is formed in response to environmental changes such as a decrease in available nutrients. Biofilms of *Escherichia coli* (421) and *Saccharomyces cerevisiae* (422) were studied and used for process improvement. Biofilms are also formed by *Clostridium* species, where they contribute to virulence in *C. perfringens* and *Clostridioides* (former *Clostridium*) *difficile* (423). Biofilms are differentiated into mono-species or multi-species depending on the number of species present in the biofilm. In the case of biofilm formation during ABE fermentation, only mono-species biofilms are studied. In solventogenic clostridia, biofilm formation was reported and studied only in *C. acetobutylicum*, for which biofilm prevented the sporulation (406). The cells in the biofilm stayed in a vegetative-like state throughout the fermentation, and few spores were observed. Biofilm-based fermentation increased solvent productivity and solvent tolerance in continuous cultures (424).

In batch cultures, the fermentation profile of biofilm-based cultures differed from planktonic cultures (425) in that butyrate reassimilation and butanol production started earlier and occurred at a faster pace in biofilm-based cultures. In this study, biofilm cells were also studied through transcriptomics; Liu et al. (425) compared the transcriptome of biofilm and planktonic cells and observed a significant change in the amino acid metabolism and iron and sulfur transport. Moreover, Fe-S-containing proteins have a central role in the ABE metabolic pathway. Thus these results could be linked to the reported increase in solvent productivity in biofilm-based cultures. Interestingly biofilm formation was not accompanied by a change in *spo0A* expression, and despite the reported

6

decrease in sporulation frequency in biofilm-based fermentation, no down-regulation of the sporulation genes was detected. Instead, an up-regulation of sporulation genes was observed in the late solventogenic phase. Still, recently a study showed that the disruption of *spo0A* impaired the strain ability to form biofilms, suggesting that Spo0A still plays a role in biofilm formation (134). In *B. subtilis,* genes involved in Spo0A activation were proven to regulate biofilm formation (426). Still, more investigations need to be carried out on biofilm formation in solventogenic clostridia regulation to identify triggers and regulatory networks. As for acetogenic clostridia, mono-species biofilms of *C. ljungdahlii* were also described, and recently, sodium chloride was identified as a trigger of biofilm formation (414).

Changes observed in co-cultures

While most studies on biofilm formation of clostridia were realized on mono-cultures, a recent study investigated the production of H_2 in a biofilm-based fermentation with *C. acetobutylicum* and *Desulfovibrio vulgaris* (427). As reported for mono-species biofilms, this method immobilized the cells and decreased the amount of washed-out cells. Another phenomenon might be involved in the stability of the biomass in the reactor. Indeed, when grown in co-cultures, in a 1:1 ratio, *C. acetobutylicum* and *D. vulgaris* cells fuse their membrane to exchange metabolites (428). This phenomenon was reported in a 1:3 ratio co-culture of *C. acetobutylicum* and *C. ljungdahlii,* and next to metabolites, also RNA was exchanged (394, 429, 430). Interestingly, this phenomenon has not been reported with *C. acetobutylicum* in other co-cultures (431–433) or other solventogenic clostridia. Still, there is a common feature in both co-cultures; the concentration of glucose is above the concentration tolerated by the partner microorganism and is consumed only by *C. acetobutylicum*. As suggested by Benomar et al. (428), microbial fusion might be induced by environmental stress in conditions in which the fusion is necessary for survival.

Open questions and future research

Most studies on solventogenic clostridia have been realized with *C. acetobutylicum*, but with the advances in gene engineering, more strains have become genetically accessible. Moreover, with more robust techniques to study clostridia, several questions raised by the work in this thesis might be answered soon.

6

On sporulation

Despite being a major cellular event, more knowledge on its trigger needs to be gathered to improve media composition. Also, it would be interesting to evaluate the energetical cost of sporulation for the cell. Sporulation involves numerous proteins, and thus, it is considered an energetically costly process. Granulose has been described as an energy

source for sporulation. However, granulose does not seem required for sporulation since some *Clostridium* species do not form granulose but still sporulate. It would also be interesting to investigate the connections between granulose formation, solvent production and sporulation. This knowledge could give clues for engineering more solvent productive strains.

On regulatory networks

Research on solventogenic clostridia has been focusing mainly on the description of phenotypes observed during batch and chemostat fermentation (except for sporulation, for which the transcriptional regulation was studied). The research focus was expanded recently to transcriptional regulation and quorum-sensing. Recent studies indicate that secondary metabolites are also involved in regulating several cellular events (82–84, 92). Many transcriptional (sRNA, σ^{54}, UTR) and post-translational regulators, mainly histidine kinases, were identified, but the interaction between the different members of the regulation networks remains to be investigated.

Signal transduction

First receptors (two-component and RRNPP systems) were identified, but their interaction with other proteins is unknown. Orphan kinases were proven to interact directly with Spo0A in *C. acetobutylicum,* but no study investigated other receptor proteins. Moreover, few studies have been carried out on other signaling systems, like chemotaxis and ECF sigma factors. Those alternative sigma factors seem to be involved in sporulation (130, 286) and granulose degradation (286), and according to the data presented in Chapter 4, they might cause the phenotype of degenerated strains. Moreover, the role of second messengers remains to be investigated. In *B. subtilis* and *C. difficile*, cyclic di-AMP (434) and cyclic di-GMP (435) were crucial for regulating stationary phase events. According to the results presented in Chapter 3, cyclic di-GMP seems to be involved in regulating sporulation in *C. beijerinckii*. Similarly, Chapter 4 calls into question the role of cyclic di-AMP in stress response and strain degeneration.

Post-translational modifications (PTM)

The role of butyryl-phosphate (ButP) in the onset of sporulation and solvent production has been partially answered by Xu et al. (71) with the butyrylation of Spo0A and Buk. Still, the consequence of butyrylation on other essential proteins (Ptb, Adc, BdhA) has not been investigated yet. It also remains to be known whether ButP also serves as a phosphate donor for the phosphorylation of Spo0A, as suggested in earlier studies (70, 71). The role of PTMs in the regulation of cellular events needs to be further investigated. For instance, it would be interesting to assess the impact of stress factors on the PTM. Investigating whether the phenotype of overproducing strains like *C. beijerinckii* BA 101 can be liked to changes in the phosphorylation or acylation of the enzymes catalyzing solvent production would also

be interesting. Phosphorylation was studied in *C. acetobutylicum* (383), but some residues that are often phosphorylated in signal transduction pathways, like serine/threonine, histidine (critical in Spo0A activation c) and asparagine (436), could not be detected because of the harsh treatment required for the analysis. However, with current proteomics techniques, more phosphorylation patterns can be studied. Studies on other PTM such as glycosylation, pupylation, lipidation in solventogenic clostridia would also be interesting; they were shown to be involved in sporulation in *B. subtilis* (437, 438) and *C. difficile* (439).

Conclusion

The work presented in this thesis is part of a global trend towards combinatorial high throughput approaches to investigate cell physiology. Through this approach, we obtained new insight into the regulation of three physiological changes occurring during the cultivation of solventogenic clostridia (sporulation, strain degeneration and BMC formation).

With the development of screening, phenotyping, and genome engineering technologies, the amount of data collected builds up, creating new research fields and deepening our knowledge on non-model organisms and microbial communities. The next revolution in biology will be driven by the integration of artificial intelligence in research approaches. Artificial intelligence is already used to design *in vitro* experiments like enzyme engineering. In biotechnology, the next steps will be to use artificial intelligence to design *in vivo* experiments and create microbial factories able to outcompete petrochemical processes. The next decade is going to be exciting as our understanding of biology and nature will expand tremendously, and together new solutions to solve global health and environmental issues will emerge.

6

Appendices

- English Summary

- References

- List of publications

- Overview of completed training activities

- About the author

- Acknowledgment

Summary

The growing global population and the rising middle class in developing countries drive the demand for consumer goods. However, the manufacture of these products is entirely resting on the petrochemical industry. Since fossil fuel resources are limited, there is a great need for sustainable alternatives. In the search for alternatives to petrochemicals, the interest in bioprocesses is growing. The production of the solvents acetone, butanol and ethanol (ABE) by fermentation of renewable resources is one of the bioprocesses being explored currently by industry and academics. The ABE fermentation converts a wide range of carbon sources such as organic wastes and complex sugars to solvents by natural solvent producing bacteria, called solventogenic clostridia.

Clostridia are gram-positive anaerobic bacteria that form spores. The *Clostridium* genus harbors around 250 species, among which 10 species can utilize sugars to produce organic solvents such as acetone, isopropanol, ethanol or butanol. These chemicals are commonly sold as paint thinners, glue additives or perfumes. Even so, the ABE fermentation process is not cost-competitive with the petrochemical processes currently used to manufacture these products. Therefore, researchers are looking for ways to improve the process and the strains used. One of the approaches investigated is strain engineering. However, the regulation and triggers of the main physiological traits (sporulation and solvent production) are not fully understood. A better knowledge of the physiology of solventogenic clostridia is necessary to generate more productive strains. This thesis describes the research on the physiology of *Clostridium beijerinckii,* which is the most studied solventogenic *Clostridium* after *Clostridium acetobutylicum*. Two intriguing physiological processes were tackled: firstly, the regulation of the sporulation, which is intimately coupled to solventogenesis and secondly, the metabolism of L-rhamnose, which involves the generation of particular cellular compartments called bacterial microcompartments (BMC).

Sporulation is a cell differentiation process that occurs in all Firmicutes. It is triggered when the cells are exposed to an unfavorable environmental condition (presence of oxygen, high pH variation). The spores are metabolically inactive and resistant to harsh conditions (UV, chemicals, heat, oxygen). Sporulation is seen as a hindrance, in industry, for several reasons: it occurs during the fermentation, stops the solvent production, and is energy costly for the cells. The regulation of sporulation and its connection to solvent production is still not elucidated. In **Chapter 1**, the research done in the last 30 years on sporulation in solventogenic clostridia is reviewed, and the latest updates on the molecular regulation of sporulation in solventogenic clostridia are presented. Common triggers and regulation mechanisms of sporulation in several solventogenic species are identified, underlining the differences and similarities between the various species. Potential links in the regulation of sporulation and solvent production are pinpointed. Around fifty non-sporulating mutant strains have been reported and studied in the literature, with, for most of these strains,

differences in solvent productivity compared to the wild-type strain. This chapter highlights the need for more studies on signal transduction pathways, transcriptional and post-translational regulation of sporulation, and solvent production. A better understanding of the connections between both physiological phenomena would provide us with relevant targets for strain engineering of solventogenic clostridia.

Despite being studied since the 1920s, many unknowns remain in our understanding of the physiology of clostridia. These gaps in the knowledge can be linked to the lack of efficient genome engineering tools for clostridia. Recent developments in CRISPR technologies have opened new possibilities for developing and improving genome editing tools dedicated to the *Clostridium* genus. A two-plasmid CRISPR genome engineering system is described in **Chapter 2**. This tool enabled scarless modification of the genome of two reference strains of *Clostridium beijerinckii*, the ABE-producing strain NCIMB 8052 and the natural IBE-producing strain DSM 6423. As a proof of principle, we performed gene deletions in both strains and even cured the endogenous pNF2 plasmid of the DSM 6423 strain. In the NCIMB 8052 strain, the disruption of the *spoIIE* coding sequence resulted in sporulation-deficient mutants. This phenotype was reverted by complementing the mutant strain with a functional *spoIIE* gene. The fungal cellulase-encoding *celA* gene was inserted into the *C. beijerinckii* NCIMB 8052 chromosome, resulting in mutants with endoglucanase activity. A similar approach was employed to engineer the genome of *C. beijerinckii* DSM 6423. The *catB* gene conferring thiamphenicol resistance was deleted with a single plasmid CRISPR tool to make this strain compatible with our dual-plasmid editing system. The dual-plasmid system was then used in *C. beijerinckii* DSM 6423 ΔcatB to remove the endogenous pNF2 plasmid, which increased transformation efficiencies.

The phenotype of the *spoIIE* mutant generated in *C. beijerinckii* NCIMB 8052, with the CRISPR-Cas9 tool described in Chapter 2, was further studied in **Chapter 3**. SpoIIE is a key sporulation factor involved in the first stage of sporulation. This phosphatase activates the first sporulation specific sigma factor, σ^F, and is involved in septum formation during asymmetric division. The Δ*spoIIE* mutant generated did not sporulate but still produced granulose and solvents. Moreover, the mutant cell morphology differed from the wild-type's; Δ*spoIIE* mutant cells are elongated with the presence of multiple septa. These observations suggest that in *C. beijerinckii*, SpoIIE is necessary to complete the sporulation process, as seen in *Bacillus* and *C. acetobutylicum*. Fermentation experiments in reactors showed that the *spoIIE* mutant produced a higher solvent titer than the wild-type strain. The impact of *spoIIE* inactivation was further assessed by comparative transcriptome analysis. As predicted, the genes belonging to the regulon of the sporulation specific transcription factors (σ^F, σ^E, σ^G, σ^K) were strongly down-regulated in the mutant. Nevertheless, the inactivation of *spoIIE* also caused differential expression of genes involved in various cellular processes. For instance, genes involved in butanol formation

and tolerance and cell motility were up-regulated in the mutant, while several genes involved in cell wall composition, oxidative stress and amino acid transport were down-regulated.

Five *spoIIE* deficient strains were obtained using the tool described in Chapter 2, and one of them, *C. beijerinckii* Δ*spoIIE*C2, did not produce solvents. The loss of solvent production combined with the loss of sporulation is characteristic of degenerated strains. Strain degeneration is observed in solventogenic clostridia after long cultivation periods. No consensus on the location, number, or type of mutations causing degeneration exists. The physiology of *C. beijerinckii* Δ*spoIIE*C2 was studied in **Chapter 4** using fermentation experiments and microscopy observations, as well as genome and transcriptome analysis to investigate the causes and consequences of this loss of solvent-producing capacity. Its physiology was compared to the wild-type *C. beijerinckii* strain, the solventogenic Δ*spoIIE*C5 strain, and degenerated *C. beijerinckii* strains described in literature. Compared to wild-type and other Δ*spoIIE*, biomass production in Δ*spoIIE*C2 was reduced and acid yields increased, as reported before for degenerated *C. beijerinckii* strains. Genome comparison of five *C. beijerinckii* Δ*spoIIE* strains revealed that the genome of Δ*spoIIE*C2 harbored seven point mutations, out of which six were located in coding sequences. Transcriptomic analysis was carried out by comparing Δ*spoIIE*C2 to the solventogenic mutant Δ*spoIIE*C5 at early exponential phase and entry into the stationary phase. The expression profile of 35 % of the genome was differentially expressed at least at one time point and most of these genes belonged (over 40%) to the COG of unknown function represented. Two mutated genes were differentially expressed, *cbei_0201* at both time points and *cbei_1712* encoding Spo0A only at the entry into stationary phase, indicating that those mutations might have caused the degeneration. Strain degeneration seems to involve several signaling pathways as genes encoding extracytoplasmic function sigma factors, quorum-sensing and chemotaxis proteins were up-regulated. Simultaneously, genes encoding carbohydrate metabolism were down-regulated at the studied time points, which corresponds to the slow growth and low sugar consumption observed during fermentation.

In **Chapter 5**, the physiology of *C. beijerinckii* was studied when grown with **L**-rhamnose as a carbon source. **L**-Rhamnose is a sugar abundant in green macroalgae (or seaweed), which represents a promising alternative to first and second-generation feedstocks for the production of renewable fuels and chemicals. We could show that *C. beijerinckii* DSM 6423 ferments **L**-rhamnose and **D**-glucose, the main sugars in hydrolysates of the green macroalgae *Ulva lactuca*, to acetic acid, butyric acid, isopropanol, butanol, and ethanol (IBE), and 1,2-propanediol. On **L**-rhamnose, the cultures displayed poor growth and low sugar consumption but produced 1,2-propanediol, propionic acid, and n-propanol in addition to acetic and butyric acids. In contrast, on **D**-glucose, IBE was the main product.

On a **D**-glucose– **L**-rhamnose mixture, both sugars were converted simultaneously and **L**-rhamnose consumption increased, leading to higher levels of 1,2-propanediol, butanol and isopropanol. Genome and transcriptomics analysis of **D**-glucose and **L**-rhamnose-grown cells revealed all genes and their transcripts involved in **L**-rhamnose utilization and those needed to form BMCs. These data provide insights into the metabolic pathway of **L**-rhamnose utilization and other physiological changes induced by growth on **L**-rhamnose (glycolysis, early sporulation, and stress response).

The work presented in the thesis and its contribution to the *Clostridium* research field is discussed in **Chapter 6**. An overview of the current knowledge on the morphological changes observed in cultures of solventogenic and acetogenic clostridia, another industrially relevant *Clostridium* group. Besides, an outlook is given on complementary research needed to increase our knowledge of the regulation of sporulation and solvent production. A better grasp of the complex regulation of the physiological changes occurring during fermentation is required to improve the ABE process further. Even though numerous fermentation studies with different feedstocks have been performed, more studies on the effects of these feedstocks on the transcriptome and proteome need to be conducted. With the expansion of the genome engineering toolbox for clostridia and the development of -omics studies, our knowledge of clostridial physiology is expected to grow exponentially in the next decades.

References

1.   European Union. 2020. EU Climate Target Plan 2030: Building a modern, sustainable.
2.   Tracy BP, Jones SW, Fast AG, Indurthi DC, Papoutsakis ET. 2012. Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. Curr Opin Biotechnol 23:364–381.
3.   Dürre P. 2005. Handbook on clostridia. CRC Press.
4.   Jones DT, Woods DR. 1986. Acetone-butanol fermentation revisited. Microbiol Rev 50:484–524.
5.   Berezina O, Zakharova N, Yarotsky C, Zverlov V. 2012. Microbial producers of butanol. Appl Biochem Microbiol 48:625–638.
6.   Sauer M. 2016. Industrial production of acetone and butanol by fermentation-100 years later. FEMS Microbiol Lett 363.
7.   Xin F, Yan W, Zhou J, Wu H, Dong W, Ma J, Zhang W, Jiang M. 2018. Exploitation of novel wild-type solventogenic strains for butanol production. Biotechnol Biofuels 11:252.
8.   Green EM. 2011. Fermentative production of butanol—the industrial perspective. Curr Opin Biotechnol 22:337–343.
9.   Tashiro Y, Yoshida T, Noguchi T, Sonomoto K. 2013. Recent advances and future prospects for increased butanol production by acetone-butanol-ethanol fermentation. Eng Life Sci 13:432–445.
10.   Vees CA, Neuendorf CS, Pflügl S. 2020. Towards continuous industrial bioprocessing with solventogenic and acetogenic clostridia: challenges, progress and perspectives. J Ind Microbiol Biotechnol 1–35.
11.   Qureshi N, Blaschek HP. 1999. Production of Acetone Butanol Ethanol (ABE) by a hyper-producing mutant strain of *Clostridium beijerinckii* BA101 and recovery by pervaporation. Biotechnol Prog 15:594–602.
12.   Liu XB, Gu QY, Yu X Bin. 2013. Repetitive domestication to enhance butanol tolerance and production in *Clostridium acetobutylicum* through artificial simulation of bio-evolution. Bioresour Technol 130:638–643.
13.   Papoutsakis ET. 2008. Engineering solventogenic clostridia. Curr Opin Biotechnol 19:420–9.
14.   Cheng C, Bao T, Yang ST. 2019. Engineering *Clostridium* for improved solvent production: recent progress and perspective. Appl Microbiol Biotechnol 103:5549–5566.
15.   Scotcher MC, Bennett GN. 2005. SpoIIE regulates sporulation but does not directly affect solventogenesis in *Clostridium acetobutylicum* ATCC 824. J Bacteriol 187:1930–1936.
16.   Bi C, Jones SW, Hess DR, Tracy MBP, Papoutsakis ET. 2011. SpoIIe Is necessary for asymmetric division, sporulation, and expression of σ^F, σ^E, and σ^G but does not control solvent production in *Clostridium acetobutylicum* ATCC 824. J Bacteriol 193:5130–5137.
17.   Al-Hinai MA, Jones SW, Papoutsakis ET. 2014. σ^K of *Clostridium acetobutylicum* is the first known sporulation-specific sigma factor with two developmentally separated roles, one early and one late in sporulation. J Bacteriol 196:287–99.
18.   Patakova P, Linhova M, Rychtera M, Paulova L, Melzoch K. 2013. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. Biotechnol Adv 31:58–67.
19.   Joseph RC, Kim NM, Sandoval NR. 2018. Recent developments of the synthetic biology toolkit for *Clostridium*. Front Microbiol 9:154.
20.   Paul C, Filippidou S, Jamil I, Kooli W, House GL, Estoppey A, Hayoz M, Junier T, Palmieri F, Wunderlin T, Lehmann A, Bindschedler S, Vennemann T, Chain PSG, Junier P. 2019. Bacterial spores, from ecology to biotechnology. Adv Appl Microbiol 106:79–111.
21.   Dürre P. 2014. Physiology and sporulation in *Clostridium*, p. 315–329. *In* The Bacterial Spore: from Molecules to Systems. American Society of Microbiology.
22.   Johnson EA. 2019. Clostridia. Encycl Microbiol 690–695.
23.   Al-Hinai MA, Jones SW, Papoutsakis ET. 2015. The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. Microbiol Mol Biol Rev 79:19–37.
24.   Collins M, Lawson P, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H, Farrow J. 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol 44:812–826.

References

25. Keis S, Bennett CF, Ward VK, Jones DT. 1995. Taxonomy and phylogeny of industrial solvent-producing clostridia. Int J Syst Bacteriol 45:693–705.
26. Keis S, Shaheen R, Jones DT. 2001. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp nov and *Clostridium saccharobutylicum* sp nov. Int J Syst Evol Microbiol 51:2095–2103.
27. Yu HY, Meade A, Liu SJ. 2019. Phylogeny of *Clostridium* spp. based on conservative genes and comparisons with other trees. Microbiology 88:469–478.
28. Cruz-Morales P, Orellana CA, Moutafis G, Moonen G, Rincon G, Nielsen LK, Marcellin E, Bapteste E. 2019. Revisiting the evolution and taxonomy of clostridia, a phylogenomic update. Genome Biol Evol 11:2035–2044.
29. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fèvre F, Longin C, Mornico D, Roche D, Rouy Z, Salvignol G, Scarpelli C, Smith AAT, Weiman M, Médigue C. 2013. MicroScope - An integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res 41:636–647.
30. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132.
31. Basu A, Xin F, Lim TK, Lin Q, Yang KL, He J. 2017. Quantitative proteome profiles help reveal efficient xylose utilization mechanisms in solventogenic *Clostridium* sp. strain BOH3. Biotechnol Bioeng 114:1959–1969.
32. Awang GM, Ingledew WM, Jones GA. 1992. The effect of fermentable carbohydrate on sporulation and butanol production by *Clostridium acetobutylicum* P262. Appl Microbiol Biotechnol 38:12–16.
33. Woods DR, Jones DT. 1987. Physiological responses of bacteroides and *Clostridium* strains to environmental stress factors. Adv Microb Physiol 28:1–64.
34. Sauer U, Santangelo JD, Treuner A, Buchholz M, Dürre P. 1995. Sigma factor and sporulation genes in *Clostridium* . FEMS Microbiol Rev 17:331–340.
35. Long S, Jones DT, Woods DR. 1983. Sporulation of *Clostridium acetobutylicum* P262 in a Defined Medium. Appl Environ Microbiol 45:1389–1393.
36. Kolek J, Diallo M, Vasylkivska M, Branska B, Sedlar K, López-Contreras AM, Patakova P. 2017. Comparison of expression of key sporulation, solventogenic and acetogenic genes in *C. beijerinckii* NRRL B-598 and its mutant strain overexpressing *spo0A*. Appl Microbiol Biotechnol. Springer Berlin Heidelberg.
37. Shaheen R, Shirley M, Jones DT. 2000. Comparative fermentation studies of industrial strains belonging to four species of solvent-producing clostridia. J Mol Microbiol Biotechnol 2:115–124.
38. Diallo M, Simons AD, van der Wal H, Collas F, Houweling-Tan B, Kengen SWM, López-Contreras AM. 2018. L-Rhamnose Metabolism in *Clostridium beijerinckii* Strain DSM 6423. Appl Environ Microbiol 85.
39. Long S, Jones D, Woods D. 1984. Initiation of solvent production, clostridial stage and endospore formation in *Clostridium acetobutylicum* P262. Appl Microbiol Biotechnol 20.
40. Varga J, Stirewalt VL, Melville SB. 2004. The CcpA protein is necessary for efficient sporulation and enterotoxin gene (*cpe*) regulation in *Clostridium perfringens*. J Bacteriol 186:5221–5229.
41. Antunes A, Martin-Verstraete I, Dupuy B. 2011. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. Mol Microbiol 79:882–899.
42. Ren C, Gu Y, Wu Y, Zhang W, Yang C, Yang S, Jiang W. 2012. Pleiotropic functions of catabolite control protein CcpA in Butanol-producing *Clostridium acetobutylicum*. BMC Genomics 13:349.
43. Lee S, Lee JH, Mitchell RJ. 2015. Analysis of *Clostridium beijerinckii* NCIMB 8052's transcriptional response to ferulic acid and its application to enhance the strain tolerance. Biotechnol Biofuels 8:68.
44. Ezeji T, Qureshi N, Blaschek HP. 2007. Butanol production from agricultural residues: Impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. Biotechnol Bioeng 97:1460–1469.
45. Liu H, Zhang J, Yuan J, Jiang X, Jiang L, Li Z, Yin Z, Du Y, Zhao G, Liu B, Huang D. 2020. Gene coexpression network analysis reveals a novel metabolic mechanism of *Clostridium acetobutylicum* responding to phenolic inhibitors from lignocellulosic hydrolysates. Biotechnol Biofuels 13:163.
46. Reeve BWP, Reid SJ. 2016. Glutamate and histidine improve both solvent yields and the acid tolerance response of *Clostridium beijerinckii* NCP 260. J Appl Microbiol 120:1271–1281.

47. Nimbalkar PR, Khedkar MA, Parulekar RS, Chandgude VK, Sonawane KD, Chavan P V., Bankar SB. 2018. Role of trace elements as cofactor: an efficient strategy toward enhanced biobutanol production. ACS Sustain Chem Eng 6:9304–9313.

48. Nimbalkar PR, Khedkar MA, Chavan P V., Bankar SB. 2019. Enhanced biobutanol production in folic acid-induced medium by using *Clostridium acetobutylicum* NRRL B-527. ACS Omega 4:12978–12982.

49. List C, Hosseini Z, Lederballe Meibom K, Hatzimanikatis V, Bernier-Latmani R. 2019. Impact of iron reduction on the metabolism of *Clostridium acetobutylicum*. Environ Microbiol 21:3548–3563.

50. Mukherjee M, Sarkar P, Goswami G, Das D. 2019. Regulation of butanol biosynthesis in *Clostridium acetobutylicum* ATCC 824 under the influence of zinc supplementation and magnesium starvation. Enzyme Microb Technol 129:109352.

51. Kiyoshi K, Kawashima S, Nobuki K, Kadokura T, Nakazato A, Suzuki KI, Nakayama S. 2017. Adenine addition restores cell viability and butanol production in *Clostridium saccharoperbutylacetonicum* N1-4 (ATCC 13564) cultivated at 37°C. Appl Environ Microbiol 83:e02960-16.

52. Lee KY, Juang TC, Lee KC. 1978. Effect of metal ions on growth and sporulation of *Clostridium perfringens* in a synthetic medium. Chinese J Microbiol Immunol 11:50–61.

53. Mah JH, Kang DH, Tang J. 2008. Effects of minerals on sporulation and heat resistance of *Clostridium sporogenes*. Int J Food Microbiol 128:385–389.

54. Martínez-Lumbreras S, Alfano C, Evans NJ, Collins KM, Flanagan KA, Atkinson RA, Krysztofinska EM, Vydyanath A, Jackter J, Fixon-Owoo S, Camp AH, Isaacson RL. 2018. Structural and Functional insights into *Bacillus subtilis* sigma factor inhibitor, CsfB. Structure 26:640-648.e5.

55. Król E, De Sousa Borges A, Kopacz M, Scheffers DJ. 2017. Metal-dependent SpoIIE oligomerization stabilizes FtsZ during asymmetric division in *Bacillus subtilis*. PLoS One 12:e0174713.

56. Huang SS, Chen D, Pelczar PL, Vepachedu VR, Setlow P, Li YQ. 2007. Levels of Ca^{2+}-dipicolinic acid in individual *Bacillus* spores determined using microfluidic Raman tweezers. J Bacteriol 189:4681–4687.

57. Paredes-Sabja D, Setlow B, Setlow P, Sarker MR. 2008. Characterization of *Clostridium perfringens* spores that lack SpoVA proteins and dipicolinic acid. J Bacteriol 190:4648–4659.

58. Jamroskovic J, Chromikova Z, List C, Bartova B, Barák I, Bernier-Latmani R. 2016. Variability in DPA and calcium content in the spores of *Clostridium* species. Front Microbiol 7:1791.

59. Vasylkivska M, Jureckova K, Branska B, Sedlar K, Kolek J, Provaznik I, Patakova P. 2019. Transcriptional analysis of amino acid, metal ion, vitamin and carbohydrate uptake in butanol-producing *Clostridium beijerinckii* NRRL B-598. PLoS One 14:e0224560.

60. Diallo M, Kint N, Monot M, Collas F, Martin-Verstraete I, van der Oost J, Kengen SWM, López-Contreras AM. 2020. Transcriptomic and phenotypic analysis of a *spoIIE* mutant in *Clostridium beijerinckii*. Front Microbiol 11.

61. Zheng Y-N, Li L-Z, Xian M, Ma Y-J, Yang J-M, Xu X, He D-Z. 2009. Problems with the microbial production of butanol. J Ind Microbiol Biotechnol 36:1127–1138.

62. Assobhei O, Kanouni A El, Ismaili M, Loutfi M, Petitdemange H. 1998. Effect of acetic and butyric acids on the stability of solvent and spore formation by *Clostridium acetobutylicum* ATCC 824 during repeated subculturing. J Ferment Bioeng 85:209–212.

63. Thorn GJ, King JR, Jabbari S. 2013. pH-induced gene regulation of solvent production by *Clostridium acetobutylicum* in continuous culture: parameter estimation and sporulation modelling. Math Biosci 241:149–166.

64. Terracciano JS, Kashket ER. 1986. Intracellular conditions required for initiation of solvent production by *Clostridium acetobutylicum*. Appl Environ Microbiol 52:86–91.

65. Yang X, Tu M, Xie R, Adhikari S, Tong Z. 2013. A comparison of three pH control methods for revealing effects of undissociated butyric acid on specific butanol production rate in batch fermentation of *Clostridium acetobutylicum*. AMB Express 3:3.

66. Desai RP, Papoutsakis ET. 1999. Antisense RNA strategies for metabolic engineering of *Clostridium acetobutylicum*. Appl Environ Microbiol 65:936–945.

67. Harris LM, Desai RP, Welker NE, Papoutsakis ET. 2000. Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: need for new phenomenological models for solventogenesis and butanol inhibition? Biotechnol Bioeng 67:1–11.

References

68. Kuit W, Minton NP, López-Contreras AM, Eggink G. 2012. Disruption of the acetate kinase (*ack*) gene of *Clostridium acetobutylicum* results in delayed acetate production. Appl Microbiol Biotechnol 94:729–41.

69. Green EM, Boynton ZL, Harris LM, Rudolph FB, Papoutsakis ET, Bennett GN. 1996. Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. Microbiology 142:2079–2086.

70. Zhao Y, Tomas CA, Rudolph FB, Papoutsakis ET, Bennett GN. 2005. Intracellular butyryl phosphate and acetyl phosphate concentrations in *Clostridium acetobutylicum* and their implications for solvent formation. Appl Environ Microbiol 71:530–537.

71. Xu J-Y, Xu Z, Liu X, Tan M, Ye B-C. 2018. Protein acetylation and butyrylation regulate the phenotype and metabolic shifts of the endospore-forming *Clostridium acetobutylicum*. Mol Cell Proteomics 17:1156–1169.

72. Kuit W. 2013. Metabolic Engineering of acid formation in *Clostridium acetobutylicum* Wouter Kuit. Wageningen University.

73. Macek B, Forchhammer K, Hardouin J, Weber-Ban E, Grangeasse C, Mijakovic I. 2019. Protein post-translational modifications in bacteria. Nat Rev Microbiol 17:651–664.

74. Sedlar K, Koscova P, Vasylkivska M, Branska B, Kolek J, Kupkova K, Patakova P, Provaznik I. 2018. Transcription profiling of butanol producer *Clostridium beijerinckii* NRRL B-598 using RNA-Seq. BMC Genomics 19:415.

75. Sedlar K, Kolek J, Gruber M, Jureckova K, Branska B, Csaba G, Vasylkivska M, Zimmer R, Patakova P, Provaznik I. 2019. A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock. Biotechnol Biofuels 12:243.

76. Tomas CA, Beamish J, Papoutsakis ET. 2004. Transcriptional Analysis of Butanol Stress and Tolerance in *Clostridium acetobutylicum*. J Bacteriol 186:2006–2018.

77. Patakova P, Branska B, Sedlar K, Vasylkivska M, Jureckova K, Kolek J, Koscova P, Provaznik I. 2019. Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level. Sci Rep 9:1371.

78. Schwarz KM, Kuit W, Grimmler C, Ehrenreich A, Kengen SWM. 2012. A transcriptional study of acidogenic chemostat cells of *Clostridium acetobutylicum* – Cellular behavior in adaptation to n-butanol. J Biotechnol 161:366–377.

79. Leggett MJ, Mcdonnell G, Denyer SP, Setlow P, Maillard JY. 2012. Bacterial spore structures and their protective role in biocide resistance. J Appl Microbiol 113:485–498.

80. Letzel A-C, Pidot SJ, Hertweck C. 2013. A genomic approach to the cryptic secondary metabolome of the anaerobic world. Nat Prod Rep 30:392–428.

81. Kosaka T, Nakayama S, Nakaya K, Yoshino S, Furukawa K. 2007. Characterization of the sol operon in butanol-hyperproducing *Clostridium saccharoperbutylacetonicum* strain N1-4 and its degeneration mechanism. Biosci Biotechnol Biochem 71:58–68.

82. Li JS, Barber CC, Herman NA, Cai W, Zafrir E, Du Y, Zhu X, Skyrud W, Zhang W. 2020. Investigation of secondary metabolism in the industrial butanol hyper-producer *Clostridium saccharoperbutylacetonicum* N1-4. J Ind Microbiol Biotechnol 47:319–328.

83. Herman NA, Kim SJ, Li JS, Cai W, Koshino H, Zhang W. 2017. The industrial anaerobe *Clostridium acetobutylicum* uses polyketides to regulate cellular differentiation. Nat Commun 8:1514.

84. Chen Y, Yang Y, Ji X, Zhao R, Li G, Gu Y, Shi A, Jiang W, Zhang Q. 2020. The SCIFF-derived ranthipeptides participate in quorum-sensing in solventogenic clostridia. Biotechnol J 15:1–8.

85. Grossman AD, Losick R. 1988. Extracellular control of spore formation in *Bacillus subtilis*. Proc Natl Acad Sci U S A 85:4369–4373.

86. Bischofs IB, Hug JA, Liu AW, Wolf DM, Arkin AP. 2009. Complexity in bacterial cell-cell communication: Quorum signal integration and subpopulation signaling in the *Bacillus subtilis* phosphorelay. Proc Natl Acad Sci 106:6459–6464.

87. Grimmler C, Janssen H, Krauße D, Fischer R-JJ, Bahl H, Dürre P, Liebl W, Ehrenreich A. 2011. Genome-wide gene expression analysis of the switch between acidogenesis and solventogenesis in continuous cultures of *Clostridium acetobutylicum*. J Mol Microbiol Biotechnol 20:1–15.

88. Heluane H, Dagher MRE, Bruno-Bárcena JM, Evans MR, Dagher SF, Bruno-Bárcena JM. 2011. Meta-Analysis and functional validation of nutritional requirements of solventogenic clostridia growing under butanol stress conditions and coutilization of D-glucose and D-xylose. Appl Environ Microbiol 77:4473–4485.

89.   Aframian N, Eldar A. 2020. A Bacterial Tower of Babel: Quorum-sensing signaling diversity and its evolution. Annu Rev Microbiol 74:annurev-micro-012220-063740.

90.   Steiner E, Scott J, Minton NP, Winzer K. 2012. An Agr quorum-sensing system that regulates granulose formation and sporulation in *Clostridium acetobutylicum*. Appl Environ Microbiol 78:1113–1122.

91.   Yang Y, Lang N, Zhang L, Wu H, Jiang W, Gu Y. 2020. A novel regulatory pathway consisting of a two-component system and an ABC-type transporter contributes to butanol tolerance in *Clostridium acetobutylicum*. Appl Microbiol Biotechnol 1–13.

92.   Kotte A-K, Severn O, Bean Z, Schwarz K, Minton NP, Winzer K. 2020. RRNPP-type quorum-sensing affects solvent formation and sporulation in *Clostridium acetobutylicum*. Microbiology 166:579–592.

93.   Feng J, Zong W, Wang P, Zhang ZT, Gu Y, Dougherty M, Borovok I, Wang Y. 2020. RRNPP-Type quorum-sensing systems regulate solvent formation, sporulation and cell motility in *Clostridium saccharoperbutylacetonicum*. Biotechnol Biofuels 13:84.

94.   Edwards AN, Krall EG, McBride SM. 2020. Strain-dependent RstA regulation of *Clostridioides difficile* toxin production and sporulation. J Bacteriol 202.

95.   Piggot PJ, Hilbert DW. 2004. Sporulation of *Bacillus subtilis*. Curr Opin Microbiol 7:579–86.

96.   Decker AR, Ramamurthi KS. 2017. Cell death pathway that monitors spore morphogenesis. Trends Microbiol 25:637–647.

97.   Davidson P, Eutsey R, Redler B, Hiller NL, Laub MT, Durand D. 2018. Flexibility and constraint: Evolutionary remodeling of the sporulation initiation pathway in Firmicutes. PLoS Genet 14:e1007470.

98.   Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ. 2012. Genomic determinants of sporulation in Bacilli and Clostridia : towards the minimal set of sporulation-specific genes. Environ Microbiol 14:2870–2890.

99.   Traag BA, Pugliese A, Eisen JA, Losick R. 2013. Gene conservation among endospore-forming bacteria reveals additional sporulation genes in *Bacillus subtilis*. J Bacteriol 195:253–60.

100.   Galperin MY. 2013. Genome diversity of spore-forming Firmicutes. Microbiol Spectr 1.

101.   Fimlaid KA, Shen A. 2015. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. Curr Opin Microbiol 24:88–95.

102.   Máté de Gérando H, Wasels F, Bisson A, Clement B, Bidard F, Jourdier E, López-Contreras AM, Lopes Ferreira N. 2018. Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423. BMC Genomics 19:242.

103.   Harris LM, Welker NE, Papoutsakis ET. 2002. Northern, morphological, and fermentation analysis of *spo0a* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. J Bacteriol 184:3586–3597.

104.   Ehsaan M, Kuit W, Zhang Y, Cartman ST, Heap JT, Winzer K, Minton NP. 2016. Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824. Biotechnol Biofuels 9:4.

105.   Li Q, Chen J, Minton NP, Zhang Y, Wen Z, Liu J, Yang H, Zeng Z, Ren X, Yang J, Gu Y, Jiang W, Jiang Y, Yang S. 2016. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. Biotechnol J 11:961–972.

106.   Alsaker K V., Spitzer TR, Papoutsakis ET. 2004. Transcriptional Analysis of spo0A Overexpression in *Clostridium acetobutylicum* and Its Effect on the Cell's Response to Butanol Stress. J Bacteriol 186:1959–1971.

107.   Foulquier C, Huang C-N, Nguyen N-P-T, Thiel A, Wilding-Steel T, Soula J, Yoo M, Ehrenreich A, Meynial-Salles I, Liebl W, Soucaille P. 2019. An efficient method for markerless mutant generation by allelic exchange in *Clostridium acetobutylicum* and *Clostridium saccharobutylicum* using suicide vectors. Biotechnol Biofuels 12:31.

108.   Liu Z, Qiao K, Tian L, Zhang Q, Liu Z-Y, Li F-L. 2015. Spontaneous large-scale autolysis in *Clostridium acetobutylicum* contributes to generation of more spores. Front Microbiol 6:950.

109.   Tracy BP, Jones SW, Papoutsakis ET. 2011. Inactivation of σ^E and σ^G in *Clostridium acetobutylicum* illuminates their roles in clostridial-cell-form biogenesis, granulose synthesis, solventogenesis, and spore morphogenesis. J Bacteriol 193:1414–26.

110.   Jones SW, Tracy BP, Gaida SM, Papoutsakis ET. 2011. Inactivation of σ^F in *Clostridium acetobutylicum* ATCC 824 blocks sporulation prior to asymmetric division and abolishes σ^E and σ^G protein expression but does not block solvent formation. J Bacteriol 193:2429–40.

References

111.   Steiner E, Dago AE, Young DI, Heap JT, Minton NP, Hoch JA, Young M. 2011. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. Mol Microbiol 80:641–54.

112.   Xu M, Zhao J, Yu L, Yang S-T. 2017. Comparative genomic analysis of *Clostridium acetobutylicum* for understanding the mutations contributing to enhanced butanol tolerance and production. J Biotechnol 263:36–44.

113.   Scotcher MC, Rudolph FB, Bennett GN. 2005. Expression of *abrB310* and *sinR*, and effects of decreased *abrB310* expression on the transition from acidogenesis to solventogenesis, in *Clostridium acetobutylicum* ATCC 824. Appl Environ Microbiol 71:1987–1995.

114.   Hu S, Zheng H, Gu Y, Zhao J, Zhang W, Yang Y, Wang S, Zhao G, Yang S, Jiang W. 2011. Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018. BMC Genomics 12:93.

115.   Jain MKMK, Beacom D, Datta R, Lansing E. January 1994. Mutant strain of *C. acetobutylicum* and process for making butanol. US5192673A. Biotechnol Adv. U.S. Patent and Trademark Office, USA.

116.   Chen Y, Zhou T, Liu D, Li A, Xu S, Liu Q, Li B, Ying H. 2013. Production of butanol from glucose and xylose with immobilized cells of *Clostridium acetobutylicum*. Biotechnol Bioprocess Eng 18:234–241.

117.   Wilkinson SR, Young M, Goodacre R, Morris JG, Farrow JAE, Collins MD. 1995. Phenotypic and genotypic differences between certain strains of *Clostridium acetobutylicum*. FEMS Microbiol Lett 125:199–204.

118.   Ravagnani A, Jennert KCB, Steiner E, Grünberg R, Jefferies JR, Wilkinson SR, Young DI, Tidswell EC, Brown DP, Youngman P, Gareth Morris J, Young M. 2000. Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. Mol Microbiol 37:1172–1185.

119.   Diallo M, Hocq R, Collas F, Chartier G, Wasels F, Wijaya HS, Werten MWT, Wolbert EJH, Kengen SWM, van der Oost J, Ferreira NL, López-Contreras AM. 2020. Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*. Methods 172:51–60.

120.   Shi Z, Blaschek HP. 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. Appl Environ Microbiol 74:7709–7714.

121.   Sandoval-Espinola WJ, Makwana ST, Chinn MS, Thon MR, Azcárate-Peril MA, Bruno-Bárcena JM. 2013. Comparative phenotypic analysis and genome sequence of *Clostridium beijerinckii* SA-1, an offspring of NCIMB 8052. Microbiology 159:2558–2570.

122.   Seo SO, Janssen H, Magis A, Wang Y, Lu T, Price ND, Jin YS, Blaschek HP. 2017. Genomic, transcriptional, and phenotypic analysis of the glucose derepressed *Clostridium beijerinckii* mutant exhibiting acid crash phenotype. Biotechnol J 12:1700182.

123.   Jiao S, Zhang Y, Wan C, Lv J, Du R, Zhang R, Han B. 2016. Transcriptional analysis of degenerate strain *Clostridium beijerinckii* DG-8052 reveals a pleiotropic response to CaCO3-associated recovery of solvent production. Sci Rep 6:38818.

124.   Zhang Y, Jiao S, Lv J, Du R, Yan X, Wan C, Zhang R, Han B. 2017. Sigma factor regulated cellular response in a non-solvent producing *Clostridium beijerinckii* degenerated strain: A comparative transcriptome analysis. Front Microbiol 8:23.

125.   Xin X, Cheng C, Du G, Chen L, Xue C. 2020. Metabolic engineering of histidine kinases in *Clostridium beijerinckii* for enhanced butanol production. Front Bioeng Biotechnol 8:214.

126.   Hocq R, Bouilloux-Lafont M, Lopes Ferreira N, Wasels F. 2019. σ^{54} (σ^{L}) plays a central role in carbon metabolism in the industrially relevant *Clostridium beijerinckii*. Sci Rep 9:7228.

127.   Sandoval NR, Venkataramanan KP, Groth TS, Papoutsakis ET. 2015. Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth. Biotechnol Biofuels 8:227.

128.   Schwarz KM, Grosse-Honebrink A, Derecka K, Rotta C, Zhang Y, Minton NP. 2017. Towards improved butanol production through targeted genetic modification of *Clostridium pasteurianum*. Metab Eng 40:124–137.

129.   Atmadjaja AN, Holby V, Harding AJ, Krabben P, Smith HK, Jenkinson ER. 2019. CRISPR-Cas, a highly effective tool for genome editing in *Clostridium saccharoperbutylacetonicum* N1-4(HMT). FEMS Microbiol Lett https://doi.org/10.1093/femsle/fnz059.

210

130. Jones SW, Paredes CJ, Tracy B, Cheng N, Sillers R, Senger RS, Papoutsakis ET. 2008. The transcriptional program underlying the physiology of clostridial sporulation. Genome Biol 9:R114.

131. Xue Q, Yang Y, Chen J, Chen L, Yang S, Jiang W, Gu Y. 2016. Roles of three AbrBs in regulating two-phase *Clostridium acetobutylicum* fermentation. Appl Microbiol Biotechnol 100:9081–9089.

132. Seo SO, Wang Y, Lu T, Jin YS, Blaschek HP. 2017. Characterization of a *Clostridium beijerinckii spo0A* mutant and its application for butyl butyrate production. Biotechnol Bioeng 114:106–112.

133. Wlkinson SR, Young M. 1994. Targeted integration of genes into the *Clostridium acetobutylicum* chromosome. Microbiology 140:89–95.

134. Yang Z, Wang Z, Lei M, Zhu J, Yang Y, Wu S, Yu B, Niu H, Ying H, Liu D, Wang Y. 2020. Effects of Spo0A on *Clostridium acetobutylicum* with an emphasis on biofilm formation. World J Microbiol Biotechnol 36:1–9.

135. Bayat H, Modarressi MH, Rahimpour A. 2018. The Conspicuity of CRISPR-Cpf1 system as a significant breakthrough in genome editing. Curr Microbiol 75:107–115.

136. Schmeisser F, Brannigan JA, Lewis RJ, Wilkinson AJ, Youngman P, BarÃ¡k I. 2000. A new mutation in *spo0A* with intragenic suppressors in the effector domain. FEMS Microbiol Lett 185:123–128.

137. Li T, He J. 2016. Simultaneous saccharification and fermentation of hemicellulose to butanol by a non-sporulating *Clostridium* species. Bioresour Technol 219:430–438.

138. Tracy BP, Gaida SM, Papoutsakis ET. 2008. Development and application of flow-cytometric techniques for analyzing and sorting endospore-forming clostridia. Appl Environ Microbiol 74:7497–506.

139. Fujita M, González-Pastor JE, Losick R. 2005. High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*. J Bacteriol 187:1357–1368.

140. Narula J, Devi SN, Fujita M, Igoshin OA. 2012. Ultrasensitivity of the *Bacillus subtilis* sporulation decision. Proc Natl Acad Sci 109:E3513–E3522.

141. Williams RHN, Whitworth DE. 2010. The genetic organisation of prokaryotic two-component system signalling pathways. BMC Genomics 11:720.

142. Paredes-Sabja D, Shen A, Sorg JA. 2014. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. Trends Microbiol 22:406–416.

143. Freedman JC, Li J, Mi E, McClane BA. 2019. Identification of an important orphan histidine kinase for the initiation of sporulation and enterotoxin production by *Clostridium perfringens* type F strain SM101. MBio 10:e02674-18.

144. Zhao H, Msadek T, Zapf J, Madhusudan, Hoch JA, Varughese KI. 2002. DNA complexed structure of the key transcription factor initiating development in sporulating bacteria. Structure 10:1041–1050.

145. Sullivan L, Bennett GN. 2006. Proteome analysis and comparison of *Clostridium acetobutylicum* ATCC 824 and Spo0A strain variants. J Ind Microbiol Biotechnol 33:298–308.

146. Barák I, Muchová K, Labajová N. 2019. Asymmetric cell division during *Bacillus subtilis* sporulation. Future Microbiol 14:353–363.

147. Patakova P, Linhova M, Vykydalova P, Branska B, Rychtera M, Melzoch K. 2014. Use of fluorescent staining and flow cytometry for monitoring physiological changes in solventogenic clostridia. Anaerobe 29:113–7.

148. Paidhungat M, Setlow B, Driks A, Setlow P. 2000. Characterization of spores of *Bacillus subtilis* which lack dipicolinic acid. J Bacteriol 182:5505–5512.

149. Shen A, Edwards AN, Sarker MR, Paredes-Sabja D. 2019. Sporulation and germination in clostridial pathogens. Microbiol Spectr 7.

150. Jabbari S, Steiner E, Heap JT, Winzer K, Minton NP, King JR. 2013. The putative influence of the agr operon upon survival mechanisms used by *Clostridium acetobutylicum*. Math Biosci 243:223–239.

151. Li J, Chen J, Vidal JE, McClane BA. 2011. The Agr-like quorum-sensing system regulates sporulation and production of enterotoxin and beta2 toxin by *Clostridium perfringens* type a non-food-borne human gastrointestinal disease strain F5603. Infect Immun 79:2451–2459.

152. Long S, Jones DT, Woods DR. 1984. The relationship between sporulation and solvent production in *Clostridium acetobutylicum* P262. Biotechnol Lett 6:529–534.

153. George HA, Johnson JL, Moore WE, Holdeman L V, Chen JS. 1983. Acetone, Isopropanol, and Butanol production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium aurantibutyricum*. Appl Environ Microbiol 45:1160–3.

References

154. Clark SW, Bennett GN, Rudolph FB. 1989. Isolation and characterization of mutants of *Clostridium acetobutylicum* ATCC 824 deficient in acetoacetyl-coenzyme A:Acetate/Butyrate:Coenzyme A-Transferase (EC 2.8.3.9) and in other solvent pathway enzymes. Appl Environ Microbiol 55:970–976.
155. Kashket ER, Zhi-Yi C. 1995. Clostridial strain degeneration. FEMS Microbiol Rev 17:307–315.
156. Cornillot E, Nair RV, Papoutsakis ET, Soucaille P. 1997. The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. J Bacteriol 179:5442–7.
157. Lv J, Jiao S, Du R, Zhang R, Zhang Y, Han B. 2016. Proteomic analysis to elucidate degeneration of *Clostridium beijerinckii* NCIMB 8052 and role of Ca$^{(2+)}$ in strain recovery from degeneration. J Ind Microbiol Biotechnol 43:741–50.
158. Woo JE, Lee SY, Jang YS. 2018. Effects of nutritional enrichment on acid production from degenerated (non-solventogenic) *Clostridium acetobutylicum* strain M5. Appl Biol Chem 61:469–472.
159. Jones DT, Keis S. 1995. Origins and relationships of industrial solvent-producing clostridial strains. FEMS Microbiol Rev 17:223–232.
160. Li S, Huang L, Ke C, Pang Z, Liu L. 2020. Pathway dissection, regulation, engineering and application: Lessons learned from biobutanol production by solventogenic clostridia. Biotechnol Biofuels 13:39.
161. Mutschlechner O, Swoboda H, Gapes JR. 2000. Continuous two-stage ABE-fermentation using *Clostridium beijerinckii* NRRL B592 operating with a growth rate in the first stage vessel close to its maximal value. J Mol Microbiol Biotechnol 2:101–5.
162. Wolken WAM, Tramper J, van der Werf MJ. 2003. What can spores do for us? Trends Biotechnol 21:338–45.
163. Ricca E, Cutting SM. 2003. Emerging Applications of bacterial spores in nanobiotechnology. J Nanobiotechnology 1:6.
164. Zhang X, Al-Dossary A, Hussain M, Setlow P, Li J. 2020. Applications of *Bacillus subtilis* spores in biotechnology and advanced materials. Appl Environ Microbiol 86:1–22.
165. Mamane-Gravetz H, Linden KG. 2004. UV disinfection of indigenous aerobic spores: Implications for UV reactor validation in unfiltered waters. Water Res 38:2898–2906.
166. Ugwuodo CJ, Nwagu TN. 2020. Stabilizing enzymes by immobilization on bacterial spores: A review of literature. Int J Biol Macromol https://doi.org/10.1016/j.ijbiomac.2020.10.171.
167. Falahati-Pour SK, Lotfi AS, Ahmadian G, Baghizadeh A. 2015. Covalent immobilization of recombinant organophosphorus hydrolase on spores of *Bacillus subtilis*. J Appl Microbiol 118:976–988.
168. Hosseini-Abari A, Kim B-G, Lee S-H, Emtiazi G, Kim W, Kim J-H. 2016. Surface display of bacterial tyrosinase on spores of *Bacillus subtilis* using CotE as an anchor protein. J Basic Microbiol 56:1331–1337.
169. Song T, Wang F, Xiong S, Jiang H. 2019. Surface display of organophosphorus-degrading enzymes on the recombinant spore of *Bacillus subtilis*. Biochem Biophys Res Commun 510:13–19.
170. Peng F, Zheng B, Zhang Y, Faheem A, Chai Y, Jiang T, Chen X, Hu Y. 2020. Biocatalytic oxidation of aromatic compounds by spore-based system. ACS Sustain Chem Eng 8:14159–14165.
171. Cartman ST. 2011. Time to consider *Clostridium* probiotics? Future Microbiol 6:969–971.
172. Guo P, Zhang K, Ma X, He P. 2020. *Clostridium* species as probiotics: potentials and challenges. J Anim Sci Biotechnol 11:24.
173. Theys J, Lambin P. 2015. *Clostridium* to treat cancer: dream or reality? Ann Transl Med 3:S21.
174. Kubiak AM, Minton NP. 2015. The potential of clostridial spores as therapeutic delivery vehicles in tumour therapy. Res Microbiol 166:244–254.
175. Krouwel PG, Groot WJ, Kossen NWF, van der Laan WFM. 1983. Continuous isopropanol-butanol-ethanol fermentation by immobilized *Clostridium beijerinckii* cells in a packed bed fermenter. Enzyme Microb Technol 5:46–54.
176. Dolejš I, Krasňan V, Stloukal R, Rosenberg M, Rebroš M. 2014. Butanol production by immobilised *Clostridium acetobutylicum* in repeated batch, fed-batch, and continuous modes of fermentation. Bioresour Technol 169:723–730.
177. Rathore S, Wan Sia Heng P, Chan LW. 2015. Microencapsulation of *Clostridium acetobutylicum* ATCC 824 spores in gellan gum microspheres for the production of biobutanol. J Microencapsul 32:290–299.

178. de Vrije T, Budde M, van der Wal H, Claassen PAM, López-Contreras AM. 2013. "In situ" removal of isopropanol, butanol and ethanol from fermentation broth by gas stripping. Bioresour Technol 137:153–159.

179. Survase SA, Jurgens G, Van Heiningen A, Granström T. 2011. Continuous production of isopropanol and butanol using *Clostridium beijerinckii* DSM 6423. Appl Microbiol Biotechnol 91:1305–1313.

180. Collas F, Kuit W, Clement B, Marchal R, López-Contreras AM, Monot F. 2012. Simultaneous production of isopropanol, butanol, ethanol and 2,3-butanediol by *Clostridium acetobutylicum* ATCC 824 engineered strains. AMB Express 2:45.

181. Dürre P. 2008. Fermentative Butanol Production. Ann N Y Acad Sci 1125:353–362.

182. Siemerink M a J, Kuit W, Contreras AML, Eggink G, van der Oost J, Kengen SWM. 2011. D-2,3-butanediol production due to heterologous expression of an acetoin reductase in *Clostridium acetobutylicum*. Appl Environ Microbiol 77:2582–2588.

183. Forsberg CW. 1987. Production of 1,3-propanediol from glycerol by *Clostridium acetobutylicum* and other *Clostridium* species. Appl Environ Microbiol 53:639–43.

184. González-Pajuelo M, Andrade JC, Vasconcelos I. 2005. Production of 1,3-Propanediol by *Clostridium butyricum* VPI 3266 in continuous cultures with high yield and productivity. J Ind Microbiol Biotechnol 32:391–6.

185. Otte B, Grunwaldt E, Mahmoud O, Jennewein S. 2009. Genome shuffling in *Clostridium diolis* DSM 15410 for improved 1,3-propanediol production. Appl Environ Microbiol 75:7610–7616.

186. Patakova P, Kolek J, Sedlar K, Koscova P, Branska B, Kupkova K, Paulova L, Provaznik I. 2017. Comparative analysis of high butanol tolerance and production in clostridia. Biotechnol Adv. Elsevier.

187. Hermann M, Fayolle F, Marchal R, Podvin L, Sebald M, Vandecasteele JP. 1985. Isolation and characterization of butanol-resistant mutants of *Clostridium acetobutylicum*. Appl Environ Microbiol 50:1238–43.

188. Maté de Gérando H, Fayolle-Guichard F, Rudant L, Millah SK, Monot F, Lopes Ferreira N, López-Contreras AM. 2016. Improving isopropanol tolerance and production of *Clostridium beijerinckii* DSM 6423 by random mutagenesis and genome shuffling. Appl Microbiol Biotechnol 100:5427–36.

189. Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol 19:1162–1167.

190. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. 2010. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. J Microbiol Methods 80:49–55.

191. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. 2007. The ClosTron: A universal gene knock-out system for the genus *Clostridium*. J Microbiol Methods 70:452–464.

192. Rodriguez SA, Yu J-J, Davis G, Arulanandam BP, Klose KE. 2008. Targeted inactivation of *Francisella tularensis* genes by Group II Introns. Appl Environ Microbiol 74:2619–2626.

193. Al-Hinai MA., Fast AG, Papoutsakis ET. 2012. Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and dna integration. Appl Environ Microbiol 78:8112–8121.

194. Heap JT, Ehsaan M, Cooksley CM, Ng YK, Cartman ST, Winzer K, Minton NP. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Res 40:e59.

195. Mougiakos I, Bosma EF, Ganguly J, van der Oost J, van Kranenburg R. 2018. Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. Curr Opin Biotechnol.

196. Xu T, Li Y, Shi Z, Hemme CL, Li Y, Zhu Y, Van Nostrand JD, He Z, Zhou J. 2015. Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. Appl Environ Microbiol 81:4423–31.

197. Wang Y, Zhang ZT, Seo SO, Lynn P, Lu T, Jin YS, Blaschek HP. 2016. Bacterial genome editing with Crispr-cCs9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in *Clostridium beijerinckii* as an example. ACS Synth Biol 5:721–732.

198. Wang Y, Zhang ZT, Seo SO, Choi K, Lu T, Jin YS, Blaschek HP. 2015. Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. J Biotechnol 200:1–5.

199. Li Q, Seys FM, Minton NP, Yang J, Jiang Y, Jiang W, Yang S. 2019. CRISPR-Cas9 D10A nickase-assisted base editing in solvent producer *Clostridium beijerinckii*. Biotechnol Bioeng https://doi.org/10.1002/bit.26949.

200. Wilson TE, Topper LM, Palmbos PL. 2003. Non-homologous end-joining: bacteria join the chromosome breakdance. Trends Biochem Sci 28:62–66.

References

201. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable Dual-RNA-Guided DNA endonuclease in adaptive bacterial immunity. Science (80-) 337:816–821.

202. Hartman AH, Liu H, Melville SB. 2011. Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*. Appl Environ Microbiol 77:471–8.

203. Wasels F, Jean-Marie J, Collas F, López-Contreras AM, Lopes Ferreira N. 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. J Microbiol Methods 140:5–11.

204. Nagaraju S, Davies NK, Walker DJF, Köpke M, Simpson SD. 2016. Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. Biotechnol Biofuels 9:219.

205. Nariya H, Miyata S, Kuwahara T, Okabe A. 2011. Development and characterization of a xylose-inducible gene expression system for *Clostridium perfringens*. Appl Environ Microbiol 77:8439–8441.

206. López-Contreras AM, Smidt H, de Vos WM, Mooibroek H, Claassen PAM, van der Oost J. 2002. *Clostridium beijerinckii* cells expressing *Neocallimastix patriciarum* glycoside hydrolases show enhanced lichenan utilization and solvent production. Appl Environ Microbiol 67:5127–5133.

207. Oultram JD, Loughlin M, Swinfield TJ, Brehm JK, Thompson DE, Minton NP. 1988. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. FEMS Microbiol Lett 56:83–88.

208. Colot H V., Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC. 2006. A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. Proc Natl Acad Sci U S A 103:10352–10357.

209. Agatep R, Kirkpatrick RD, Parchaliuk DL, Woods RA, Gietz RD. 1998. Transformation of *Saccharomyces cerevisiae* by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol. Tech Tips Online 3:133–137.

210. Quan J, Tian J. 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441.

211. Wang Y, Li X, Mao Y, Blaschek HP. 2012. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics 13:102.

212. Dong H, Tao W, Zhang Y, Li Y. 2012. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. Metab Eng 14:59–67.

213. Mermelstein LD, Papoutsakis ET. 1993. In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. Appl Environ Microbiol 59:1077–81.

214. Barák I, Behari J, Olmedo G, Guzmán P, Brown DP, Castro E, Walker D, Westpheling J, Youngman P. 1996. Structure and function of the *Bacillus* SpoIIE protein and its localization to sites of sporulation septum assembly. Mol Microbiol 19:1047–60.

215. Denman S, Xue GP, Patel B. 1996. Characterization of a *Neocallimastix patriciarum* cellulase cDNA (celA) homologous to *Trichoderma reesei* cellobiohydrolase II. Appl Environ Microbiol 62:1889–96.

216. Kolek J, Sedlar K, Provaznik I, Patakova P. 2016. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol Biofuels 9:14.

217. Schwarz K, Zhang Y, Kuit W, Ehsaan M, Kovács K, Winzer K, Minton NP. 2015. New tools for the genetic modification of industrial clostridia, p. 241–289. *In* Direct Microbial Conversion of Biomass to Advanced Biofuels. Elsevier.

218. Steffen C, Matzura H. 1989. Nucleotide sequence analysis and expression studies of a chloramphenicol-acetyltransferase-coding gene from *Clostridium perfringens*. Gene 75:349–354.

219. Barák I, Youngman P. 1996. SpoIIE mutants of *Bacillus subtilis* comprise two distinct phenotypic classes consistent with a dual functional role for the SpoIIE protein. J Bacteriol 178:4984–4989.

220. Carniol K, Ben-Yehuda S, King N, Losick R. 2005. Genetic dissection of the sporulation protein SpoIIE and its role in asymmetric division in *Bacillus subtilis*. J Bacteriol 187:3511–3520.

221. Mingardon F, Chanal A, López-Contreras AM, Dray C, Bayer EA, Fierobe HP. 2007. Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. Appl Environ Microbiol 73:3822–3832.

222. Poehlein A, Solano JDM, Flitsch SK, Krabben P, Winzer K, Reid SJ, Jones DT, Green E, Minton NP, Daniel R, Dürre P. 2017. Microbial solvent formation revisited by comparative genome analysis. Biotechnol Biofuels 10:58.

223. Gu Y, Feng J, Zhang ZT, Wang S, Guo L, Wang Y, Wang Y. 2019. Curing the endogenous megaplasmid in *Clostridium saccharoperbutylacetonicum* N1-4 (HMT) using CRISPR-Cas9 and preliminary investigation of the role of the plasmid for the strain metabolism. Fuel 236:1559–1566.

224. Paredes CJ, Alsaker K V., Papoutsakis ET. 2005. A comparative genomic view of clostridial sporulation and physiology. Nat Rev Microbiol 3:969–978.

225. Feucht A, Abbotts L, Errington J. 2002. The cell differentiation protein SpoIIE contains a regulatory site that controls its phosphatase activity in response to asymmetric septation. Mol Microbiol 45:1119–1130.

226. Muchová K, Chromiková Z, Bradshaw N, Wilkinson AJ, Barák I. 2016. Morphogenic protein RodZ interacts with sporulation specific SpoIIE in *Bacillus subtilis*. PLoS One 11:e0159076.

227. Muchová K, Chromiková Z, Barák I. 2020. Linking the peptidoglycan synthesis protein complex with asymmetric cell division during *Bacillus subtilis* sporulation. Int J Mol Sci 21:4513.

228. Louie P, Lee A, Stansmore K, Grant R, Ginther C, Leighton T. 1992. Roles of *rpoD*, *spoIIF*, *spoIIJ*, *spoIIN*, and sin in regulation of *Bacillus subtilis* stage II sporulation-specific transcription. J Bacteriol 174:3570–6.

229. Levin PA, Losick R. 1994. Characterization of a cell division gene from *Bacillus subtilis* that is required for vegetative and sporulation septum formation. J Bacteriol 176:1451–9.

230. van der Wal H, Sperber BLHM, Houweling-Tan B, Bakker RRC, Brandenburg W, López-Contreras AM. 2013. Production of acetone, butanol, and ethanol from biomass of the green seaweed *Ulva lactuca*. Bioresour Technol 128:431–437.

231. Gu Y, Jiang Y, Yang S, Jiang W. 2014. Utilization of economical substrate-derived carbohydrates by solventogenic clostridia: Pathway dissection, regulation and engineering. Curr Opin Biotechnol. Elsevier Ltd.

232. Quan J, Tian J. 2014. Circular polymerase extension cloning. Methods Mol Biol 1116:103–17.

233. Ransom EM, Weiss DS, Ellermeier CD. 2016. Use of mCherryOpt fluorescent protein in *Clostridium difficile*, p. 53–67. *In* Methods in Molecular Biology. Humana Press, New York, NY.

234. Ransom EM, Ellermeier CD, Weiss DS. 2015. Use of mCherry red fluorescent protein for studies of protein localization and gene expression in *Clostridium difficile*. Appl Environ Microbiol 81:1652–1660.

235. Robson RL, Robson RM, Morris JG. 1974. The biosynthesis of granulose by *Clostridium pasteurianum*. Biochem J 144:503–11.

236. Schaeffer AB, Fulton MD. 1933. A simplified method of staining endospores. Science (80-) 77.

237. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

238. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

239. Varet H, Brillet-Guéguen L, Coppée JY, Dillies MA. 2016. SARTools: A DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. PLoS One 11:e0157022.

240. Monot M, Orgeur M, Camiade E, Brehier C, Dupuy B. 2014. COV2HTML: A visualization and analysis tool of bacterial Next Generation Sequencing (NGS) data for postgenomics life scientists. Omi A J Integr Biol 18:184–195.

241. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45.

242. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305:567–580.

243. Grage K, McDermott P, Rehm BHA. 2017. Engineering *Bacillus megaterium* for production of functional intracellular materials. Microb Cell Fact 16:211.

244. Tang H, Bomhoff MD, Briones E, Zhang L, Schnable JC, Lyons E. 2015. SynFind: Compiling syntenic regions across any set of genomes on demand. Genome Biol Evol 7:3286–3298.

245. Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin P V., Gelfand MS, Dupuy B, Henriques AO, Martin-Verstraete I. 2013. Genome-Wide analysis of cell type-specific gene transcription during spore formation in *Clostridium difficile*. PLoS Genet 9:e1003756.

References

246. Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, Shen A. 2013. Global analysis of the sporulation pathway of *Clostridium difficile*. PLoS Genet 9:e1003660.

247. Wang ST, Setlow B, Conlon EM, Lyon JL, Imamura D, Sato T, Setlow P, Losick R, Eichenberger P. 2006. The forespore line of gene expression in *Bacillus subtilis*. J Mol Biol 358:16–37.

248. Zhu B, Stülke J. 2018. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism *Bacillus subtilis*. Nucleic Acids Res 46:D743–D748.

249. Saujet L, Pereira FC, Henriques AO, Martin-Verstraete I. 2014. The regulatory network controlling spore formation in *Clostridium difficile*. FEMS Microbiol Lett 358:1–10.

250. Barák I, Muchová K. 2018. The positioning of the asymmetric septum during sporulation in *Bacillus subtilis*. PLoS One 13:e0201979.

251. Brown EE, Miller AK, Krieger I V., Otto RM, Sacchettini JC, Herman JK. 2019. A DNA-Binding Protein Tunes Septum Placement during *Bacillus subtilis* Sporulation . J Bacteriol 201.

252. Kotte A-K, Severn O, Bean Z, Schwarz K, Minton NP, Winzer K. 2017. RNPP-type quorum-sensing regulates solvent formation and sporulation in *Clostridium acetobutylicum*. bioRxiv 44:106666.

253. Kemperman R, Jonker M, Nauta A, Kuipers OP, Kok J. 2003. Functional analysis of the gene cluster involved in production of the bacteriocin circularin a by *Clostridium beijerinckii* ATCC 25752. Appl Environ Microbiol 69:5839–5848.

254. Gutierrez NA, Maddox IS. 1987. Role of Chemotaxis in Solvent Production by *Clostridium acetobutylicum*. Appl Environ Microbiol 53:1924–7.

255. Eydallin G, Viale AM, Morán-Zorzano MT, Muñoz FJ, Montero M, Baroja-Fernández E, Pozueta-Romero J. 2007. Genome-wide screening of genes affecting glycogen metabolism in *Escherichia coli* K-12. FEBS Lett 581:2947–2953.

256. Montero M, Eydallin G, Viale AM, Almagro G, Muñoz FJ, Rahimpour M, Sesma MT, Baroja-Fernández E, Pozueta-Romero J. 2009. *Escherichia coli* glycogen metabolism is controlled by the PhoP-PhoQ regulatory system at submillimolar environmental Mg^{2+} concentrations, and is highly interconnected with a wide variety of cellular processes. Biochem J 424:129–141.

257. Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, Viale AM, Pozueta-Romero J. 2010. Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev. Oxford Academic.

258. Li D, Meng C, Wu G, Xie B, Han Y, Guo Y, Song C, Gao Z, Huang Z. 2018. Effects of zinc on the production of alcohol by *Clostridium carboxidivorans* P7 using model syngas. J Ind Microbiol Biotechnol 45:61–69.

259. Edwards AN, Nawrocki KL, McBride SM. 2014. Conserved oligopeptide permeases modulate sporulation initiation in *Clostridium difficile*. Infect Immun 82:4276–4291.

260. Takagi H. 2008. Proline as a stress protectant in yeast: Physiological functions, metabolic regulations, and biotechnological applications. Appl Microbiol Biotechnol.

261. Li T, Yan Y, He J. 2014. Reducing cofactors contribute to the increase of butanol production by a wild-type *Clostridium* sp. strain BOH3. Bioresour Technol 155:220–228.

262. De Hoon MJL, Eichenberger P, Vitkup D. 2010. Hierarchical evolution of the bacterial sporulation network. Curr Biol. NIH Public Access.

263. Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics 12:116.

264. Lépiz-Aguilar L, Rodríguez-Rodríguez CE, Arias ML, Lutz G, Ulate W. 2011. Butanol production by *Clostridium beijerinckii* BA101 using cassava flour as fermentation substrate: enzymatic versus chemical pretreatments. World J Microbiol Biotechnol 27:1933–1939.

265. Maiti S, Gallastegui G, Sarma SJ, Brar SK, Le Bihan Y, Drogui P, Buelna G, Verma M. 2016. A re-look at the biochemical strategies to enhance butanol production. Biomass and Bioenergy 94:187–200.

266. Reysenbach AL, Ravenscroft N, Long S, Jones DT, Woods DR. 1986. Characterization, Biosynthesis, and Regulation of Granulose in *Clostridium acetobutylicum*. Appl Environ Microbiol 52:185–90.

267. Tomas CA, Welker NE, Papoutsakis ET. 2003. Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. Appl Environ Microbiol 69:4951–4965.

268. Liao Z, Zhang Y, Luo S, Suo Y, Zhang S, Wang J. 2017. Improving cellular robustness and butanol titers of *Clostridium acetobutylicum* ATCC824 by introducing heat shock proteins from an extremophilic bacterium. J Biotechnol 252:1–10.

216

269. Reusser F. 1963. Stability and degeneration of microbial cultures on repeated transfer. Adv Appl Microbiol 5:189–215.

270. Peng M, Liang Z. 2020. Degeneration of industrial bacteria caused by genetic instability. World J Microbiol Biotechnol. Springer.

271. Petit E, LaTouf WG, Coppi M V., Warnick TA, Currie D, Romashko I, Deshpande S, Haas K, Alvelo-Maurosa JG, Wardman C, Schnell DJ, Leschine SB, Blanchard JL. 2013. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. PLoS One 8:e54337.

272. Hayashida S, Yoshino S. 1990. Degeneration of solventogenic *Clostridium* caused by a defect in NADH Generation. Agric Biol Chem 54:427–435.

273. Guerrero K, Gallardo R, Paredes I, Quintero J, Mau S, Conejeros R, Gentina JC, Aroca G. 2020. Continuous biohydrogen production by a degenerated strain of *Clostridium acetobutylicum* ATCC 824. Int J Hydrogen Energy https://doi.org/10.1016/j.ijhydene.2020.11.104.

274. Woolley RC, Morris JG. 1990. Stability of solvent production by *Clostridium acetobutylicum* in continuous culture: strain differences. J Appl Bacteriol 69:718–728.

275. Schuster K, Goodacre R, Gapes J, Young M. 2001. Degeneration of solventogenic *Clostridium* strains monitored by Fourier transform infrared spectroscopy of bacterial cells. J Ind Microbiol Biotechnol 27:314–321.

276. Aguilar C, Vlamakis H, Losick R, Kolter R. 2007. Thinking about *Bacillus subtilis* as a multicellular organism. Curr Opin Microbiol 10:638–643.

277. Hamze K, Julkowska D, Autret S, Hinc K, Nagorska K, Sekowska A, Holland IB, Séror SJ. 2009. Identification of genes required for different stages of dendritic swarming in *Bacillus subtilis*, with a novel role for phrC. Microbiology 155:398–412.

278. Tomas CA, Alsaker K V., Bonarius HPJ, Hendriksen WT, Yang H, Beamish JA, Paredes CJ, Papoutsakis ET. 2003. DNA array-based transcriptional analysis of asporogenous, nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5. J Bacteriol 185:4539–4547.

279. Kim M, Park S, Lee S-J. 2016. Global transcriptional regulator TrmB family members in prokaryotes. J Microbiol 54:639–645.

280. Gundlach J, Mehne FMP, Herzberg C, Kampf J, Valerius O, Kaever V, Stülke J. 2015. An essential poison: synthesis and degradation of cyclic di-AMP in *Bacillus subtilis*. J Bacteriol 197:3265–3274.

281. Commichau FM, Dickmanns A, Gundlach J, Ficner R, Stülke J. 2015. A jack of all trades: The multiple roles of the unique essential second messenger cyclic di-AMP. Mol Microbiol. Blackwell Publishing Ltd.

282. Mehne FMP, Gunka K, Eilers H, Herzberg C, Kaever V, Stülke J. 2013. Cyclic Di-AMP homeostasis in *Bacillus subtilis*: Both lack and high level accumulation of the nucleotide are detrimental for cell growth. J Biol Chem 288:2004–2017.

283. Heimann JD. 2002. The extracytoplasmic function (ECF) sigma factors. Adv Microb Physiol. Academic Press.

284. Asai K, Ishiwata K, Matsuzaki K, Sadaie Y. 2008. A viable *Bacillus subtilis* strain without functional extracytoplasmic function sigma genes. J Bacteriol 190:2633–2636.

285. Asai K. 2017. Anti-Sigma factor-mediated cell surface stress responses in *Bacillus subtilis*. Genes Genet Syst. Genetics Society of Japan.

286. Behrens S, Meyer U, Schankin H, Lonetto MA, Fischer RJ, Bahl H. 2000. Identification of two genes encoding putative new members of the ECF subfamily of eubacterial RNA polymerase sigma factors in *Clostridium acetobutylicum*. J Mol Microbiol Biotechnol 2:265–269.

287. Ho TD, Ellermeier CD. 2011. PrsW is required for colonization, resistance to antimicrobial peptides, and expression of extracytoplasmic function σ factors in *Clostridium difficile*. Infect Immun 79:3229–3238.

288. Woods EC, Nawrocki KL, Suárez JM, McBride SM. 2016. The *Clostridium difficile* Dlt pathway is controlled by the extracytoplasmic function sigma factor σ^V in response to lysozyme. Infect Immun 84:1902–1916.

289. Yang B, Nie X, Gu Y, Jiang W, Yang C. 2019. Control of solvent production by sigma-54 factor and the transcriptional activator AdhR in *Clostridium beijerinckii*. Microb Biotechnol 1751-7915.13505.

290. Walter BM, Rupnik M, Hodnik V, Anderluh G, Dupuy B, Paulič N, Žgur-Bertok D, Butala M. 2014. The LexA regulated genes of the *Clostridium difficile*. BMC Microbiol 14:88.

References

291. Wang Q, Venkataramanan KP, Huang H, Papoutsakis ET, Wu CH. 2013. Transcription factors and genetic circuits orchestrating the complex, multilayered response of *Clostridium acetobutylicum* to butanol and butyrate stress. BMC Syst Biol 7:120.

292. Alsaker K V., Papoutsakis ET. 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. J Bacteriol 187:7103–7118.

293. Nawrocki KL, Edwards AN, Daou N, Bouillaut L, McBride SM. 2016. CodY-dependent regulation of sporulation in *Clostridium difficile*. J Bacteriol 198:2113–2130.

294. Dürre P. 2011. Ancestral sporulation initiation. Mol Microbiol 80:584–587.

295. Cherubini F. 2010. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. Energy Convers Manag 51:1412–1421.

296. IEA. Aims & Objectives of Task 42 | Task 42.

297. Burg S Van Den, Stuiver M, Veenstra F, Bikker P, Contreras AL, Palstra A, Broeze J, Jansen H, Jak R, Gerritsen A, Harmsen P, Kals J, Blanco A, Brandenburg W, van Krimpen M, van Duijn AP, Mulder W, van Raamsdonk L. 2013. A Triple P review of the feasibility of sustainable offshore seaweed production in the North SeaLEI Report 13-077.

298. Kraan S. 2013. Mass-cultivation of carbohydrate rich macroalgae, a possible solution for sustainable biofuel production. Mitig Adapt Strateg Glob Chang 18:27–46.

299. Van Hal JW, Huijgen WJJ, López-Contreras AM. 2014. Opportunities and challenges for seaweed in the biobased economy. Trends in Biotechnology. Elsevier Ltd.

300. Kim JK, Yarish C, Hwang EK, Park M, Kim Y. 2017. Seaweed aquaculture: cultivation technologies, challenges and its ecosystem services. Algae 2017:1–13.

301. Rioux L-EE, Turgeon SL. 2015. Seaweed carbohydrates, p. 141–192. *In* .

302. Forsberg CW, Donaldson L, Gibbins LN. 1987. Metabolism of rhamnose and other sugars by strains of *Clostridium acetobutylicum* and other *Clostridium* species. Can J Microbiol 33:21–26.

303. Boronat A, Aguilar J. 1981. Metabolism of L-fucose and L-rhamnose in *Escherichia coli:* Differences in induction of propanediol oxidoreductase. J Bacteriol 147:181–185.

304. Badía J, Ros J, Aguilar J. 1985. Fermentation mechanism of fucose and rhamnose in *Salmonella typhimurium* and *Klebsiella pneumoniae*. J Bacteriol 161:435–437.

305. Petit E, Coppi M V., Hayes JC, Tolonen AC, Warnick T, Latouf WG, Amisano D, Biddle A, Mukherjee S, Ivanova N, Lykidis A, Land M, Hauser L, Kyrpides N, Henrissat B, Lau J, Schnell DJ, Church GM, Leschine SB, Blanchard JL. 2015. Genome and transcriptome of *Clostridium phytofermentans*, catalyst for the direct conversion of plant feedstocks to fuels. PLoS One 10.

306. Bennett GN, San KY. 2001. Microbial formation, biotechnological production and applications of 1,2-propanediol. Appl Microbiol Biotechnol. Springer.

307. Rodionova IA, Li X, Thiel V, Stolyar S, Stanton K, Fredrickson JK, Bryant DA, Osterman AL, Best AA, Rodionov DA. 2013. Comparative genomics and functional analysis of rhamnose catabolic pathways and regulons in bacteria 4:407.

308. Bikker P, van Krimpen MM, van Wikselaar P, Houweling-Tan B, Scaccia N, van Hal JW, Huijgen WJJ, Cone JW, López-Contreras AM. 2016. Biorefinery of the green seaweed *Ulva lactuca* to produce animal feed, chemicals and biofuels. J Appl Phycol 28:3511–3525.

309. Mate de Gerando H, Wasels F, Bisson A, Clement B, Bidard F, Jourdier E. 2018. Genome sequence of the natural Isopropanol producer *Clostridium beijerinckii* DSM6423.

310. Altschul S, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

311. Richardson JS, Hynes MF, Oresnik IJ. 2004. A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum bv. trifolii*. J Bacteriol 186:8433–8442.

312. Axen SD, Erbilgin O, Kerfeld CA. 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring Method. PLOS Comput Biol 10:e1003898.

313. Zarzycki J, Erbilgin O, Kerfeld CA. 2015. Bioinformatic characterization of glycyl radical enzyme-associated bacterial microcompartments. Appl Environ Microbiol 81:8315–29.

314. Moralejo P, Egan SM, Hidalgo E, Aguilar J. 1993. Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. J Bacteriol 175:5585–5594.

315. Chen YM, Lu Z, Lin CC. 1989. Constitutive activation of the *fucAO* operon and silencing of the divergently transcribed *fucPIK* operon by an IS5 element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. J Bacteriol 171:6097–6105.

316. Hugouvieux-Cotte-Pattat N. 2004. The RhaS activator controls the *Erwinia chrysanthemi* 3937 genes *rhiN*, *rhiT* and *rhiE* involved in rhamnogalacturonan catabolism. Mol Microbiol 51:1361–1374.

317. Hugouvieux-Cotte-Pattat N, Reverchon S. 2001. Two transporters, TogT and TogMNAB, are responsible for oligogalacturonide uptake in *Erwinia chrysanthemi* 3937. Mol Microbiol 41:1125–1132.

318. Mate de Gérando H, Bidard-Michelot F, Jourdier E, López-Contreras AM, Lopes Ferreira N. 2018. RNA-seq analysis of glucose fermentation by the natural Isopropanol producer *Clostridium beijerinckii* DSM 6423.

319. Skraly FA, Lytle BL, Cameron DC. 1998. Construction and characterization of a 1,3-propanediol operon. Appl Environ Microbiol 64:98–105.

320. Raynaud C, Sarcabal P, Meynial-Salles I, Croux C, Soucaille P. 2003. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. Proc Natl Acad Sci 100:5010–5015.

321. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. 2018. Bacterial microcompartments. Nat Rev Microbiol 16:277–290.

322. EAGON RG. 1961. Bacterial dissimilation of L-fucose and L-rhamnose. J Bacteriol 82:548–550.

323. Nakas JP, Schaedle M, Parkinson CM, Coonley CE, Tanenbaum SW. 1983. System development for linked-fermentation production of solvents from algal biomass. Appl Environ Microbiol 46.

324. Houweling-Tan GBN, Sperber BLHM, Wal H van der, Bakker RRC, López-Contreras AM. 2016. Barley Distillers Dried Grains with Solubles (DDGS) as feedstock for production of acetone, butanol and ethanol. BAOJ Microbiol 2.

325. Chowdhury C, Chun S, Sawaya MR, Yeates TO, Bobik TA. 2016. The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. Mol Microbiol 101:770–783.

326. Millat T, Janssen H, Thorn GJ, King JR, Bahl H, Fischer RJ, Wolkenhauer O. 2013. A shift in the dominant phenotype governs the pH-induced metabolic switch of *Clostridium acetobutylicum* in phosphate-limited continuous cultures. Appl Microbiol Biotechnol 97:6451–6466.

327. Cheng S, Liu Y, Crowley CS, Yeates TO, Bobik TA. 2008. Bacterial microcompartments: Their properties and paradoxes. BioEssays. John Wiley & Sons, Ltd.

328. Von Stockar U, Liu JS. 1999. Does microbial life always feed on negative entropy? Thermodynamic analysis of microbial growth. Biochim Biophys Acta - Bioenerg. Biochim Biophys Acta.

329. Filippidou S, Wunderlin T, Junier T, Jeanneret N, Dorador C, Molina V, Johnson DR, Junier P. 2016. A combination of extreme environmental conditions favor the prevalence of endospore-forming firmicutes. Front Microbiol 7:1707.

330. Adams BL. 2016. The next generation of synthetic biology chassis: moving synthetic biology from the laboratory to the field. ACS Synth Biol 5:1328–1330.

331. Kaushik SS, Kaushik SS, Sharma D. 2018. Functional genomics. Encycl Bioinforma Comput Biol ABC Bioinforma 1–3:118–133.

332. Williams DR, Young DI, Young M. 1990. Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. J Gen Microbiol 136:819–826.

333. Heap JT, Pennington OJ, Cartman ST, Minton NP. 2009. A modular system for *Clostridium* shuttle plasmids. J Microbiol Methods 78:79–85.

334. Oh YH, Eom GT, Kang KH, Choi JW, Song BK, Lee SH, Park SJ. 2015. Optimized transformation of newly constructed *Escherichia coli*-Clostridia shuttle vectors into *Clostridium beijerinckii*. Appl Biochem Biotechnol 177:226–236.

335. Raghavan R, Kelkar YD, Ochman H. 2012. A selective force favoring increased G+C content in bacterial genes. Proc Natl Acad Sci U S A 109:14504–14507.

336. Lamberte LE, Baniulyte G, Singh SS, Stringer AM, Bonocora RP, Stracy M, Kapanidis AN, Wade JT, Grainger DC. 2017. Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. Nat Microbiol 2.

337. Mordaka PM, Heap JT. 2018. Stringency of synthetic promoter sequences in *Clostridium* revealed and circumvented by tuning promoter library mutation rates. ACS Synth Biol 7:672–681.

References

338.   Pyne ME, Bruder M, Moo-Young M, Chung D a., Chou CP. 2014. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. Biotechnol Adv 32:623–641.

339.   McAllister KN, Sorg JA. 2019. CRISPR genome editing systems in the genus *Clostridium*: a timely advancement. J Bacteriol. American Society for Microbiology Journals.

340.   Kwon SW, Paari KA, Malaviya A, Jang Y-S. 2020. Synthetic biology tools for genome and transcriptome engineering of solventogenic *Clostridium*. Front Bioeng Biotechnol 8.

341.   Merino N, Zhang S, Tomita M, Suzuki H. 2019. Comparative genomics of bacteria commonly identified in the built environment. BMC Genomics 20:1–17.

342.   Krejci L, Altmannova V, Spirek M, Zhao X. 2012. Homologous recombination and its regulation. Nucleic Acids Res. Oxford Academic.

343.   Michel B, Leach D. 2012. Homologous Recombination—Enzymes and Pathways. EcoSal Plus 5.

344.   Rocha EPC, Cornet E, Michel B. 2005. Comparative and Evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet 1:e15.

345.   Burgers PMJ, Percival KJ. 1987. Transformation of yeast spheroplasts without cell fusion. Anal Biochem 163:391–397.

346.   Hanahan D, Jessee J, Bloom FR. 1991. Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol 204:63–113.

347.   New england biolabs. Cloning Competent Cell Strains | NEB.

348.   Zhang XZ, Zhang YHP. 2011. Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. Microb Biotechnol 4:98–105.

349.   Meddeb-Mouelhi F, Dulcey C, Beauregard M. 2012. High transformation efficiency of *Bacillus subtilis* with integrative DNA using glycine betaine as osmoprotectant. Anal Biochem 424:127–129.

350.   You C, Zhang X-Z, Zhang Y-HP. 2012. Simple Cloning via Direct Transformation of PCR Product (DNA Multimer) to *Escherichia coli* and *Bacillus subtilis*. Am Soc Microbiol https://doi.org/10.1128/AEM.07105-11.

351.   Li Y, Tan D, Blaschek HPH. 2003. Molecular characterization and utilization of the CAK1 filamentous viruslike particle derived from *Clostridium beijerinckii*. J Ind Microbiol Biotechnol 30:133–133.

352.   Little GT, Willson BJ, Heap JT, Winzer K, Minton NP. 2018. The Butanol Producing Microbe *Clostridium beijerinckii* NCIMB 14988 Manipulated Using Forward and Reverse Genetic Tools. Biotechnol J 13:1700711.

353.   Wang Y, Li X, Milne CB, Janssen H, Lin W, Phan G, Hu H, Jin Y-S, Price ND, Blaschek HP. 2013. Development of a gene knockout system using mobile group ii introns (targetron) and genetic disruption of acid production pathways in *Clostridium beijerinckii*. Appl Environ Microbiol 79:5853–5863.

354.   Wen Z, Lu M, Ledesma-Amaro R, Li Q, Jin M, Yang S. 2020. TargeTron technology applicable in solventogenic clostridia: revisiting 12 Years' advances. Biotechnol J 15:1900284.

355.   Zhang N, Shao L, Jiang Y, Gu Y, Li Q, Liu J, Jiang W, Yang S. 2015. I-SceI-mediated scarless gene modification via allelic exchange in *Clostridium*. J Microbiol Methods 108:49–60.

356.   Wang Y, Zhang ZT, Seo SO, Lynn P, Lu T, Jin YS, Blaschek HP. 2016. Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. Biotechnol Bioeng 113:2739–2743.

357.   Cañadas IC, Groothuis D, Zygouropoulou M, Rodrigues R, Minton NP. 2019. RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. ACS Synth Biol 8:1379–1390.

358.   Metzger MJ, McConnell-Smith A, Stoddard BL, Miller AD. 2011. Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. Nucleic Acids Res 39:926–935.

359.   Wasels F, Chartier G, Hocq R, Lopes Ferreira N. 2020. A CRISPR/Anti-CRISPR genome editing approach underlines the synergy of butanol dehydrogenases in *Clostridium acetobutylicum* DSM 792. Appl Environ Microbiol https://doi.org/10.1128/aem.00408-20.

360.   Huang H, Chai C, Yang S, Jiang W, Gu Y. 2019. Phage serine integrase-mediated genome engineering for efficient expression of chemical biosynthetic pathway in gas-fermenting *Clostridium ljungdahlii*. Metab Eng 52:293–302.

361.   Tummala SB, Welker NE, Papoutsakis ET. 2003. Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. J Bacteriol 185:1923–34.

362. Cho C, Lee SY. 2017. Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs. Biotechnol Bioeng 114:374–383.

363. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY. 2013. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. Nat Biotechnol 31:170–174.

364. Lee J, Jang YS, Papoutsakis ET, Lee SY. 2016. Stable and enhanced gene expression in *Clostridium acetobutylicum* using synthetic untranslated regions with a stem-loop. J Biotechnol 230:40–43.

365. Vasylkivska M, Branska B, Sedlar K, Jureckova K, Provaznik I, Patakova P. 2020. Phenotypic and genomic analysis of *Clostridium beijerinckii* NRRL B-598 mutants with increased butanol Tolerance. Front Bioeng Biotechnol 8:598392.

366. Gao X, Zhao H, Zhang G, He K, Jin Y. 2012. Genome shuffling of *Clostridium acetobutylicum* CICC 8012 for improved production of acetone-butanol-ethanol (ABE). Curr Microbiol 65:128–132.

367. Cartman ST, Minton NP. 2010. A mariner-based transposon system for in vivo random mutagenesis of *Clostridium difficile*. Appl Environ Microbiol 76:1103–1109.

368. Gyulev IS, Willson BJ, Hennessy RC, Krabben P, Jenkinson ER, Thomas GH. 2018. Part by Part: synthetic biology parts used in solventogenic clostridia. ACS Synth Biol 7:311–327.

369. Yang G, Jia D, Jin L, Jiang Y, Wang Y, Jiang W, Gu Y. 2017. Rapid generation of universal synthetic promoters for controlled gene expression in both gas-fermenting and saccharolytic *Clostridium* species. ACS Synth Biol 6:1672–1678.

370. Buckley AM, Jukes C, Candlish D, Irvine JJ, Spencer J, Fagan RP, Roe AJ, Christie JM, Fairweather NF, Douce GR. 2016. Lighting Up *Clostridium* Difficile: Reporting Gene Expression Using Fluorescent Lov Domains. Sci Rep 6:23463.

371. Seo SO, Lu T, Jin YS, Blaschek HP. 2018. Development of an oxygen-independent flavin mononucleotide-based fluorescent reporter system in *Clostridium beijerinckii* and its potential applications. J Biotechnol 265:119–126.

372. Streett HE, Kalis KM, Papoutsakis ET. 2019. A Strongly fluorescing anaerobic reporter and protein-tagging system for *Clostridium* organisms based on the fluorescence-activating and absorption-shifting tag protein (FAST). Appl Environ Microbiol 85:AEM.00622-19.

373. Charubin K, Streett H, Papoutsakis ET. 2020. Development of strong anaerobic fluorescent reporters for *Clostridium acetobutylicum* and *Clostridium ljungdahlii* using halotag and snap-tag proteins. Appl Environ Microbiol 86:1–19.

374. Teng L, Wang K, Xu J, Xu C. 2015. Flavin mononucleotide (FMN)-based fluorescent protein (FbFP) as reporter for promoter screening in *Clostridium cellulolyticum*. J Microbiol Methods 119:37–43.

375. Molitor B, Kirchner K, Henrich AW, Schmitz S, Rosenbaum MA. 2016. Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. Sci Rep 6.

376. Yoo M, Nguyen N-P-T, Soucaille P. 2020. Trends in systems biology for the analysis and engineering of *Clostridium acetobutylicum* metabolism. Trends Microbiol 28:118–140.

377. Venkataramanan KP, Min L, Hou S, Jones SW, Ralston MT, Lee KH, Papoutsakis ET. 2015. Complex and extensive post-transcriptional regulation revealed by integrative proteomic and transcriptomic analysis of metabolite stress response in *Clostridium acetobutylicum*. Biotechnol Biofuels 8:81.

378. Janssen H, Döring C, Ehrenreich A, Voigt B, Hecker M, Bahl H, Fischer R-J. 2010. A proteomic and transcriptional view of acidogenic and solventogenic steady-state cells of *Clostridium acetobutylicum* in a chemostat culture. Appl Microbiol Biotechnol 87:2209–26.

379. Munir RI, Spicer V, Krokhin O V., Shamshurin D, Zhang X, Taillefer M, Blunt W, Cicek N, Sparling R, Levin DB. 2016. Transcriptomic and proteomic analyses of core metabolism in *Clostridium termitidis* CT1112 during growth on α-cellulose, xylan, cellobiose and xylose. BMC Microbiol 16.

380. Lyristis M, Boynton ZL, Petersen D, Kan Z, Bennett GN, Rudolph FB. 2000. Cloning, sequencing, and characterization of the gene encoding flagellin, *flaC,* and the post-translational modification of flagellin, FlaC, from *Clostridium acetobutylicum* ATCC824. Anaerobe 6:69–79.

381. Alsaker K V., Paredes C, Papoutsakis ET. 2010. Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. Biotechnol Bioeng 105:1131–47.

382. Balodimos IA, Rapaport E, Kashket ER. 1990. Protein phosphorylation in response to stress in *Clostridium acetobutylicum*. Appl Environ Microbiol 56:2170–2173.

383. Bai X, Ji Z. 2012. Phosphoproteomic investigation of a solvent producing bacterium *Clostridium acetobutylicum*. Appl Microbiol Biotechnol 95:201–211.

References

384. Salimi F, Mandal R, Wishart D, Mahadevan R. 2010. Understanding *Clostridium acetobutylicum* ATCC 824 metabolism using genome-scale thermodynamics and metabolomics-based modeling, p. 126–131. *In* IFAC Proceedings Volumes (IFAC-PapersOnline). Elsevier.

385. Yoo M, Bestel-Corre G, Croux C, Riviere A, Meynial-Salles I, Soucaille P. 2015. A quantitative system-scale characterization of the metabolism of *Clostridium acetobutylicum*. MBio 6:e01808-15-.

386. Rabinowitz JD, Aristilde L, Amador-Noguez D. 2015. Metabolomics of clostridial biofuel production. Argonne, IL (United States).

387. Aristilde L, Lewis IA, Park JO, Rabinowitz JD. 2015. Hierarchy in pentose sugar metabolism in *Clostridium acetobutylicum*. Appl Environ Microbiol 81:1452–1462.

388. Kolek J, Patáková P, Melzoch K, Sigler K, Řezanka T. 2015. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as determined by lipidomic analysis. PLoS One 10:e0122058.

389. Zhao X, Condruz S, Chen J, Jolicoeur M. 2016. A quantitative metabolomics study of high sodium response in *Clostridium acetobutylicum* ATCC 824 acetone-butanol-ethanol (ABE) fermentation. Sci Rep 6:1–13.

390. Adler HI, Crow W. 1987. A Technique for Predicting the Solvent-Producing Ability of *Clostridium acetobutylicum*. Appl Environ Microbiol 53:2496–2499.

391. Tracy BP, Gaida SM, Papoutsakis ET. 2010. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. Curr Opin Biotechnol 21:85–99.

392. González-Peñas H, Lu-Chau TA, Moreira MT, Lema JM. 2015. Assessment of morphological changes of *Clostridium acetobutylicum* by flow cytometry during acetone/butanol/ethanol extractive fermentation. Biotechnol Lett 37:577–584.

393. Kolek J, Branska B, Drahokoupil M, Patakova P, Melzoch K. 2016. Evaluation of viability, metabolic activity and spore quantity in clostridial cultures during ABE fermentation. FEMS Microbiol Lett 363:fnw031.

394. Charubin K, Modla S, Caplan JL, Papoutsakis ET. 2020. Interspecies microbial fusion and large-scale exchange of cytoplasmic proteins and rna in a syntrophic *Clostridium* coculture. MBio 11:1–20.

395. Xu X, Qi LS. 2018. A CRISPR–dCas toolbox for genetic engineering and synthetic biology. J Mol Biol. Academic Press.

396. Minton NP, Ehsaan M, Humphreys CM, Little GT, Baker J, Henstra AM, Liew F, Kelly ML, Sheng L, Schwarz K, Zhang Y. 2016. A roadmap for gene system development in *Clostridium*. Anaerobe https://doi.org/10.1016/j.anaerobe.2016.05.011.

397. Sandoval-Espinola WJ, Chinn MS, Thon MR, Bruno-Bárcena JM. 2017. Evidence of mixotrophic carbon-capture by n-butanol-producer *Clostridium beijerinckii*. Sci Rep 7:12759.

398. Wang Y, Li X, Mao Y, Blaschek HP. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. BMC Genomics 12:479.

399. Potel CM, Lin M-H, Prust N, van den Toorn HWP, Heck AJR, Lemeer S. 2019. Gaining Confidence in the elusive histidine phosphoproteome. Anal Chem 91:5542–5547.

400. Hardman G, Eyers CE. 2020. High-throughput characterization of histidine phosphorylation sites using upax and tandem mass spectrometry, p. 225–235. *In* Methods in Molecular Biology. Humana Press Inc.

401. Rohani A, Moore JH, Su YH, Stagnaro V, Warren C, Swami NS. 2018. Single-cell electro-phenotyping for rapid assessment of *Clostridium difficile* heterogeneity under vancomycin treatment at sub-MIC (minimum inhibitory concentration) levels. Sensors Actuators, B Chem 276:472–480.

402. Jehan Z. 2019. Single-cell omics: An overview Single-Cell Omics: Volume 1: Technological Advances and Applications. Elsevier Inc.

403. Lundberg E, Borner GHH. 2019. Spatial proteomics: a powerful discovery tool for cell biology. Nat Rev Mol Cell Biol 20:285–302.

404. Narendra DP, Steinhauser ML. 2020. Metabolic analysis at the nanoscale with Multi-Isotope Imaging Mass Spectrometry (MIMS). Curr Protoc Cell Biol 88:1–24.

405. Schuster KC, Van Den Heuvel R, Gutierrez NA, Maddox IS. 1998. Development of markers for product formation and cell cycle in batch cultivation of *Clostridium acetobutylicum* ATCC 824. Appl Microbiol Biotechnol 49:669–676.

406. Liu D, Yang Z, Chen Y, Zhuang W, Niu H, Wu J, Ying H. 2018. *Clostridium acetobutylicum* grows vegetatively in a biofilm rich in heteropolysaccharides and cytoplasmic proteins. Biotechnol Biofuels 11:315.

407. Liew F, Martin ME, Tappel RC, Heijstra BD, Mihalcea C, Köpke M. 2016. Gas Fermentation—A flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. Front Microbiol 7.

408. LanzaTech. 2019. LanzaTech | Capturing carbon. Fueling growth. LanzaTech.

409. Tan Y, Liu Z-Y, Liu Z, Zheng H-J, Li F-L. 2015. Comparative transcriptome analysis between *csrA*-disruption *Clostridium acetobutylicum* and its parent strain. Mol Biosyst 11:1434–1442.

410. Abrini J, Naveau H, Nyns E-J. 1994. *Clostridium autoethanogenum,* sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. Arch Microbiol 161:345–351.

411. Liou JS-C, Balkwill DL, Drake GR, Tanner RS. 2005. *Clostridium carboxidivorans* sp. nov., a solvent-producing *Clostridium* isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. Int J Syst Evol Microbiol 55:2085–2091.

412. Huhnke RL, Lewis RS, Tanner RS. August 2008. Isolation and characterization of novel clostridial species. US7704723 B2. Patent. U.S. Patent and Trademark Office.

413. Zahn JA, Saxena J. August 2011. Novel ethanologenic *Clostridium* species, *Clostridium coskatii*. Patent.

414. Philips J, Rabaey K, Lovley DR, Vargas M. 2017. Biofilm formation by *Clostridium ljungdahlii* is induced by sodium chloride stress: experimental evaluation and transcriptome analysis. PLoS One 12:e0170406.

415. Aklujkar M, Leang C, Shrestha PM, Shrestha M, Lovley DR. 2017. Transcriptomic profiles of *Clostridium ljungdahlii* during lithotrophic growth with syngas or H_2 and CO_2 compared to organotrophic growth with fructose. Sci Rep 7:13135.

416. Whitham JM, Tirado-Acevedo O, Chinn MS, Pawlak JJ, Grunden AM. 2015. Metabolic response of *Clostridium ljungdahlii* to oxygen exposure. Appl Environ Microbiol 81:8379–8391.

417. Diender M, Parera Olm I, Gelderloos M, Koehorst JJ, Schaap PJ, Stams AJM, Sousa DZ. 2019. Metabolic shift induced by synthetic co-cultivation promotes high yield of chain elongated acids from syngas. Sci Rep 9:18081.

418. Hiral S, Abhishek M, Aruna G, Annamma A, Arvind L. 2017. Enhanced acidogenesis by the degenerated *Clostridium* sp. strain on a continuous membrane cell recycle reactor. Adv Biotechnol Microbiol 7.

419. Kashket ER, Cao ZY. 1993. Isolation of a degeneration-resistant mutant of *Clostridium acetobutylicum* NCIMB 8052. Appl Environ Microbiol 59:4198–4202.

420. Lee J, Blaschek HP. 2001. Glucose Uptake in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. Appl Environ Microbiol 67:5025–5031.

421. Chen T, Liu N, Ren P, Xi X, Yang L, Sun W, Yu B, Ying H, Ouyang P, Liu D, Chen Y. 2019. Efficient biofilm-based fermentation strategies for l-threonine production by *Escherichia coli*. Front Microbiol 10:1773.

422. Liang C, Ding S, Sun W, Liu L, Zhao W, Zhang D, Ying H, Liu D, Chen Y. 2020. Biofilm-based fermentation: a novel immobilisation strategy for *Saccharomyces cerevisiae* cell cycle progression during ethanol production. Appl Microbiol Biotechnol 104:7495–7505.

423. Pantaléon V, Bouttier S, Soavelomandroso AP, Janoir C, Candela T. 2014. Biofilms of *Clostridium* species. Anaerobe 30:193–198.

424. Liu D, Chen Y, Ding FY, Zhao T, Wu JL, Guo T, Ren HF, Li BB, Niu HQ, Cao Z, Lin XQ, Xie JJ, He XJ, Ying HJ. 2014. Biobutanol production in a *Clostridium acetobutylicum* biofilm reactor integrated with simultaneous product recovery by adsorption. Biotechnol Biofuels 7:5.

425. Liu D, Xu J, Wang Y, Chen Y, Shen X, Niu H, Guo T, Ying H. 2016. Comparative transcriptomic analysis of *Clostridium acetobutylicum* biofilm and planktonic cells. J Biotechnol 218:1–12.

426. García García T, Ventroux M, Derouiche A, Bidnenko V, Correia Santos S, Henry C, Mijakovic I, Noirot-Gros M-F, Poncet S. 2018. Phosphorylation of the *Bacillus subtilis* replication controller YabA plays a role in regulation of sporulation and biofilm formation. Front Microbiol 9:486.

References

427. Barca C, Ranava D, Bauzan M, Ferrasse JH, Giudici-Orticoni MT, Soric A. 2016. Fermentative hydrogen production in an up-flow anaerobic biofilm reactor inoculated with a co-culture of *Clostridium acetobutylicum* and *Desulfovibrio vulgaris*. Bioresour Technol 221:526–533.

428. Benomar S, Ranava D, Cárdenas ML, Trably E, Rafrafi Y, Ducret A, Hamelin J, Lojou E, Steyer J-PP, Giudici-Orticoni M-TT. 2015. Nutritional stress induces exchange of cell material and energetic coupling between bacterial species. Nat Commun 6:1–10.

429. Engel M, Gemünde A, Holtmann D, Müller-Renno C, Ziegler C, Tippkötter N, Ulber R. 2020. *Clostridium acetobutylicum*'s connecting world: cell appendage formation in bioelectrochemical systems. ChemElectroChem 7:414–420.

430. Ranava D, Backes C, Karthikeyan G, Ouari O, Soric A, Guiral M, Cárdenas ML, Giudici-Orticoni MT. 2020. Metabolic exchange and energetic coupling between nutritionally stressed bacterial species, the possible role of QS molecules. bioRxiv 2020.04.27.063917.

431. Yu EKC, Chan MKH, Saddler JN. 1985. Butanol production from cellulosic substrates by sequential co-culture of *Clostridium thermocellum* and *C. acetobutylicum*. Biotechnol Lett 7:509–514.

432. Wu P, Wang G, Wang G, Børresen BT, Liu H, Zhang J. 2016. Butanol production under microaerobic conditions with a symbiotic system of *Clostridium acetobutylicum* and *Bacillus cereus*. Microb Cell Fact 15:8.

433. Zagrodnik R, Łaniecki M. 2017. Hydrogen production from starch by co-culture of *Clostridium acetobutylicum* and *Rhodobacter sphaeroides* in one step hybrid dark- and photofermentation in repeated fed-batch reactor. Bioresour Technol 224:298–306.

434. Gundlach J, Rath H, Herzberg C, Mäder U, Stülke J. 2016. Second messenger signaling in *Bacillus subtilis*: accumulation of cyclic di-amp inhibits biofilm formation. Front Microbiol 7.

435. Purcell EB, McKee RW, McBride SM, Waters CM, Tamayo R. 2012. Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. J Bacteriol 194:3307–3316.

436. Kobir A, Shi L, Boskovic A, Grangeasse C, Franjevic D, Mijakovic I. 2011. Protein phosphorylation in bacterial signal transduction. Biochim Biophys Acta - Gen Subj 1810:989–994.

437. Dartois V, Djavakhishvili T, Hoch JA. 1997. KapB is a lipoprotein required for KinB signal transduction and activation of the phosphorelay to sporulation in *Bacillus subtilis*. Mol Microbiol 26:1097–1108.

438. Trouve J, Mohamed A, Leisico F, Contreras-Martel C, Liu B, Mas C, Rudner DZ, Rodrigues CDA, Morlot C. 2018. Structural characterization of the sporulation protein GerM from *Bacillus subtilis*. J Struct Biol 204:481–490.

439. Charlton TM, Kovacs-Simon A, Michell SL, Fairweather NF, Tate EW. 2015. Quantitative lipoproteomics in *Clostridium difficile* reveals a role for lipoproteins in sporulation. Chem Biol 22:1562–1573.

List of publications

M. Diallo, N. Kint, M. Monot, I. Martin-Verstraete, J. van der Oost, S. W.M. Kengen, and A. M. López-Contreras. "Characterization of a degenerated *Clostridium beijerinckii* mutant strain." (manuscript in preparation) -This Thesis

M. Diallo, S. W. M. Kengen, and A. M. López-Contreras. "Sporulation in solventogenic clostridia and acetogenic" (submitted) -This Thesis

M. Diallo, N. Kint, M. Monot, F.t Collas, I. Martin-Verstraete, J. van der Oost, S. W. M. Kengen, and Ana M. López-Contreras. 2020. "Transcriptomic and phenotypic analysis of a *spoIIE* mutant in *Clostridium beijerinckii*." *Frontiers in Microbiology* 11. https://doi.org/10.3389/fmicb.2020.556064. -This Thesis

M. Diallo *, R. Hocq*, F. Collas, G. Chartier, F. Wasels, H. Surya Wijaya, M. W. T. Werten, et al. 2020. "Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*." Methods (San Diego, Calif.) 172, 51–60. https://doi.org/10.1016/j.ymeth.2019.07.022. -This Thesis

M. Diallo *, A. D. Simons*, H. van der Wal, F. Collas, B. Houweling-Tan, S. W. M. Kengen, and Ana M. López-Contreras. 2018. "L-Rhamnose metabolism in *Clostridium beijerinckii* strain DSM 6423." Edited by Maia Kivisaar. *Applied and Environmental Microbiology* 85 (5). https://doi.org/10.1128/AEM.02656-18. -This Thesis

J. Kolek, **M. Diallo**, M. Vasylkivska, B. Branska, K. Sedlar, A. M. López-Contreras, and P. Patakova. 2017. "Comparison of expression of key sporulation, solventogenic and acetogenic genes in *C. beijerinckii* NRRL B-598 and Its mutant strain overexpressing Spo0A." *Applied Microbiology and Biotechnology* 101(22), 1–13. https://doi.org/10.1007/s00253-017-8555-3. (not in this thesis)

N. Pakharukova, J. A. Garnett, M. Tuittila, S. Paavilainen, **M. Diallo**, Y. Xu, S. J. Matthews, A.V. Zavialov (2015). Structural Insight into Archaic and Alternative Chaperone-Usher Pathways Reveals a Novel Mechanism of Pilus Biogenesis. PLoS Pathogens, 11(11). https://doi.org/10.1371/journal.ppat.1005269 (not in this thesis)

S. Rumpel, J. F. Siebel, **M. Diallo**, C. Farès, E. J. Reijerse, W. Lubitz (2015). Structural Insight into the Complex of Ferredoxin and [FeFe] Hydrogenase from *Chlamydomonas reinhardtii*. ChemBioChem, 1663–1669. https://doi.org/10.1002/cbic.201500130 (not in this thesis)

J. A. Garnett, **M. Diallo**, & S. J. Matthews. (2015). Purification, crystallization and preliminary X-ray diffraction analysis of the *Escherichia coli* common pilus chaperone EcpB. Acta Crystallographica Section F: Structural Biology Communications, 71(6), 676–679. https://doi.org/10.1107/S2053230X15006354 (not in this thesis)

*Authors contributed equally

Overview of completed training activities

Discipline-specific activities

Courses

- Gene Road Map Training (University of Nottingham, Nottingham, UK, 2015)

- Insight into Biofuel Production (Green Biologics Ltd., Oxford, UK, 2015)

- Annotation and analysis of prokaryotic genomes (LABGeM, Evry. FR, 2016)

- Corbion Training Workshop (Corbion, Gorinchem, NL, 2016)

- Functional Genomics & RNA.seq Workshop (Institut Pasteur, Paris, FR, 2016)

- Cell and Developmental Biology of *Clostridium* (ITQB Nova Lisbon, Lisabon, PT, 2017)

- Translation - From science to applications (Nizo Food Research, Ede, NL, 2018)

- Biorefineries - the role of microbiology and biotechnology (Wageningen Food and Biobased Reasearch, Wageningen, NL, 2018)

Secondments

- Functional Genomics & RNA.seq (Institut Pasteur, Paris, FR, 2016)

- Genome engineering of *Clostridium beijerinckii* (IFPEN, Rueil Malmaison, 2018)

Meetings and Conferences

- Mid-term progress review (European Commission, Brussel, BE, 2016)

- Midterm Conference (ITQB Nova Lisbon, Lisabon, PT, 2017) **

- BioTech 2017 and the 7th Czech-Swiss Symposium (Vysoká škola chemicko-technologická v Praze, Prague, CZ 2017) ***

- Clospore International Showcase Conference (University of Nottingham, Nottingham, UK, 2018) **

- Clostridium XV conference (Technische Universität München, Freising, DE, 2018)*

- FEMS 2019 (FEMS, Glasgow, UK, 2019)*

General courses

- VLAG Ph.D. week (VLAG, Baarlo, NL, 2016)

- Scientific Writing (Wageningen graduate school, NL, 2017)

- Efficient Writing Strategies (Wageningen graduate school, NL, 2018)

- Teaching and supervising Thesis students (Wageningen graduate school, NL, 2018)

- Responsible research and innovation Workshop (Nottingham, UK, 2018)

- Design thinking and innovation (Lorentz Center, NL, 2019)

- Innovation management (Erasmus University Rotterdam, NL, 2020)

- Graphic design (Udemy, 2020)

Optional

- Research proposal (WUR, NL, 2015)

- Bi-weekly group meetings

- Clospore Netherlands meetings (once per semester)

- European Researchers Night (Marie Skłodowska-Curie actions, 2018)*

- Ph.D. study trip in the USA 2019 ***

*Poster presentation, **Oral presentation

About the author

Mamou Diallo was born on October 21th, 1991, in Boulogne-Billancourt, in France. Through her father, she discovered the world of microbiology and molecular biology at a very young age. In 2001, she moved with her family to Vienna, Austria, where she studied at the French Lycée and obtained her Baccalauréat in 2009. She moved back to France for her undergraduate studies in a "classe preparatoire" in Biology, Chemistry, Physics, and Earth Science (BCPST) at the Lycée St Louis, in Paris. There, she prepared the examination to enter the "Grandes Ecoles". After passing the exam, she was admitted to Montpellier SupAgro, in the south of France, where she pursued a master's degree in agricultural sciences engineering. During her studies, she did an academic exchange at the Imperial College in London, where she took courses in biochemistry and worked in the Department of Life Sciences under Dr. James Garnett and Prof. Steve Matthews on the characterization of the *Escherichia coli* common pilus (Ecp).

She further took a gap year (2013-2014) to explore different research focuses in the field of biobased products. As a research assistant at the Max Planck Institute for Chemical Energy Conversion in Germany, Mamou studied the [FeFe] Hydrogenase from *Chlamydomonas reinhardtii* under the supervision of Dr. Sigrun Rumpel and Dr. Judith Siebel. She also investigated biochar characteristics when working for the International Research and Cooperation Center in Agronomy and Development (CIRAD) and biofuel policies in developing countries as a junior analyst at the French Development Agency (AFD). These experiences convinced her to specialize in chemical and bioprocesses for sustainable development (Green Chemistry) to deepen her knowledge in bioengineering and bioprocesses. She did her MSc thesis at the agro-polymers and emerging technologies (IATE) research unit at Montpellier SupAgro under Prof. Eric Dubreucq's supervision. During her thesis, she worked on a depolymerization and functionalization process of bark tannins to produce biobased phenolic monomers.

After graduating in 2015 from the Montpellier SupAgro, she moved to Wageningen to start her Ph.D. research at the Wageningen Food and Biobased Research Institute (2015-2020). Her doctoral research was part of the European Union-funded Marie Skłodowska-Curie Innovative Training Network CLOSPORE, which was focused on sporulation in *Clostridium*. Her research focused on genome engineering tools for *C. beijerinckii* and the regulatory networks governing sporulation and solvent formation in *C. beijerinckii*. The main results of this research have been described in this thesis.

Acknowledgments

Despite being a personal journey, this thesis would not have been made possible without the help and support of many. Thus, I would like to use few lines to thank those who made this thesis possible, even though words are pale compared to their contribution.

First, I would like to thank my promotor and co-promotors, **Prof. John van der Oost, Dr. Ana López Contreras**, and **Dr. Servé Kengen**, without whom this booklet would not have seen the light. Thanks to your words and support, I matured as a researcher. **John**, thank you for accepting me as a member of the Bacterial Genetics group (BacGen) and for being my promotor. I am grateful for your great feedback and warmth that made me always feel welcome in BacGen. **Ana**, I am truly indebted to you for giving me the opportunity to pursue my academic career in your team in the Bioconversion group at the Wageningen Food and Biobased Research Institute (WFBR). You introduced me to the world of clostridia, and I am grateful for your guidance and encouragement throughout my thesis journey. **Servé**, I am thankful for your sound advice and showed me how to better articulate my thoughts and appreciate my results. I am also grateful to the other members of the committee for joining the committee and taking the time to evaluate my work.

In addition to my promotors, I was lucky to be surrounded by experienced researchers who mentored me through my research. **Florent**, thank you for taking the time to teach me your know-how on clostridia and molecular biology. Thanks to you, my first year of Ph.D. was very insightful, and I acquired the ground knowledge to tackle my thesis. I am also grateful for your advice and for taking the time to help me even after you left Wageningen.

Marc, **Emil**, **Jan S.**, thank you for keeping the labs running and sharing your molecular biology expertise with me. **Myriam**, **Hetty**, **Bwee**, sorry for the mess I might have caused sometimes in the fermentation laboratory, and thank you as well for keeping the labs running and for helping me whenever I needed advice or guidance.

Marjo, **Toine**, **Hans v. d. K.** thank you for sharing your good mood with me. **Truus**, **Jan** and **Ruud**, sharing our green office space with you was enjoyable. Thanks to the atmosphere you created, I could focus on my research and enjoy tasty sweets during tea breaks. Next to our pleasant conversation and laughs, you provided me with many advice and encouraging words; thank you.

Andre, I enjoyed working and exchanging with you; thank you for being a great colleague and still available whenever I need your advice. I am pleased that our work resulted in an article!